University of Reading

Unlocking the potential of ensemble forecasts in tropical cyclone forecasting and early flood warning

A thesis submitted for the degree of Doctor of Philosophy

Department of Geography & Environmental Science

Helen Ann Titley

July 2023

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Helen Titley

Abstract

Ensemble forecasting is widely recognised as a vital method to improve the skill and utility of weather forecasts. This research aims to unlock the potential of dynamic ensemble forecasts to improve operational tropical cyclone (TC) forecasting and early flood warnings. Firstly, a survey of operational TC forecast centres reveals the huge potential to improve the pull-through of ensemble forecasts in operational TC forecasts and warnings, identifying hurdles that need to be overcome to allow their full utilisation. A global evaluation of multi-model ensemble TC track probability forecasts highlights the benefit of using ensembles over a deterministic consensus forecast and demonstrates the added value gained from using multi-model ensembles. Focus then shifts to the potential use of ensembles in TC fluvial flood forecasting via the Global Flood Awareness System (GloFAS). The severity of flooding in 280 TC cases in terms of flood area, duration and magnitude is calculated using GloFAS-ERA5 reanalyses, and the key factors influencing this severity are assessed. Slow-moving, large, and intense cyclones, affecting areas with wet antecedent conditions, have the highest likelihood of causing widespread flooding, although flooding can be severe even in weaker storms. Finally, novel ensemble-based methods are developed to analyse predictability links along the GloFAS TC flood forecasting chain, demonstrated using the case of Hurricane Iota. This provides an in-depth look at the ability of GloFAS to forecast river flooding in a TC case, and reveals how the strength of the predictability links between TC track, intensity, rainfall and river discharge varies depending on catchment location. As numerical weather prediction (NWP) centres move towards purely ensemble-based forecast systems, the importance of unlocking their potential in TC forecasts and warnings grows. This project leaves a legacy of increased focus on this important topic within international forecasting and research communities via its links to World Meteorological Organisation (WMO) policy and wider collaborative projects.

Acknowledgments

I would first and foremost like to thank my supervisors, Hannah Cloke and Liz Stephens. Their support, advice and guidance throughout my PhD was crucial to enable me to make it through the last five years. They have a natural ability to know when I need to be pushed further and outside of my comfort zone, but perhaps even more importantly when to encourage me to take time away to rest and recuperate when juggling PhD, work and family life all gets too much. They have always believed in me and helped me to build up my confidence, for which I will always be indebted to them. I am also incredibly grateful to my Met Office line manager and friend Joanne Robbins, who has always wholeheartedly supported me with my PhD studies and has helped me to fit the PhD work with my work schedule. Jo is also my inspiration, having done her own PhD a few years ago. Without seeing her succeed in this I would never have taken the leap to study for one myself. Further thanks go to Florian Pappenberger, Christel Prudhomme, and the whole of the GloFAS team at ECMWF, for their support during my PhD. Ervin Zsótér, David Richardson, and Rebecca Emerton are owed a special mention for being such helpful PhD buddies at various stages of my journey.

I would also like to thank my friends and colleagues in the Weather Impacts team at the Met Office, and in my wider Met Office friendship and support group. Special mentions to Julian Heming for introducing me to the tropical cyclone world and from whom I have learnt so much, and Becky Bowyer for her help with the tropical cyclone verification. I have also really enjoyed being part of the Water@Reading research group at the University of Reading; I am grateful to have felt so included by everyone in spite of me being part time and remote from Reading.

I am also thankful to all my friends who have been a constant support through the PhD, often by understanding that I really didn't want to talk about how it was going in those times when things were more of a struggle. I am really looking forward to celebrating with you all with a glass (or two) of prosecco when it is finally complete. My lifelong friend Lizzie has been another valued inspiration and guide, having completed her own part-time PhD several years ago.

I am incredibly grateful to my family: to my devoted parents, David and Judith Watkin, and my amazing sister Emma, for their support throughout my life; to my lovely late grandparents; and to my very supportive wider family and family-in-law. Finally, the biggest thank you of all has to go to my hugely supportive husband, Neil, and my two wonderful children, Evie and Ben, who have had to suffer the consequences of me being busy and stressed with work and the PhD and who have always supported me enthusiastically every step of the way, regardless.

A huge thank you goes to you all!

Contents

Declaration i
Abstract iii
Acknowledgments v
Contents vii
List of acronyms xi
My personal story of the PhD xv
Chapter 1 Introduction 1
1.1 Motivation and aim 1
1.2 Objectives and research questions 2
1.2.1 Use of ensemble forecasts in operational tropical cyclone forecasting 2
1.2.2 Evaluation of tropical cyclone track probability forecasts 2
1.2.3 Key factors influencing severity of tropical cyclone flood hazard 3
1.2.4 Analysing predictability links in the tropical cyclone flood forecast chain 4
1.3 Structure of the thesis 4
Chapter 2 Literature review 7
2.1 Ensemble forecasting 7
2.1.1 Motivation and history 7
2.1.2 Current status 7
2.2 Tropical cyclones 8
2.2.1 Climatology 8
2.2.2 Rainfall 9
2.2.3 Fluvial flood hazard and impact 11
2.3 Forecasting of tropical cyclones 13
2.3.1 Operational forecasting and global co-operation 13
2.3.2 Forecasting track and intensity 14
2.3.3 Forecasting precipitation and flooding 15
2.3.4 Probabilistic forecasting of tropical cyclones 17
2.3.5 Forecasting for humanitarian preparedness and response 19
2.4 Verification methods for tropical cyclone forecasts 19
2.4.1 Traditional verification metrics 19
2.4.2 Ensemble-based verification metrics 20
2.4.3 Feature-based verification metrics 21
2.5 River flood forecasting 21
2.5.1 Background 21

2.5.2 Ensemble flood forecasting 23
2.5.3 GloFAS 23
2.5.4 Verification methods 25
2.6 Summary 26
Chapter 3 Current and potential use of ensemble forecasts in operational TC forecasting: results from a global forecaster survey 27
3.1 Introduction and aims of the questionnaire 28
3.2 Results of the questionnaire 30
3.2.1 Current availability and use of ensemble forecasts 30
3.2.2 How is uncertainty represented in operational warnings? 35
3.2.3 Integration of probabilistic forecasts into operations 39
3.2.4 Priorities for future research and development 41
3.3 Conclusions and recommendations 44
Appendix 3-1: Survey questions 46
Chapter 4 A global evaluation of multi-model ensemble tropical cyclone track probability forecasts 51
4.1 Introduction 52
4.2 Verification methodology 55
4.3 Results 58
4.3.1 Overall track probability forecasts for named tropical cyclones 58
4.3.2 Track probability forecasts for named tropical cyclones over each 24-hour forecast period 62
4.3.3 Verification for two high-profile tropical cyclones 67
4.4 Discussion 68
4.5 Conclusions and future plans 72
Chapter 5 Key factors influencing the severity of fluvial flood hazard from tropical cyclones 75
5.1 Introduction 76
5.2 Data sources 78
5.2.1 IBTrACS 78
5.2.2 Global Precipitation Measurement (GPM) IMERG 78
5.2.3 ERA5 78
5.2.4 GloFAS-ERA5 reanalysis 79
5.3 Methods 80
5.3.1 Selecting tropical cyclone cases and calculating their land footprint 80
5.3.2 Tropical cyclone characteristics 82
5.3.3 Catchment characteristics 84
5.3.4 Calculating flood severity 84

5.4 Results 85
5.4.1 Comparison of flood severity measures 85
5.4.2 Relationships between flood severity and case characteristics 87
5.5 Discussion of the key outcomes 94
5.5.1 Applying flood severity metrics to tropical cyclones footprints 94
5.5.2 Importance of precipitation and translation speed 95
5.5.3 Presence of flood hazard at all levels of storm intensity 95
5.5.4 Complex relationship between flood severity and antecedent soil moisture 95
5.5.5 Interplay of characteristics in modulating flood severity 96
5.5.6 Implications for research into the predictability of fluvial flooding from tropical cyclones 96
5.5.7 Implications for disaster risk reduction, preparedness and response 96
5.6 Conclusions 97
Chapter 6 Using ensembles to analyse predictability links in the tropical cyclone flood forecast chain 101
6.1 Introduction 102
6.2 Case background 103
6.3 Data and methods 104
6.3.1 Tropical cyclone track (position and intensity) 106
6.3.2 Tropical cyclone precipitation 107
6.3.3 River discharge 107
6.3.4 Proposed methodology to evaluate the predictability linkages along the forecast chain 108
6.4 Results 109
6.4.1 Evolution of forecasts for Iota 109
6.4.2 Flow of uncertainty and predictability through the forecast chain 112
6.4.3 Member-specific analysis 115
6.5 Discussion 121
6.6 Conclusion 124
Chapter 7 Discussion 127
7.1 Use of ensemble forecasts in operational tropical cyclone forecasting 127
7.1.1 Scientific knowledge contribution 127
7.1.2 Limitations 128
7.1.3 Additional outcomes and impact of the research 129
7.2 Evaluation of tropical cyclone track probability forecasts 132
7.2.1 Scientific knowledge contribution 132
7.2.2 Limitations 133

7.2.3 Additional outcomes and impact of the research 133
7.3 Key factors influencing severity of tropical cyclone flood hazard 135
7.3.1 Scientific knowledge contribution 135
7.3.2 Limitations 136
7.3.3 Additional outcomes and impact of the research 137
7.4 Analysing predictability links in the tropical cyclone flood forecast chain 138
7.4.1 Scientific knowledge contribution 138
7.4.2 Limitations 139
7.4.3 Additional outcomes and impact of the research 140
7.5 Additional research towards the aim of this thesis 141
7.5.1 WMO activities on ensemble pull-through in TC forecasting 141
7.5.2 WCSSP research in support of this thesis 143
7.5.3 Emergency flood bulletins to support humanitarian decision-making in TC cases .. 145
7.6 Recommended next steps 146
Chapter 8 Conclusions 151
References 155
Appendices 179
A1: Published Article: Current and potential use of ensemble forecasts in operational TC forecasting: results from a global forecaster survey 181
A2: Published Article: A global evaluation of multi-model ensemble tropical cyclone track probability forecasts 199
A3: Published Article: Key Factors Influencing the Severity of Fluvial Flood Hazard from Tropical Cyclones 217
A4: Submitted Article: Using ensembles to analyse predictability links in the tropical cyclone flood forecast chain 237
A5: Published Article: Track forecast: Operational capability and new techniques - Summary from the Tenth International Workshop on Tropical Cyclones (IWTC-10) 273
A6: WCSSP India report: The predictability of fluvial flooding from recent Indian cyclones in the Global Flood Awareness System (GloFAS) 293
A7: WCSSP India report: Exploring the use of feature-based verification techniques on ensemble forecasts of tropical cyclone rainfall 323
A8: Published Article: Emergency flood bulletins for Cyclones Idai and Kenneth: A critical evaluation of the use of global flood forecasts for international humanitarian preparedness and response 367
A9: Submitted Article: Recommendations to improve the interpretation of global flood forecasts to support international humanitarian operations for tropical cyclones 399

List of acronyms

ADAM	Advanced Disaster Analysis & Mapping
AG-TC	Advisory Group on Tropical Cyclones
AROME	Numerical prediction model at Meteo-France
ATCF	Automated Tropical Cyclone Forecasting System
AUC	Area under the ROC Curve
AUS	Australian TC basin
BSS	Brier Skill Score
BUFR	Binary Universal Form for the Representation of meteorological data
C3S	Copernicus Climate Change Service
CDS	Copernicus Climate Data Store
CEMS	Copernicus Emergency Management Service
CFS	NCEP Climate Forecast System
CLARE	Climate Adaptation and Resilience research programme
CLIPER	Climatology and Persistence
CRA	Contiguous Rain Area
CRPS	Continuous Ranked Probability Score
CXML	Cyclone XML format data in TIGGE archive
DNGRH	National Directorate of Water Resources Management (Mozambique)
ECLand	ECMWF's land-surface modelling platform
ECMWF	European Centre for Medium-range Weather Forecasting
EDA	Ensemble of data assimilations
ENS	ECMWF ensemble forecasts
ERA5	ECMWF reanalysis version 5
ERA-Interim	ECMWF reanalysis dataset (predecessor to ERA-5)
ERO	Excessive Rainfall Outlook
ETKF	Ensemble Transform Kalman Filter
ETQPF	Ensemble Typhoon Quantitative Precipitation Forecast
FbF	Forecast-based Financing
FCDO	Foreign, Commonwealth and Development Office
FHRFI	Flood Hazard Risk Forecasting Index
FSS	Fractions Skill Score

FV3	Finite-Volume Cubed-Sphere Dynamical Core
GDACS	Global Disaster Alert and Coordination System
GDPFS	Global Data-processing and Forecasting System
GEFS	Global Ensemble Forecast System
GEPS	Global Ensemble Prediction System
GFS	Global Forecast System
GGU	Global Guidance Unit
GloFAS	Global Flood Awareness System
GPM	NASA's Global Precipitation Measurement Mission
GRDC	Global Runoff Data Centre
GTS	WMO Global Telecommunications System
HEPEX	Hydrologic Ensemble Prediction Experiment
HIWeather	High-Impact Weather Project
HTESSEL	Hydrology Tiled ECWMF Scheme for Surface Exchange over Land
HRRR	High Resolution Rapid Refresh (model)
HWRF	Hurricane Weather Research and Forecasting (model)
IMERG	Integrated Multi-satellitE Retrievals for GPM
IFS	Integrated Forecasting System
IMD	India Meteorological Department
INAM	Mozambique National Institute of Meteorology
IPCC	Intergovernmental Panel on Climate Change
IWTC	International Workshop on Tropical Cyclones
JASMIN	The UK's data analysis facility for environmental science
JHM	Journal of Hydrometeorology
JMA	Japanese Meteorological Agency
JRC	Joint Research Centre of the European Commission
KGE	Kling–Gupta Efficiency
Lisflood	Spatially distributed water resources model, developed by the Joint Research Centre of the European Commission
MAE	Mean Absolute Error
MODE	Method for Object-based Diagnostic Evaluation
MOGREPS	Met Office Global and Regional Ensemble Prediction System
MOGREPS-G	Met Office Global Ensemble

MSLP	Mean sea level pressure
NASA	National Aeronautics and Space Administration
NAT	North Atlantic Ocean TC basin
NCEP	National Centres for Environmental Predictions
NCMRWF	National Centre for Medium Range Weather Forecasting
NEP	North East Pacific Ocean TC basin
NEPS-G	NCMRWF Global Ensemble Prediction System
NHC	U.S. National Hurricane Center
NHM	Nonhydrostatic Model
NI	North Indian Ocean TC basin
NMHS	National Meteorological and Hydrological Services
NOAA	National Oceanic and Atmospheric Administration
NSE	Nash–Sutcliffe model Efficiency
NWC	National Water Center
NWP	Numerical Weather Prediction
NWP	North West Pacific Ocean TC basin
NWP-TCEFP	North Western Pacific Tropical Cyclone Ensemble Forecast Project
NWS	U.S. National Weather Service
PAGASA	Philippine Atmospheric, Geophysical and Astronomical Services Administration
PDEF	Predictability, Dynamics and Ensemble Forecasting
QJRMS	Quarterly Journal of the Royal Meteorological Society
QPF	Quantitative Precipitation Forecasting
R2O	Research to Operations
REPRESA	REsilience and PReparedness to tropical cyclones across Southern Africa
RFC	River Forecast Centers (U.S. NWS)
RMSE	Root mean square error
RMW	Radius of the maximum wind
ROC	Relative Operating Characteristic
ROCI	Radius of the outermost closed isobar
RSMC	Regional Specialized Meteorological Center
SAL	Structure Amplitude Location method
SHEAR	Science for Humanitarian Emergencies and Resilience

SPPT	Stochastically Perturbed Parameterization Tendencies
SST	Sea surface temperature
SWI	South West Indian Ocean TC basin
TC	Tropical Cyclone
TCP	Tropical Cyclone Programme
TC-PFP	Tropical Cyclone Probabilistic Forecast Products
TCWC	Tropical Cyclone Warning Center
TIGGE	THORPEX Interactive Grand Global Ensemble
TT-TCPara	Task Team on Tropical Low/Cyclone Vortex Parameters
UTC	Coordinated Universal Time
WCSSP	Weather and Climate Science to Service Partnership
WFP	World Food Programme
WMO	World Meteorological Organization
WPC	Weather Prediction Center
WRF	Weather Research and Forecasting (model)
WWRP	World Weather Research Programme

My personal story of the PhD

My name is Helen Titley and I had the privilege to write this PhD thesis. This section gives a background to me, my decision to undertake this PhD journey, and my experience during the five long years from 2018 to 2023.

I came to a PhD much later in life than most, starting it around twenty years after I completed my undergraduate degree in Geography at the University of Cambridge, and seventeen years after completing my Meteorology MSc at the University of Birmingham and beginning my career in the Met Office in 2001. I didn't feel that not having a PhD had held me back in my Met Office career, although name badges at conferences sometimes had a habit of making me feel like an old washer woman, with my 'Mrs' name rather than everyone else's 'Dr' or 'Professor'!

However, I became inspired by my friend (and now manager) Joanne Robbins, after watching her successfully complete a part-time PhD a few years before, and by my best friend from school, Lizzie, who also completed a part-time PhD a few years ago. This coincided with me developing an idea for a research topic that could potentially form a PhD project. For many years my Met Office work had focussed on developing tools and applications using ensemble forecasts to help our forecasters to predict high-impact weather. I developed a special interest in tropical cyclones (TCs), and a visit to the 32nd Conference on Hurricanes and Tropical Meteorology in Puerto Rico in 2016 revealed to me that ensemble forecasts were really under-utilised by the operational TC forecasting community, and my personal crusade to try to change this was born! My time for research and writing papers was limited due to ongoing product development and maintenance commitments, and carrying out a part-time PhD would give me a way to change this and research my topic of interest in more depth.

The next step was finding a PhD supervisor, and I responded to a call from Hannah Cloke at the University of Reading, and together we formulated a PhD around increasing the pull-through of ensemble forecast data in operational TC forecasting and early flood warning. I was incredibly lucky to find Hannah, and my second supervisor, Liz Stephens, as I couldn't have wished for nicer and more supportive supervisors. I am incredibly grateful to the Met Office, who have funded the PhD and given me the space to carry out the additional research while still continuing in my existing Met Office role.

From a personal perspective, my husband and children were really supportive of the idea of me doing a PhD. I had worked part-time since having my children in 2008 and 2010, and we decided that now the children were a bit older (10 and 7 at the start of the PhD), I could go back full-time

to enable me to do something for me and start the PhD. Luckily their support has never wavered through the years, even in the tough times.

The detailed plan of the PhD has changed a little over the years, as I'm sure is the case in most PhDs, but the overall idea, to start by looking at the general use of ensembles in operational TC forecasting and the benefit of multi-model ensemble track probability forecasts, before moving on to investigate the predictability of river flooding in TCs using the Global Flood Awareness System (GloFAS), has remained. Hannah is a hydrologist, and a particular appeal of the PhD was the chance to look in detail at river flooding in TCs, and work with ECMWF to investigate the use of their GloFAS system in TCs, which no-one had previously done.

For the first couple of years all went well. I managed to create a sustainable mode of working where I was based in Exeter for half of the week to carry out my Met Office work, and based at home for the rest of the week to carry out the PhD. The initial research that became Chapters 3 and 4 was progressing well and I really enjoyed and valued my monthly visits to Reading to see Hannah and visit the GloFAS team in ECMWF to start the flooding side of the PhD. Then, in March 2020, the world and my PhD entered a new more challenging phase as Covid hit and everything changed. Suddenly I was also balancing home-schooling two children (one at secondary and one at primary school, with very differing needs), the trips to Reading and ECMWF had to stop, and the home/work/PhD split became an impossible juggling act and a chaotic mix of competing demands. Although eventually lockdowns and home-schooling came to an end, somehow I've never quite got back to the equilibrium that I had before in this new post-covid world, so the final three years have definitely been more challenging than the first two.

That being said, I have continued to enjoy the opportunities that the PhD has given me. I have really valued the wider collaborations that have resulted, such as the opportunities to join important WMO activities and the chance to be involved in the flood forecast bulletin work in support of humanitarian activities. I'm incredibly proud that in spite of the challenges that have come my way, I have completed the PhD in five years, with four published papers resulting from the research, and momentum to take the ideas forward in the future.

Hopefully, I have inspired my children, Evie and Ben (who are now 15 and 12), to always believe in yourself, stick at things even when the road gets bumpy, and follow your dreams.



Figure 0-1. My beloved helper team, the Titley family, at the start of my PhD journey in 2018 (left), and near the end of it in 2023 (right). We've all aged quite a bit during the process!



Figure 0-2. My supervisor Hannah and I manning the University of Reading "Beware: floods ahead!" exhibit at the Royal Society's Summer Science Exhibition in London in July 2023.

Chapter 1 Introduction

1.1 Motivation and aim

Ensemble forecasting is a method developed in numerical weather prediction (NWP), where instead of running just a single forecast of the most likely weather, the computer model is run a number of times from slightly different starting conditions and/or model formulations, aiming to take account of the main sources of forecast uncertainty. The resulting set (or ensemble) of forecasts gives a much better idea of possible forecast outcomes at a particular time, and can be used to create probabilistic forecasts and to assess how much confidence there is in the forecast (Sivillo et al., 1997). Thirty years on from the production of the first ensemble forecasts in 1992 (Buizza, 2019), ensemble forecasting has become widely recognised as a key future path for improving the forecast skill and utility of weather forecasts and warnings (Palmer, 2019). A recent review of the impact of risk messages that include probability information strongly suggested that people do understand the probabilistic information and can use it to make better decisions providing it is appropriately presented (Ripberger et al., 2022b). However, in operational tropical cyclone (TC) forecasting there has been only limited pull-through of ensemble-based probability and uncertainty information into operational forecasts and warnings (Yamaguchi et al., 2009; Dupont et al., 2011; Elliott and Yamaguchi, 2014).

More than a decade after the publication of Hamill et al.'s (2012) paper with the strapline "The rapid maturation of ensemble prediction of tropical cyclones allows the possibility of new uncertainty products", the flagship "cone of uncertainty" product from the National Hurricane Center (NHC) still only utilises historical forecast errors rather than exploiting the benefits of the dynamic situation-based uncertainty information that ensembles can provide (NHC, 2023a) (Figure 1-1). So there is huge scope to improve the use of ensembles in TC track forecasting beyond the confines of using the ensemble mean track forecasts in consensus forecasts (e.g. Cangialosi, 2019).

Ensemble forecasts also have largely untapped potential to help translate increased track and intensity forecast skill (Landsea and Cangialosi, 2018; Cangialosi et al., 2020) into TC hazard forecasting, such as flooding from TC rainfall. Recent TC cases where river flooding has led to significant loss of life, such as Cyclones Idai (March 2020) and Freddy (February/March 2023) in Southern Africa and Hurricane Iota (November 2020) in Central America, emphasise the importance of improving the focus on early flood warning in TC cases. Therefore the main aim of this PhD thesis is as follows:

- **Aim: to design and conduct research to facilitate increased pull-through of ensemble forecast data in tropical cyclone forecasting and early flood warning**.

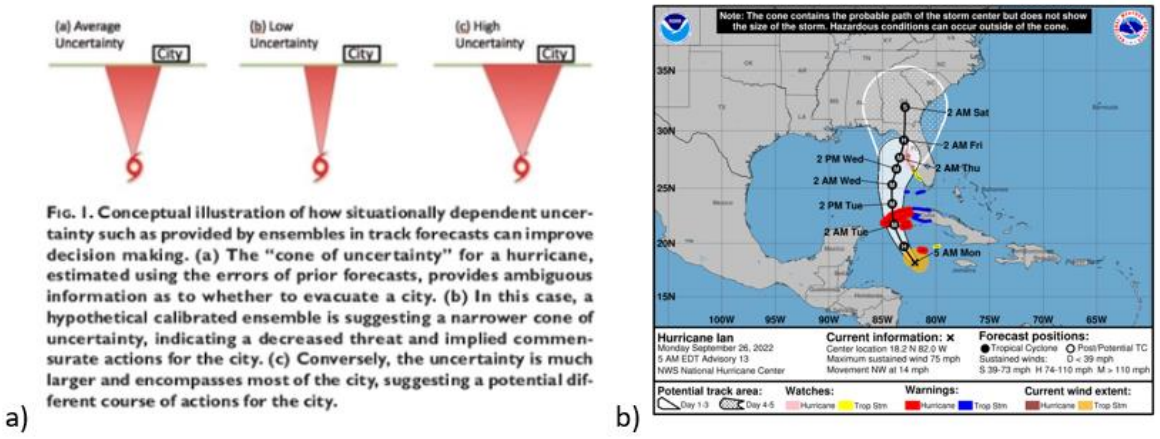


Figure 1-1. Illustration of the lack of progress in pulling through ensemble forecasts into operational TC forecasting: a) Fig. 1 from Hamill et al. (2012) proposing that the cone width in hurricane warnings could vary dependent on ensemble-based situation-dependent uncertainty information; b) Figure from NHC graphic archive for Hurricane Ian ten years later in 2022, where cone width is still purely based on historical error statistics (NHC, 2022).

1.2 Objectives and research questions

1.2.1 *Use of ensemble forecasts in operational tropical cyclone forecasting*

The first essential step was to obtain a baseline knowledge of the current and potential use of ensemble forecast data by operational TC forecast centres such as Regional Specialised Met Centres (RSMCs) and Tropical Cyclone Warning Centres (TCWCs). This led to the first key question: *How are ensemble forecasts currently used in operational tropical cyclone forecasting and warnings, and what potential is there for this to be improved?* This motivated my first objective:

- **Objective 1: Identify the current and potential use of ensemble forecasts by operational tropical cyclone forecast centres.**

This objective was addressed in the study contained in Chapter 3, entitled "*Current and potential use of ensemble forecasts in operational TC forecasting: results from a global forecaster survey*" and published in Tropical Cyclone Research and Review in 2019.

1.2.2 *Evaluation of tropical cyclone track probability forecasts*

One key element to motivate the pull-through in the use of ensemble forecast data in operational TC forecasting is to illustrate the benefits of ensembles in helping to improve the skill and reliability of TC track forecasts. Previous studies evaluating named TC track probability forecasts using multi-model ensembles have done this for only one or two Northern Hemisphere basins (e.g. Majumdar and Finocchio, 2010; Yamaguchi et al., 2012; and Leonardo and Colle, 2017), leaving a research gap to thoroughly analyse multi-model ensemble skill for all named storms worldwide. This led to the second key question: *What benefit is there in using ensemble forecast models in TC track*

forecasting, and does combining global ensembles into a multi-model (super) ensemble increase skill and value?, which inspired my second objective:

- **Objective 2: Evaluate the benefits of multi-model ensemble tropical cyclone track probability forecasts.**

This objective was tackled by the study contained in Chapter 4, entitled "*A global evaluation of multi-model ensemble tropical cyclone track probability forecasts*" and published in the Quarterly Journal of the Royal Meteorological Society in 2020.

1.2.3 *Key factors influencing severity of tropical cyclone flood hazard*

Primary attention in TC forecasting and evaluation has often been on the track, the intensity (the mean sea level pressure (mslp) minima and wind maxima) and the associated wind hazard (Emanuel, 2018). However, flooding due to heavy rain is a key contributor to the total number of deaths and impacts from TCs (Rappaport, 2000, 2014; Czajkowski and Kennedy, 2010; Czajkowski et al., 2017). Despite this, research into heavy rainfall and flooding associated with TCs has received more limited attention (Villarini et al., 2014b), so it is important to increase the focus on the rainfall from TCs, and the associated flood hazard.

Information on which TCs are likely to have storm surge and winds as their main hazard and which will also have substantial flood hazard, is of vital importance for disaster risk reduction. While the strongest winds and largest storm surge usually occur near the centre of intense TCs, rainfall and flooding often occur far from the centre, spread far inland, and last beyond when the cyclone has dissipated (Villarini et al., 2014a; Khouakhi et al., 2017), which has important consequences for evacuation, emergency management and recovery programs. Through case study or regional analysis there is broad understanding (Villarini et al., 2011; Saha et al., 2015; Hernández Ayala and Matyas, 2016; Yu et al., 2017; Jiang et al., 2018; Touma et al., 2019; Li and Liu, 2022; Racoma et al., 2022) of the factors that influence rainfall patterns in landfalling TCs, and how hydrological factors, such as soil conditions and orography, help to determine which TC cases had elevated impacts from flooding (Rappaport, 2000; Sturdevant-Rees et al., 2001), but systematic global analysis to objectively confirm these drivers and their relationships to downstream flood hazard is lacking. This initiated the third key question: *Can a global analysis of TC cases and their associated fluvial flood hazard reveal the key meteorological and hydrological factors that lead to an increased TC flood severity?,* which led to my third objective:

- **Objective 3: Identify the key factors that influence flood-related hazards from tropical cyclones using observations and reanalyses.**

This objective was addressed in the study contained in Chapter 5, entitled "*Key factors influencing the severity of fluvial flood hazard from tropical cyclones.*" and published in the Journal of Hydrometeorology in 2021.

1.2.4 *Analysing predictability links in the tropical cyclone flood forecast chain*

If planners and responders know the locations likely to be at greatest risk from flooding in TC cases, this can help with evacuation and emergency planning, response and recovery efforts. Hydrometeorological ensemble flood forecasts, where probabilistic river flow predictions are produced by forcing a hydrological model with ensemble precipitation forecast data (Cloke and Pappenberger, 2009; Cloke et al., 2013a), have become available at a global scale over the last decade (Emerton et al., 2016; Pappenberger et al., 2016). Therefore the final study of the thesis considers the potential to use ensemble flood forecasts to help provide early warning of river flooding in TC cases and explore their predictability. It looks in detail at the forecast chain using an ensemble-based sensitivity technique to see how the predictability and uncertainty of the overall flooding is impacted by the predictability of the track, intensity, and precipitation in TCs.

The final key question is: *How are forecasts for river flooding in TC cases affected by TC track, intensity, and precipitation forecasts, and can ensembles help to investigate this?* This motivated my fourth objective:

- **Objective 4: Analyse predictability links in the tropical cyclone flood forecast chain using an ensemble-based method.**

This final objective is addressed in Chapter 6, which analyses predictability links in the TC flood forecast chain, using an ensemble-based method illustrated through the case of Hurricane Iota. This work was accepted for publication in the Journal for Hydrometeorology in November 2023 with the following title: "*Using ensembles to analyse predictability links in the tropical cyclone flood forecast chain*".

1.3 Structure of the thesis

Table 1-1 summarises the research questions and the corresponding objectives as introduced in Section 1.1, along with the chapter title of where the associated research can be found. Each chapter has an allocated circular icon, which is contained in the header for that Chapter, to make it easier to navigate the thesis. The thesis is structured around five main chapters (Figure 1-2): a literature review and then the four main papers published during the PhD, which are re-formatted into the format of the thesis in order to preserve the literary unity of the thesis. Following this brief introduction (Chapter 1), Chapter 2 provides a literature review on ensemble forecasting, TCs (hazards, forecasting and verification), and river flood forecasting, to substantiate the motivation

for carrying out the research and introduce the main datasets and methods used in the thesis. Chapters 3 to 6 contain the four main research outputs from this PhD thesis, and address the four objectives stated in Section 1.1. Chapter 7 gives a discussion of the key knowledge contributions, limitations, and impacts of each stage of the research, and also includes a summary of additional papers and reports that I have been involved in during the period of the PhD that connect to this thesis. Recommended next steps based on the research findings are also outlined. Finally, Chapter 8 provides conclusions to the thesis, including a summary of the findings and the wider outcomes of this thesis. At the end of the thesis I provide a reference list and an appendix that contains Chapters 3, 4, 5 and 6 in published form, along with other important reports and papers.

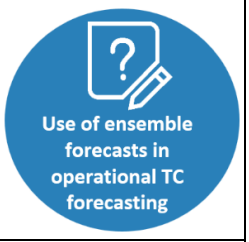
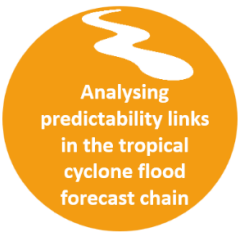
Chapter details			**Icon**
Chapter 2: Literature review; substantiating the motivation for carrying out the study and introducing the main datasets and methods used in the thesis.			
Research Question	**Objective**	**Chapter details**	**Icon**
How are ensemble forecasts currently used in operational tropical cyclone forecasting and warnings, and what potential is there for this to be improved?	Identify the current and potential use of ensemble forecasts by operational tropical cyclone forecast centres.	**Chapter 3**: Current and potential use of ensemble forecasts in operational TC forecasting: results from a global forecaster survey.	
What benefit is there in using ensemble forecast models in TC track forecasting, and does combining global ensembles into a multi-model (super) ensemble increase skill and value?	Evaluate the benefits of multi-model ensemble tropical cyclone track probability forecasts.	**Chapter 4**: A global evaluation of multi-model ensemble tropical cyclone track probability forecasts.	
Can a global analysis of TC cases and their associated fluvial flood hazard reveal the key meteorological and hydrological factors that lead to an increased TC flood severity?	Identify the key factors that influence flood-related hazards from tropical cyclones using observations and reanalyses.	**Chapter 5**: Key factors influencing the severity of fluvial flood hazard from tropical cyclones.	
How are forecasts for river flooding in TC cases affected by TC track, intensity, and precipitation forecasts, and can ensembles help to investigate this?	Analyse predictability links in the tropical cyclone flood forecast chain using an ensemble-based method.	**Chapter 6**: Using ensembles to analyse predictability links in the tropical cyclone flood forecast chain.	

Table 1-1. Summary of research questions and the associated objectives and chapters.

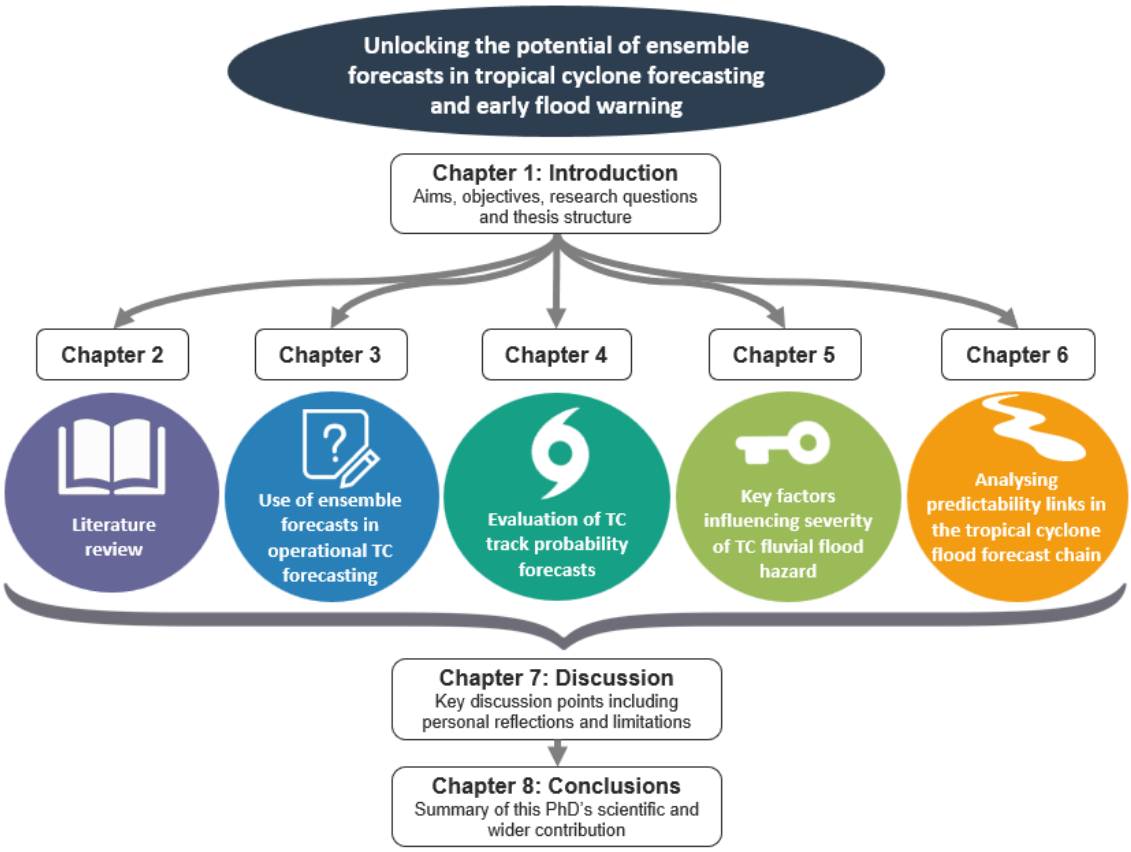


Figure 1-2. Schematic structure of the PhD.

Chapter 2 Literature review

2.1 Ensemble forecasting

2.1.1 *Motivation and history*

Traditional numerical weather prediction (NWP) computer models began in the 1950s and 1960s (Young and Grahame, 2023), taking a single best estimate of the current atmospheric state and using mathematical models to predict future weather. However, due to the chaotic nature of the atmosphere (Lorenz, 1963), even very small initial errors can grow very rapidly (Lorenz, 1969) and affect forecast quality from NWP models. In the 1970s and 1980s, scientists started investigating possible objective methods that could provide a level of forecast confidence (e.g. Palmer and Tibaldi, 1988), realising that there is no single deterministic solution, only an array of possible future atmospheric states. Ensembles began to be developed to provide a rational basis for assessing the range and likelihood of alternative scenarios. By 1992, the first operational ensemble forecasts were issued, allowing users to see a range of possible future scenarios rather than a single forecast realization (Buizza, 2019). These forecasts, by the European Centre for Medium-Range Weather Forecasts (ECMWF) (Molteni et al., 1996) and the National Centers for Environmental Prediction (NCEP) (Tracton and Kalnay, 1993), included multiple simulations starting from slightly differing initial conditions where dynamically defined perturbations had been added to the operational analysis, and allowed the creation of ensemble-based forecast products such as postage stamps, clusters, and probability fields.

2.1.2 *Current status*

Global and regional ensemble forecast systems are now run at most NWP centres across the world (e.g. Bowler et al., 2008; Houtekamer et al., 2009; Mamgain et al., 2020) and methods for perturbing the initial conditions and model physics continue to evolve (Leutbecher et al., 2017; Palmer, 2019; Zhou et al., 2022a). Innovative probabilistic forecast products have been developed using ensemble forecast data such as the Extreme Forecast Index (Lalaurette, 2003), Cyclone Database (Hewson and Titley, 2010), Decider regime analysis (Neal et al., 2016) and Global Hazard Map (Robbins and Titley, 2018), and a range of sector-specific applications developed (Taylor and Buizza, 2003; Storer et al., 2020; Mylne et al., 2021).

Many studies have demonstrated that probabilistic forecast skill and reliability can be improved through the use of multi-model (or 'super') ensembles, where multiple ensemble forecast models are combined together into a grand ensemble (e.g. Park et al., 2008; Hamill, 2012; Matsueda and Nakazawa, 2015). Ensemble forecasts can also be used to carry out ensemble-based sensitivity

analyses, to explore key factors controlling the nature and predictability of weather events (Hakim and Torn, 2008; Wu et al., 2013; Lynch and Schumacher, 2014; Torn et al., 2018; Zhang and Meng, 2018; Hu and Wu, 2020; Shen et al., 2020). Research has consistently concluded that if the probabilistic information that can be derived from ensemble forecasts is appropriately presented to people, they can understand it and use it to improve their decision making (Ripberger et al., 2022b).

Upgraded supercomputing facilities are allowing several NWP centres to plan to drop their 'high-resolution deterministic forecast' in favour of high-resolution ensemble forecasts. The first centre to do this is ECMWF in June 2023, when the horizontal resolution of the medium-range ensemble (ENS) increased to 9km (ECMWF, 2023). Therefore efforts to increase the pull-through of the benefits of ensemble forecasting into forecasts and warnings of extreme weather phenomena such as TCs have never been more important.

The following global ensemble forecast models are used in this thesis: Met Office Global and Regional Ensemble Prediction System (MOGREPS-G) (in Chapter 4), European Centre for Medium-Range Weather Forecasts Ensemble (ECMWF ENS) (in Chapters 4 and 6) and National Centers for Environmental Prediction Global Ensemble Forecast System (NCEP GEFS) (in Chapter 4). See Table 2.1 for further details.

Model	Number of members	Resolution	Forecast length
MOGREPS-G	36 (18+18 time-lagged)	20 km	8 days
ECMWF ENS	51	18 km	15 days
NCEP GEFS	31	25 km	35 days

Table 2-1. Summary of the ensemble forecast models used in this thesis (prior to ECMWF ENS upgrade in June 2023).

2.2 Tropical cyclones

2.2.1 *Climatology*

Tropical cyclones (TCs) are one of the most powerful and destructive weather systems on earth, bringing surface winds as high as 300km/hr, torrential rain and powerful storm surges. Defined by Ramsay (2017) as a "non-frontal synoptic scale low-pressure system over tropical and subtropical waters with organized convection and a closed cyclonic low-level wind circulation", they typically form in tropical oceans at latitudes between 5° and 30° north or south of the equator where converging winds near the ocean surface force air to rise to form a cluster of thunderstorms. If conditions are right (including sea surface temperature (SST) of at least 26.5°C, a steep vertical temperature gradient, high lower-tropospheric relative humidity and low wind shear), the convective area can grow in size, begin to rotate and become an intense self-sustaining TC (Gray, 1968; DeMaria et al., 2001). There are around 80 TCs each year globally; two thirds of these form

in the Northern Hemisphere from about June to November, and the remaining third form in the Southern Hemisphere normally from November to May.

2.2.2 *Rainfall*

The precipitation structure in mature TCs is composed of several distinct structural areas: the eyewall (a ring of heavy convective precipitation and high winds wrapped around the centre of circulation); the spiral rainbands outside of the eyewall, which contain a mixture of convective and stratiform rainfall; and primarily stratiform precipitating areas not within the eyewall or rainbands (Rogers et al., 2009). See Figure 2-1 taken from Houze (2010) for an illustration of an idealised but typical array of rainbands and eyewalls in a TC. The heaviest rain rates occur in the eyewall region, with peak rain rates occurring at radii less than 50km from the centre of the storm, and the mean rain rate decreasing from there with increasing radius (Lonfat et al., 2004). A study of three landfalling TCs in the U.S. found that: i) before landfall, the maximum rainfall tends to be concentrated between 50 km and 100 km from the centre of circulation; ii) after landfall, the maximum rainfall moves towards the storm centre as the eye collapses; iii) as the storm moves inland, the peak rainfall reduces and the rainfall distribution moves away from the centre with orographic amplification of rainfall playing a key role (Villarini et al., 2011). A recent study of landfalling TCs in China showed that TCs with a smaller inner core (measured by the radius of the maximum wind (RMW)) tend to have higher rain rates with higher axisymmetry than larger TCs, but that irrespective of size rainfall occurs within a 5° radius (Yu et al., 2022c). Most studies that have assigned a set radius for defining TC-related rainfall, have used a radius of either 500km (Prat and Nelson, 2013, 2016; Jiang et al., 2011; Luitel et al., 2018), 550km (Zhou and Matyas, 2017) or 5 degrees (Guo et al., 2017). A radius of 500km from the centre of the storm track is used to define TC rainfall in Chapters 5 and 6.

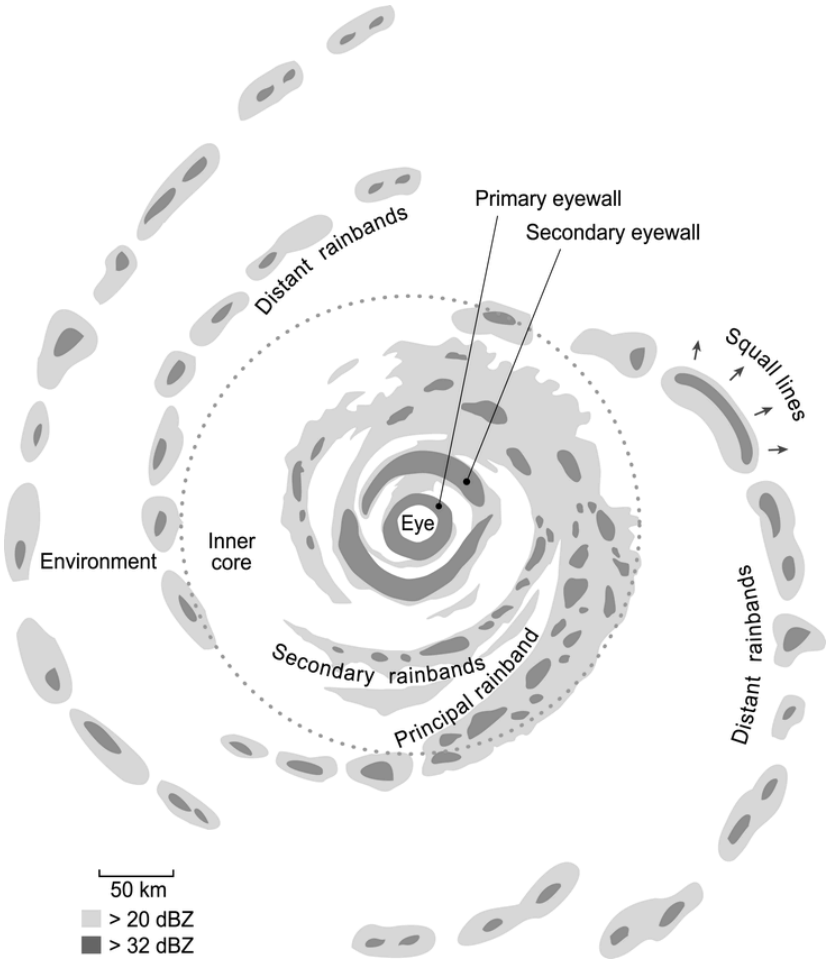


Figure 2-1. Schematic illustration of radar reflectivity in a Northern Hemisphere TC. Taken from Houze (2010).

The relationship between the intensity (maximum wind and mslp minima) of the TC and its associated rainfall is not straightforward. TC rain areas do not always correlate well with the spatial extent of strong winds, and are often larger, meaning that moderately heavy rainfall may commence before damaging winds arrive, decreasing the time available for preparedness activities (Guo and Matyas, 2016). The mean rain rate close to the centre of the storm tends to increase with increasing storm intensity but differences in rain rates among the different storm intensities tend to vanish at greater distances from the centre, e.g. by 300km from the centre (Lonfat et al., 2004). Racoma et al. (2022) did find a positive relationship between TC intensity and extreme precipitation, while Yu et al. (2015) found that although on average higher intensity TCs have larger peak axisymmetric rain, total rain, and rain areas, stronger TCs do not have systematically higher maximum rain rates than weaker storms.

The total rainfall produced by a TC is influenced by the length of time a TC spends over a given location, which is dependent on the size of the rainfall area and the translational speed of the storm (Rogers et al., 2009). A slow along-track motion, or stalling near or after landfall, can lead to higher amounts of rainfall (Racoma et al., 2022), while larger TCs also increase the rainfall hazard as they

precipitate upon one spot for a longer time and bring rainfall to a wider area. Strong upper-level atmospheric forcing (i.e. dynamic processes that force air to rise, such as upper level divergence) can lead to high rainfall amounts even from systems moving at an average forward motion. TC rain fields also contain asymmetries caused by environmental effects such as topography, storm motion, vertical wind shear, and interaction with midlatitude weather systems such as upper-level troughs. There are many orographic effects on TC rainfall, including modulation of the microphysics and dominant rain processes in addition to the direct orographic enhancement of the rainfall due to the lifting effect of the terrain (DeHart and Houze, 2017; Qiu et al., 2019). Storm motion generally leads to an area of low-level convergence and greater rainfall in the front-right quadrant of the storm (Zhou et al., 2018). In weak vertical wind shear, the TC eyewall and rainband precipitation are more symmetric around the centre, but they become increasingly concentrated downshear and left-of-shear (for Northern Hemisphere storms) as the vertical wind shear increases (Yu et al., 2015; Bell, 2019), leading to an exposed circulation centre and a reduction of the potential magnitude of rainfall. Wider environmental interactions can modify the TC rainfall pattern substantially and generate remote rainfall events, which is one of the challenges in forecasting TC rainfall (Bao et al., 2015; Deng et al., 2017; Cheung et al., 2018).

Climatologically, TC rainfall contributes significantly to the total and extreme precipitation events in all global regions where TC landfalls occur (Prat and Nelson, 2016; Villarini and Denniston, 2016; Bagtasa, 2017; Khouakhi et al., 2017). The changing climate is likely to lead to an increase in the hazards and impacts from TC rainfall, with increasing TC rain rates already being detected (Guzman and Jiang, 2021), and projections that rainfall over land may be particularly increased due to stronger updrafts (Scoccimarro et al., 2017). The latest comprehensive assessment of the projected responses of TCs to climate change (Knutson et al., 2020) found medium-to-high confidence in an increase in TC rainfall rates, with a median projected increase of 14% for a 2°C global warming. Other factors important to TC rainfall such as translation speed may also be impacted, although confidence in projections of TC translation speed is low due to a lack of consensus in climate projections and in recent observed studies (Kossin, 2018; Moon et al., 2019; Yamaguchi et al., 2020).

2.2.3 *Fluvial flood hazard and impact*

The hazards associated with a TC include strong winds, tornadoes, heavy rain, storm surge, waves, flooding (coastal inundation, pluvial and fluvial), and landslides. These often occur simultaneously in space and time, leading to difficulties for emergency preparedness, response and recovery programmes, and a need to consider hazards and impacts in a multi-hazard context. Yet the focus on TC prediction has traditionally been on the track and intensity, and the associated wind hazard

close to the landfall location, rather than on the multiple and cascading hazards such as fluvial flooding.

Flooding is defined by the Intergovernmental Panel on Climate Change (IPCC) as “The overflowing of the normal confines of a stream or other body of water, or the accumulation of water over areas that are not normally submerged” (IPCC, 2012). Fluvial floods occur when a watercourses' capacity is exceeded resulting in water overflowing on to surrounding land, and are affected by the intensity, duration, amount, and timing of precipitation. They are also affected by catchment conditions such as water levels in the rivers, soil character and status (soil moisture content and vertical distribution), urbanisation, and the existence of dikes, dams, and reservoirs (Bates et al., 2008).

Flooding due to heavy rain in TCs is a significant and dangerous hazard. In a large study of TC related fatalities in the U.S., Rappaport (2014) found that rainfall-induced flood deaths occurred in more tropical cyclones than deaths from any other TC hazard. An earlier study (Rappaport, 2000) found that freshwater floods caused more than half of the 600 U.S. deaths directly associated with TCs or their remnants in the 30-year period 1970-99, with more than three-quarters of the victims under age 13 dying in rain-induced floods and most fatalities occurring in inland counties. River flooding from landfalling TCs is also responsible for severe socioeconomic losses around the globe (Yang et al., 2020). TCs account for a significant percentage of the total number of flood-related fatalities in countries most often directly impacted by landfalling events (Hu et al., 2018),

TC-related flooding often occurs far from the centre of storm, spreads far inland, and lasts beyond when the cyclone has dissipated (Villarini et al., 2014a; Khouakhi et al., 2017), with insurance data indicating twice as many residential losses from freshwater flooding compared to storm surge losses (Czajkowski et al., 2017). Excluding Hurricane Katrina, 80% of hurricane fatalities between 1970 and 2007 occurred outside of landfall counties (Czajkowski and Kennedy, 2010). However, despite the role of inland flooding in TC-related deaths, evacuation decision-making does not often include a consideration of the risk of inland flooding (Stein et al., 2010). Residents often misconstrue the greatest threat as coming from wind rather than water (Meyer et al., 2014), for example using the intensity of the storm on the Saffir Simpson scale was a key determining factor in the evacuation decision-making during hurricanes (Whitehead et al., 2000). The potential for heavy rainfall is present regardless of the storm’s category (NHC, 2023b; NWS, 2023), as TCs that are considered “weak” by traditional intensity measures can still produce extremely heavy rainfall and flooding impacts (Tang et al., 2021). This shows the importance of increasing public awareness of the dangers of inland flooding, and of a better understanding of those factors that influence the severity of flood-related hazards.

While there is broad understanding, largely through case study or regional analysis (Saha et al., 2015; Hernández Ayala and Matyas, 2016; Yu et al., 2017; Jiang et al., 2018; Touma et al., 2019; Li and Liu, 2022; Racoma et al., 2022, 2023) of those factors that influence rainfall patterns in landfalling TCs, why seemingly similar landfalls and tracks have different rainfall distributions (Cheung et al., 2018), and how hydrological factors, such as soil conditions and orography, help to determine which TC cases had elevated impacts from flooding (Rappaport, 2000; Sturdevant-Rees et al., 2001), systematic global analysis to objectively confirm these drivers and their relationships to downstream flood hazard is lacking. This is addressed in Chapter 5, which undertakes a global analysis of the key meteorological and hydrological factors that lead to an increased fluvial flood hazard from TCs. For example, the importance of translation speed, highlighted in a study of cyclone-induced rainfall in Australia in Saha et al. (2015), is confirmed in the systematic global analysis of TC flooding in Chapter 5. The idea of assessing the extent to which multiple key TC characteristics combine to affect the likelihood of extreme precipitation in northern Philippines (Racoma, 2022), is also taken further in chapter 5, to look at how key factors influencing TC rainfall combine to modulate flood severity across a large global sample of TC cases.

2.3 Forecasting of tropical cyclones

2.3.1 *Operational forecasting and global co-operation*

For more than 50 years, the World Meteorological Organization (WMO) has led global collaboration efforts with an aim of improving TC forecasts and warnings. In 1972, the WMO Tropical Cyclone Project (now Programme) (TCP) was created, which established six Regional Specialized Meteorological Centers (RSMCs) and six Tropical Cyclone Warning Centers (TCWCs) responsible for providing real-time warnings to the public (Ramsay, 2017) (see Figure 2-2). It is these centres that are surveyed in Chapter 3. The International Best Track Archive for Climate Stewardship (IBTrACS) (Knapp et al., 2010) collects from these international centres their best estimates of TC location and intensity to create a merged global dataset for public use. This data is used in Chapters 4, 5, and 6.

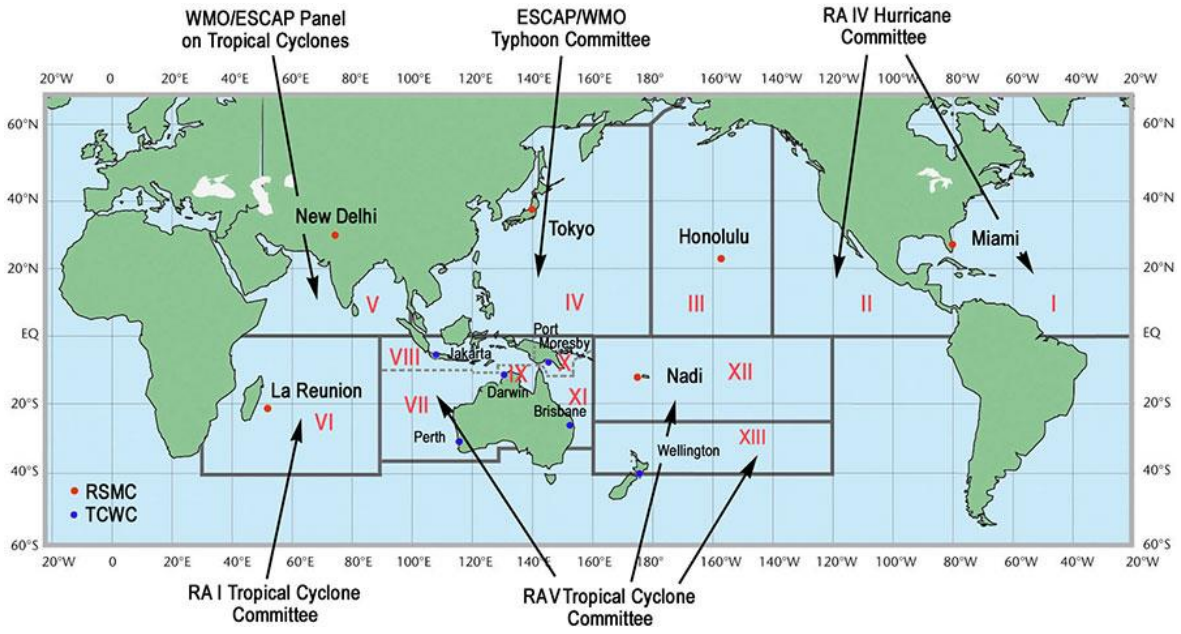


Figure 2-2. Map showing the six Regional Specialized Meteorological Centers (RSMCs) and the six Tropical Cyclone Warning Centers (TCWCs) assigned by the World Meteorological Tropical Cyclone Programme. Taken from Ramsay (2017).

2.3.2 *Forecasting track and intensity*

TC forecasts have been issued by RSMCs for over 50 years, focussed on forecasting the track, landfall location and intensity. Over that time there have been remarkable improvements in the forecasting of the TC track (Figure 2-3a), with track errors reducing by around two thirds in just 25 years (Landsea and Cangialosi, 2018; Yu et al., 2022a). A variety of TC tracking methods are used, including statistical forecasting techniques, dynamical modelling, statistical-dynamical techniques, and hybrid techniques (Roy and Kovordányi, 2012), with a traditional emphasis on multimodel deterministic consensus forecast guidance at many operational forecasting centres (Goerss, 2000; Simon et al., 2018). Track improvements over recent years are primarily due to advances in observational technology, data assimilation, and numerical weather prediction models (Yamaguchi et al., 2017; Heming et al., 2019).

Forecasting the intensity of TCs, and in particular forecasting rapid intensification, has traditionally been more challenging, although Figure 2-3b shows that there have been significant improvements in intensity forecasts from RSMCs in recent years (Cangialosi et al., 2020), while work to improve TC intensity forecasts from NWP models continues, including higher resolution ensemble forecasts (Magnusson et al., 2019).

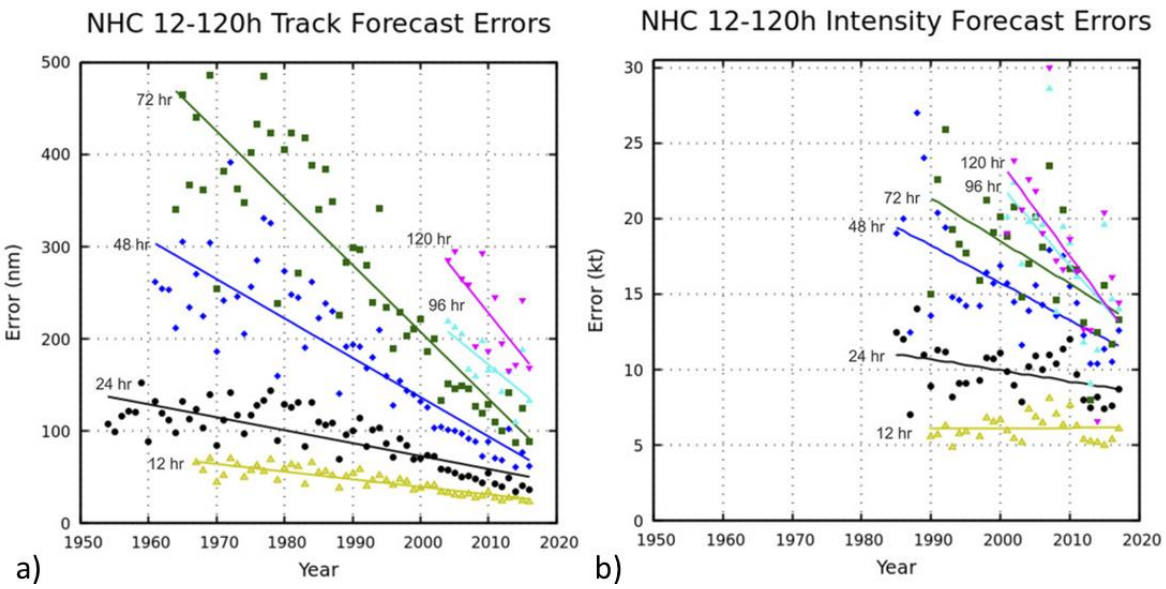


Figure 2-3. National Oceanic and Atmospheric Administration (NOAA) National Hurricane Center (NHC) Atlantic forecast errors: (a) track errors in nautical miles, from 1950 to 2017 and (b) intensity errors in knots from 1990 to 2017. Forecast lead time colours: 12 hr: yellow; 24 hr: black; 48 hr: blue; 72 hr: green; 96 hr: cyan; 120 hr: magenta; linear regression lines are also shown. Taken from Emanuel (2018).

Despite the improvements in TC forecast skill, and predictions that track skill could improve further in the future (Zhou and Toth, 2020; Yu et al., 2022a), it is widely accepted that making perfect forecasts will never be possible. This is because, as first demonstrated by Lorenz (1969), there exists a limit of predictability beyond which further increases of skill are unattainable due to the cascading of errors up to the synoptic scale. Therefore a shift in focus of the research and forecasting community towards providing probabilistic forecast information about TC track uncertainty, and pulling this dynamic forecast information into probabilistic TC hazard and impact forecasts, is recommended to maximise the usefulness of forecasts (Landsea and Cangialosi, 2018).

To date, most metrics of TC forecast skill have focused on the track and intensity described in this section. Yet, in practice, these are far removed from the metrics of most utility to society, which are more concerned with the probabilities of hazardous conditions at fixed points in space (Emanuel, 2018). With ensemble forecasts now more widely available, probabilistic TC forecasts have great potential, both as effective ways of communicating risk in advance of TC landfall and as better metrics by which model improvements may be judged.

2.3.3 *Forecasting precipitation and flooding*

Lamers et al. (2022) prepared a summary of recent research studies and the forecasting challenges of TC rainfall for the Tenth International Workshop on Tropical Cyclones (IWTC-10) in Bali in December 2022. They state that forecasts of TC rainfall are generally influenced by eight main

factors: movement, storm size, topography, moisture, instability, diurnal cycle, vertical wind shear and wider interactions. Considering the time required for local governments to issue an emergency alert or organize extensive evacuations, rainfall or flooding forecasts have to be provided to decision makers at least 2-3 days in advance, which given the complexity of the factors influencing TC rainfall is still a challenging task (Cheung et al., 2018).

There are many techniques for quantitative precipitation forecasting (QPF) of TC rainfall including statistical techniques (Pfost, 2000), satellite-based forecasts (Kidder et al., 2005; Ebert et al., 2011), climatology-based techniques (Marks et al., 2002; Lonfat et al., 2007), and analog-based forecast methods (Kim et al., 2020b, 2020a; Bagtasa, 2021; Jia et al., 2022), with machine learning techniques showing promise in enhancing these analog and statistical methods (Liu et al., 2021; Wang et al., 2022).

More complex NWP models are widely cited by operational TC forecast centres as a critical source of guidance for TC rainfall prediction, in particular global models and their ensembles that can provide coverage across a whole basin (Lamers, 2022). Global NWP models have been found provide skilful forecasts of TC rainfall with lead-times up to 48 h, without a consistently best model (Luitel et al., 2018). However, global models can suffer from insufficient horizontal and vertical resolution to resolve orographic effects and physical processes and higher-resolution regional models such as HRRR (High Resolution Rapid Refresh) and HWRF (Hurricane Weather Research and Forecasting) in the U.S. (Ko et al., 2020; Dowell et al., 2022), and AROME in the south west Indian ocean (Faure et al., 2020) can help to provide more detailed forecasts at short lead times.

Not all RSMCs are responsible for providing precipitation forecasts and warnings from TCs. For example, for TCs in the U.S. another NWS national centre, the Weather Prediction Center (WPC), serves as the specialized centre of precipitation expertise and creates QPF based on the NHC official track forecasts. It also issues rainfall and flooding forecast products based on these such as the Excessive Rainfall Outlook (ERO), which graphically maps the probability of rainfall exceeding flash flood guidance thresholds, exhibiting good reliability (Erickson et al., 2021).

Lamers (2022) provides an international summary of the current state of warnings for rainfall and associated flooding. Many countries have adopted a tiered warning strategy for rainfall, represented by different colours and/or names. For example, India (IMD) provides impact-based heavy rainfall warnings along with designated colour-code based actionable warnings for all major cities, and is also implementing an integrated flood warning system for major flood prone cities such as Chennai. In the U.S., the forecast process at WPC has included exchange of hydrometeorological insights with the newly created National Water Center (NWC), and these

centres have engaged in event-driven joint product creation (e.g. tropical cyclone key messages for flooding) (Burke et al., 2023). In Japan, a “real-time risk map” for flooding was launched via smart phone in 2019, with five coloured levels to indicate flood risk at each location.

Approaches for hydrological modelling also vary from country to country, and are mostly dependent on individual national meteorological and hydrological services, rather than being carried out at the regional TC forecast centre, with a trend toward trying to account for more of the complexity in modelling approaches (Lamers et al., 2022). In the U.S. the NWS River Forecast Centers (RFC) use the QPF operationally to forecast river levels at key points along U.S. (Adams, 2016), while the new national centre, the NWC, is currently building operational capacity and will be the centre of expertise for hydrology (Lamers et al., 2022). There are similar arrangements at other RSMCs and national meteorological services. Flooding in particular can sometimes be handled in a totally different department. For example, in Mozambique, the official TC forecasts are received from RSMC La Reunion, but do not provide information on rainfall or flooding. The Mozambique National Institute of Meteorology (INAM) then issue TC warnings and rainfall forecasts based on a variety of rainfall forecast products. Flood warnings are the responsibility of the National Directorate of Water Resources Management (DNGRH), who issue them based on observed trends in river levels and a qualitative assessment of the TC and rainfall forecasts (Emerton et al., 2020).

2.3.4 *Probabilistic forecasting of tropical cyclones*

Dynamic ensemble forecast models have a vital role to play in TC forecasting, through their ability to highlight the situation dependent uncertainty and provide probabilistic forecast information to help inform decision makers. Objectively identifying the forecast track of each ensemble member is essential both for post-event model evaluation and for the generation of forecast guidance products in real time. Various tracking techniques are used by operational centres around the world (Vitart and Stockdale, 2001; Van der Grijn, 2002; Tallapragada et al., 2013) and in research (Hodges, 1995). At the Met Office, a bivariate (850 relative vorticity and mean sea-level pressure) TC tracker known as MOTCTracker (Heming, 2017) is run in real time on the Met Office MOGREPS-G ensemble, and also on the ECMWF ENS and NCEP GEFS ensembles. The resulting track information from the three global ensembles are also combined into a 108-member multi-model ensemble, often referred to as a “super ensemble”. A range of products, including track probability and intensity forecasts for both named and forming storms, are produced and used by forecasters in the Met Office Global Guidance Unit. They are also distributed to several operational TC forecasting centres, including the Regional Specialised Met Centres (RSMCs) in Miami, La Reunion, and Delhi. It is these track probability forecasts that are verified in Chapter 4.

Many global modelling centres produce TC track forecasts from their ensemble forecast models and share the forecast track data via the TIGGE (The International Grand Global Ensemble) cyclone exchange programme (Swinbank et al., 2016). This TIGGE data is used to source the ECMWF ENS TC track forecasts used in Chapter 6.

Previous studies evaluating named storm track probability forecasts using multi-model ensembles have done this for only one or two Northern Hemisphere basins (e.g. Majumdar and Finocchio, 2010; Yamaguchi et al., 2012; and Leonardo and Colle, 2017). Therefore there is a research gap to thoroughly analyse multi-model ensemble skill for all named storms worldwide, to find if there is benefit in terms of probabilistic forecast skill, reliability and value, from combining the ensembles into a multi-model (super) ensemble. To further establish the benefit of ensemble track probability forecasts over traditional deterministic techniques, a further research gap is to evaluate whether using the full probability forecast information via the multi-model ensemble adds skill compared to a consensus forecast of the parent deterministic models. These research gaps are addressed in Chapter 4, with the global study confirming that the results of the earlier basin-specific studies, such as the under dispersion of the tracks in the MOGREPS-G and NCEP GEFS ensemble found in Leonardo and Colle (2017) and the additional skill found from combining the individual ensembles into a multi-model ensemble (Yamaguchi et al., 2012), hold true when examined across all global TC basins.

Ensemble forecasting of TC hazards has received less attention, although this is changing in recent years. For example, in Asia, ensemble forecasting systems show promising forecast skill, including the ensemble typhoon quantitative precipitation forecast (ETQPF) used in Taiwan (Hong et al., 2015) and the Global Ensemble Prediction System (GEPS) precipitation forecasts for Japan (Cheung et al., 2018). In the Southwest Indian Ocean basin, the Système de Prévision des Inondations en contexte Cyclonique (SPICy) project aimed to explore new probabilistic forecast products for TC hazards such as coastal inundations using both ECMWF ENS forecasts (Bonnardot et al., 2019) and high-resolution regional ensemble AROME-IO EPS (Bousquet et al., 2020). In the U.S., a regional HWRF-based ensemble was found to have good forecast skill and reliability for probabilistic TC wind and precipitation forecasts (Bachmann and Torn, 2021). An ECMWF ENS-based combined meteorological and land-surface index, the flood hazard risk forecasting index (FHRFI) was found to provide skilful forecasts up to four days ahead for the flash flooding in Hurricane Harvey (Jurlina et al., 2020). For river flooding in TC cases a handful of studies have successfully illustrated the potential of forcing hydrological models with ensemble forecast data: A WRF-based ensemble was used to drive a hydrological model for the Lanyang basin during Typhoon Nanmadol (Hsiao et al., 2013); NCEP GEFS data was used to forecast inundation extents for Hurricanes Irene and Sandy

(Saleh et al., 2017); and an ensemble based on the JMA Nonhydrostatic Model (NHM) was used to force a hydrological model for the Shingu river basin in Japan (Yu et al., 2016).

2.3.5 *Forecasting for humanitarian preparedness and response*

Humanitarian and development agencies need to prepare for and respond to events such as TCs in order to prevent human suffering and losses. Over the last few years progress has been made in developing systems that provide funding based on a forecast of an extreme event. Initiatives such as Forecast-based Financing (FbF) and the UK's Science for Humanitarian Emergencies and Resilience (SHEAR) research programme have helped to ensure that forecasts can be used to take early action ahead of a disaster (Coughlan de Perez et al., 2016). For TC cases many humanitarian agencies rely on forecasts from the Global Disaster Alert and Coordination System (GDACS), which classifies TCs based on their winds, but do not currently include rainfall forecasts or any uncertainty information (GDACS, 2023). The World Food Programme (WFP) Advanced Disaster Analysis & Mapping (ADAM) system includes forecasts for TCs and includes precipitation forecasts, but only using deterministic forecast data.

Part of the motivation behind this PhD project was to look at the potential for ensemble flood forecasts from the Global Flood Awareness System (GloFAS, see Section 2.5.3) to be used in TC cases. GloFAS flood forecasts had previously been used for triggering humanitarian action based in Uganda (Coughlan de Perez et al., 2016) but their potential for TC cases was as yet untapped. During the timeline of this PhD project, GloFAS forecasts have started to be used to produce forecast bulletins for humanitarian agencies in TC cases, with the knowledge gained from this PhD able to contribute to this work (see Emerton et al., 2020 Speight et al., 2022, and summarised in Chapter 7). GloFAS forecasts are also now included in Pacific Disaster Centre Disaster AWARE platform and WFP ADAM flooding dashboards, so the relevance of the work of this thesis in aiding the interpretation of flood forecasts in TC cases continues to grow.

2.4 Verification methods for tropical cyclone forecasts

2.4.1 *Traditional verification metrics*

In order to evaluate how good our forecasts are scientists often use statistical metrics to verify their accuracy, skill and other characteristics. The most common verification metric used to analyse a TC forecast from an NWP model is to verify the TC forecast track error, defined as the great-circle difference (the shortest distance between two points on the surface of a sphere, measured along the surface of the sphere) between a TC's forecast centre position and the best-track position at the verification time. This is a vector quantity, which is sometimes decomposed into components of along- and cross-track error, with respect to the observed best track (JWGFVR, 2013). TC intensity

is often represented by either the maximum surface wind speed averaged over a particular time interval or by a minimum surface pressure estimate. The mean value of the raw intensity errors provides an estimate of the bias in the forecast intensity values, whereas the mean of the absolute errors indicates the average magnitude of the error. The size of a TC is typically given by "wind-speed radii", expressed as the distance from the centre of the TC to the maximum extent of winds exceeding 34 and 64 knots (tropical storm and hurricane strength) but the uncertainty associated with the analysis of these values is quite large due to the limited density of available wind observations, making verification challenging. TC track and intensity errors are calculated in Chapter 6.

When verifying rainfall, the accuracy of predicted rainfall accumulations over fixed periods and for the duration of the storm are of interest. Observations of TC rainfall are important for analysing the rainfall patterns in TCs and for verifying their forecasts, and are generally gauge-based, satellite-based or from reanalyses (Sun et al., 2018). Rain gauges are indispensable for measuring precipitation directly, but global gauge density is limited and their distribution is uneven (Kidd et al., 2017). Satellite data provides a more uniform global coverage, and the international Global Precipitation Measurement (GPM) mission uses a constellation of satellites to improve sampling, using its Integrated Multi-satellitE Retrievals for GPM (IMERG) algorithm to combine microwave and IR-based observations to increase accuracy and reliability (Hou et al., 2014; Huffman et al., 2019). IMERG offers global rainfall estimates with a spatiotemporal resolution of 0.1°×0.1° every 30 minutes, meaning it is able to better represent TC rainfall compared to its predecessor, the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) (Chen and Huang, 2019; Yuan et al., 2021). IMERG has been found to capture the detailed spatial structure of rainfall associated with TCs, especially within the first 100 km from the TC centre outwards (Rios Gaona et al., 2018), and for severe flooding cases such as Hurricane Harvey (Omranian et al., 2018). GPM IMERG data is used in Chapters 5 and 6. Reanalyses aim to optimally combine observations and models; ERA5 is ECMWF's latest comprehensive atmospheric reanalysis with a global horizontal grid resolution of about 31km/0.28125° (Hersbach et al., 2020) and is used in Chapter 5. A simple gridpoint-based error measure (Mean Absolute Error) was calculated against IMERG data in Chapter 6 in order to provide a basic precipitation score to compare against track, intensity, and discharge scores.

2.4.2 *Ensemble-based verification metrics*

Verification methods applied to ensemble forecasts have two main objectives: 1) to assess the characteristics of the ensemble distribution; and 2) to verify the probability forecasts. Several metrics are required to fully evaluate the various attributes of ensemble forecast quality such as

skill, reliability, sharpness and value. Detailed reviews of ensemble forecast verification can be found in Wilks (2011) and Jolliffe and Stephenson (2012). Commonly used verification metrics include rank histograms, reliability diagrams, spread vs. skill plots, Relative Operating Characteristic (ROC) plots, Brier Skill Score (BSS) and the Continuous Ranked Probability Score (CRPS). Relative economic value plots are another useful verification tool, displaying the relative improvement in economic value between the sample climatology and a perfect forecast for a range of cost-loss ratios (Richardson, 2000). The track probability verification carried out in Chapter 4 includes ROC diagrams, Relative Economic Value plots, reliability diagrams and Brier Skill Score calculations.

2.4.3 *Feature-based verification metrics*

Mislocated features, which can be a problem in TC rainfall forecasts due to track errors, give rise to the "double penalty", where the forecast is penalized for predicting something where it did not occur, and again for failing to predict it where it did occur. Spatial verification approaches such as the Fractions Skill Score (FSS) were developed to address this issue by explicitly considering the spatial coherence in the forecast and observed fields (Roberts and Lean, 2008; Mittermaier et al., 2013). Feature-based verification metrics take this idea further, aiming to provide objective information that mimics human judgment when comparing mapped forecasts and observations (Gilleland et al., 2009, 2010), providing information on how well forecasts have captured the rainfall area, location, and intensity of threshold-based objects identified in the observed and forecast fields. Examples of object-based verification methods include the Contiguous Rain Area method (CRA) (Ebert and McBride, 2000; Ebert and Gallus, 2009), the Method for Object-based Diagnostic Evaluation (MODE) (Davis et al., 2006, 2009), and the Structure Amplitude Location method (SAL) (Wernli et al., 2008, 2009). Some object-based methods have been further adapted and applied to ensemble verification (Gallus, 2010; Radanovics et al., 2018; Ji et al., 2020; Dube et al., 2022),

Only a handful of studies have applied feature-based verification to TC rainfall forecasts (Marchok et al., 2007; Chen et al., 2018; Yu et al., 2020), and only to deterministic forecasts. To my knowledge there has not been any attempt to apply object-based methods to ensemble TC rainfall forecasts. An initial employment of SAL, MODE and CRA methods to verify ensemble TC rainfall forecasts is introduced and discussed in Chapter 7.

2.5 River flood forecasting

2.5.1 *Background*

Flood forecasting and warning systems have been shown to reduce impacts and save lives and these systems are an increasingly important tool for water managers and emergency preparedness and response activities (Adams and Pagano, 2016). Hydrological forecasting techniques include simple

statistical and water-balance approaches, rainfall-runoff modelling, and hydrological and hydraulic flow routing (Sene, 2010). Some approaches are purely based on statistical linkages between upstream and downstream gauges (Franchini et al., 2011), which can be effective on longer rivers or where only short lead times are needed. However, most hydrological forecasts aim to translate meteorological observations and forecasts into estimates of river flows, using rainfall-runoff models.

The rainfall-runoff models used for flood risk assessment use rainfall inputs to provide runoff outputs, are generally built around simplifying assumptions of natural hydrological system and have evolved greatly in the last few decades (Beven, 2012). These models can be divided into statistical, conceptual or physically based models. Data-driven statistical models lack a description of the underlying hydrological processes and are usually expressed as empirical models (Todini, 2007). Conceptual models attempt to describe the underlying physical processes using more complex empirical equations. Physically based models are more strongly grounded in the physics of the underlying hydrological processes and have equations that are based on the laws of conservation of mass and energy. Routing techniques used in hydrological models are broadly classified as either hydrological or hydraulic routing. In hydrological routing, the time-varying flow is computed at a single specified location downstream, while in hydraulic flood routing the hydrograph is simultaneously computed at several cross-sections along the watercourse, requiring detailed channel bathymetry and roughness data (Jain et al., 2018). Hydrological models can be further categorised by scale. Lumped models are the simplest and assume homogeneity of hydrological response across the entire catchment. More complex distributed models consider the spatial variability of the catchment physical properties, while semi-distributed models seek to find a middle ground benefitting from the strengths of both lumped and distributed models (Okiria et al., 2022). Best practice in flood forecast and warning systems has increasingly moved towards the use of physically based distributed models (WMO, 2011).

In recent decades better scientific understanding of physical processes and increased computing power has allowed the development of increasingly complex and realistic hydrological models, while satellite data has improved calibration, and technological advances have helped with warning distribution. However, there are still many challenges to overcome, including ongoing difficulties in the availability of reliable, spatially distributed, river discharge observations from river gauges, particularly in more vulnerable countries (Lavers et al., 2019).

It is also crucial to be able to estimate and communicate to flood forecast users the uncertainty in the river discharge and flood forecasts. In hydrological forecasting, the key sources of uncertainty are the input data, model structure and model parameters (Zappa et al., 2011), with the largest

uncertainty usually coming from the meteorological forecast input (Moore, 2002; He et al., 2009; Emerton et al., 2016). Therefore there is an increasing focus on ensemble flood forecasting (Cloke and Pappenberger, 2009), where ensemble weather forecast models provide a representation of that meteorological forecast uncertainty.

2.5.2 *Ensemble flood forecasting*

In ensemble flood forecasts, probabilistic river flow predictions are produced by forcing a hydrological model with ensemble precipitation forecast data (Clark and Hay, 2004; Cloke and Pappenberger, 2009; Cloke et al., 2013b). Ensemble flood forecasting has gained significant momentum over the past decade due to the growth of ensemble weather forecast models, an increase in the availability of high-performance computing, and a gradual shift from deterministic to risk-based decision-making to account for forecast uncertainty (Wu et al., 2020). International community-building activities, such as the Hydrologic Ensemble Prediction Experiment (HEPEX), which seeks to advance the science and practice of hydrological ensemble prediction and its use in impact- and risk-based decision making, have also helped to build this momentum (Schaake et al., 2007).

A review of ensemble flood forecasting research from 2001 to 2019 found 61 papers focussing on various aspects of ensemble flood forecasting around the globe (Wu et al., 2020). Although there was an increasing trend in the number of papers per year through the period, the vast majority of studies are conducted in countries in the northern hemisphere or in the developed world. There was also an increasing trend in studies applying ensemble flood forecasting to solve a real-world problem, although these remain fewer than those studies focussing on modelling techniques and evaluation. The papers reviewed did not include studies on the use of ensemble forecasting in emerging application areas, such as evacuation planning in emergency management and risk mitigation for humanitarian actions. The work of Chapter 6, looking at how global ensemble flood forecasts can be used to help understand the cascading predictability of flooding in TC cases, whilst not directly looking at these areas, can hopefully provide useful insight for these communities to help advance early flood warnings in TC cases.

2.5.3 *GloFAS*

The forecasting of floods at the global scale is crucial to providing early awareness of potential impactful flood events in those regions where local flood models and warning services may not exist (Emerton et al., 2016). The Global Flood Awareness System (GloFAS) is the global flood service of the European Commission's Copernicus Emergency Management Service (CEMS), and provides an operational system for monitoring and forecasting floods across the world (Harrigan et al., 2023)

with over 5000 registered users (Harrigan et al., 2020). However, prior to the start of this PhD research, the potential of GloFAS to help to provide early warning of river flooding in TC cases was largely unexplored.

GloFAS is one of only two global scale ensemble operational flood forecasting systems (Emerton et al., 2016); The other is the Global Flood Forecasting and Information System (GLOFFIS) run by Deltares (den Toom et al. 2020). GloFAS was jointly developed by the Joint Research Centre (JRC) of the European Commission and the European Centre for Medium-Range Weather Forecasts (ECMWF), and went pre-operational in 2011 (Alfieri et al., 2013), becoming a fully operational, supported service in 2018. It couples global EC ENS input data with hydrological modelling to produce river discharge on a global river network, currently at 0.1 degree resolution. In this thesis, the version of GloFAS used is v2.1, implemented in November 2019, so this will be described in more detail. GloFAS v2.1 couples the ECLand surface and subsurface runoff outputs to Lisflood-routing to produce river discharge. ECLand is the land surface modelling platform (Boussetta et al., 2021) that processes ECMWF ENS meteorological forecasts through the land surface module HTESSEL (The Hydrology-Tiled ECMWF Scheme for Surface Exchange over Land) (Balsamo et al., 2011), to create surface and sub-surface runoff fields. The Lisflood-routing model is an open source distributed semi-physically based model developed at the Joint Research Centre (JRC) of the European Commission (Van Der Knijff et al., 2010), with a calibrated hydrological routing scheme (Hirpa et al., 2018). GloFAS v2.1 included a revised hydrological reanalysis based on the officially released ERA5 meteorological forcing from 1979 to 2019 (Harrigan et al., 2020). ERA5 is the most recent atmospheric reanalysis product of ECMWF, and covers the 1950-present period with a horizontal resolution of ~31 km (Hersbach et al., 2020).

A first systematic evaluation of GloFAS hydrological forecast skill globally found that GloFAS is skilful in over 93% of catchments in the short (1 to 3 days) and medium range (5 to 15 days) against a persistence benchmark forecast and skilful in over 80% of catchments out to the extended range (16 to 30 d) against a climatological benchmark forecast (Harrigan et al., 2023). The GloFAS ensemble river discharge reanalysis and real-time forecast datasets are freely and openly available through the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) (Zsótér et al., 2019; Harrigan et al., 2021). The GloFAS v2.1 reanalysis is used in Chapters 5 and 6, with the GloFAS v2.1 forecasts forming a key part of Chapter 6.

GloFAS can provide early warnings of upcoming flood events for decision makers ranging from water authorities and hydropower companies to civil protection and international humanitarian aid organizations, available via an interactive map viewer (https://www.globalfloods.eu/glofas-

forecasting/). For example, the flood summary map displayed in Figure 2-4 combines the 2- (yellow), 5- (red) and 20-year (purple) return period exceedance probabilities into categorical information: the return period thresholds are calculated by fitting a Gumbel extreme value distribution to the GloFAS-ERA5 river discharge reanalysis over the 1979-2018 period (Zsótér et al., 2020). Reporting points are chosen at fixed (selected from 4000 gauged river stations included in the Global Runoff Data Centre (GRDC, http://grdc.bafg.de/)) and dynamic (chosen to provide similar information in river reaches where no gauge data is available) locations in the river network where upcoming flood hazard is detected. These points are assigned an alert level colour depending on the mean discharge forecast peak compared to the return period thresholds, and have an upward-pointing triangle if the discharge forecast is rising, and a downward-pointing triangle if it is falling. When you click on a point the discharge hydrograph for that river point is displayed.

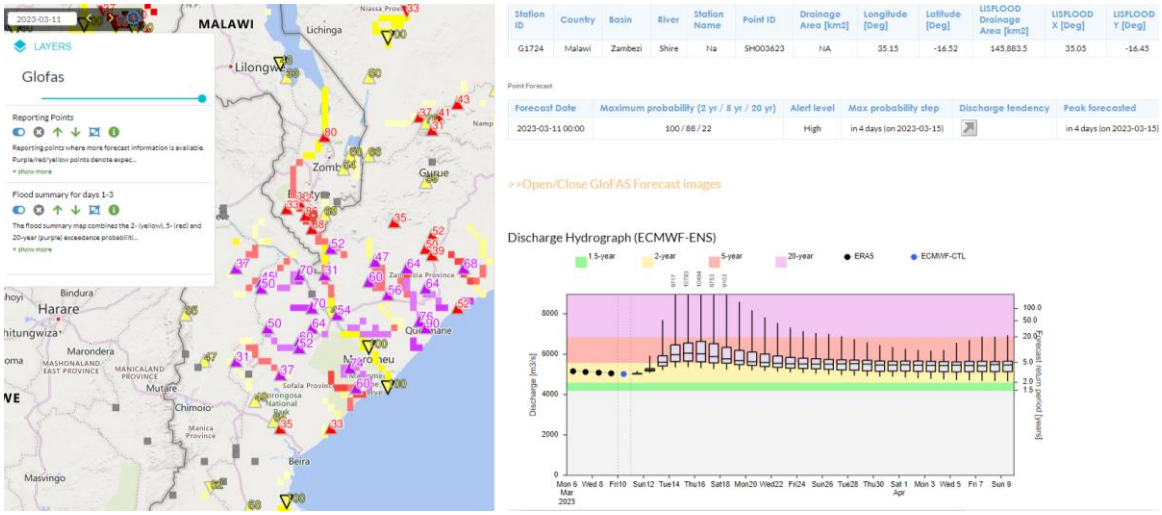


Figure 2-4. Screenshot of the GloFAS map viewer for the recent case of Cyclone Freddy, which made its final landfall in Mozambique on 11th March 2023. The selected point is on the Shire River in Malawi, which did experience significant flooding and impacts (NASA earth observatory, 2023).

Several studies have investigated the potential for GloFAS forecasts to be used in humanitarian preparedness activities and flood early warning in particular regions such as the Amazon basin (Towner et al., 2019), Bangladesh (Hossain et al., 2023), Uganda and Kenya (Mitheu et al., 2023), and Myanmar and the Philippines (MacLeod et al., 2021). The research presented in Chapters 5 and 6 provided the first opportunity to investigate the potential of GloFAS forecasts and GloFAS-ERA5 reanalyses to help in flood early warning and flood predictability assessment specifically in flooding caused by TCs.

2.5.4 *Verification methods*

Traditional verification measures of hydrological performance include the root-mean square error (RMSE), mean absolute error (MAE), Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970), Pearson correlation coefficient and modified Kling–Gupta efficiency (KGE) (Gupta et al., 2009; Kling

et al., 2012). The KGE is increasingly considered as the standard performance metric in hydrology (Knoben et al., 2019; Lin et al., 2019). When considering the output of ensemble flood forecasts in a probabilistic form, other measures suited for probabilistic forecasts such as those described in Section 2.4.2 and initially developed for use with meteorological ensemble data, are now increasingly being applied to hydrological ensemble forecasts (Brown et al., 2010; Demargne et al., 2010).

In Chapter 6 the modified KGE is used to score the hydrographs from each ensemble member in order to compare performance with other stages of the TC flood forecast chain such as the track and precipitation errors.

2.6 Summary

This chapter has provided additional background information on the current state of ensemble forecasting, TC forecasting and research, and river flood forecasting, providing additional motivating material behind the aim and objectives of this research, and introducing the main datasets and methods used in the thesis. The following four chapters contain the four main papers published as part of this research, with the details of the datasets and methods that are used in each of these studies described in the relevant section of each of these chapters.

Chapter 3 Current and potential use of ensemble forecasts in operational TC forecasting: results from a global forecaster survey

The first objective of the PhD was to identify the current and potential use of ensemble forecasts by operational TC forecast centres. The paper addressing this objective has been published in journal Tropical Cyclone Research and Review (TCRR) with the following reference:

Titley, H.A., Yamaguchi, M. and Magnusson, L. (2019) Current and potential use of ensemble forecasts in operational TC forecasting: results from a global forecaster survey. Tropical Cyclone Research and Review, 8 (3): 166–180. doi:10.1016/j.tcrr.2019.10.005.

The contributions of the authors of this paper are as follows: H.T. designed the experiment, produced the datasets, carried out the analysis, and led the writing of the manuscript. M.Y. and L.M. assisted with posing the research question and designing the analysis. All authors assisted with writing the manuscript. Overall, 90% of the writing was undertaken by H.T..

The published article can be found in the Appendix A1.

Abstract. In order to understand the current and potential use of ensemble forecasts in operational tropical cyclone (TC) forecasting, a questionnaire on the use of dynamic ensembles was conducted at operational TC forecast centers across the world, in association with the World Meteorological Organisation (WMO) High-Impact Weather Project (HIWeather). The results of the survey are presented, and show that ensemble forecasts are used by nearly all respondents, particularly in TC track and genesis forecasting, with several examples of where ensemble forecasts have been pulled through successfully into the operational TC forecasting process. There is still, however, a notable difference between the high proportion of operational TC forecasters who use and value ensemble forecast information, and the slower pull-through into operational forecast warnings and products of the probabilistic guidance and uncertainty information that ensembles can provide. Those areas of research and development that would help TC forecasters to make increased use of ensemble forecast information in the future include improved access to ensemble forecast data, verification and visualizations, the development of hazard and impact-based products, an improvement in the skill of the ensembles (particularly for intensity and structure), and improved guidance on how to use ensembles and optimally combine forecasts from all available models. A change in operational working practices towards using probabilistic information, and providing and communicating dynamic uncertainty information in operational forecasts and warnings, is also recommended.

3.1 Introduction and aims of the questionnaire

For over 25 years, dynamical ensemble model forecasts have been routinely produced by several global numerical weather prediction (NWP) modelling centers (Palmer, 2019). Instead of making a single forecast of the most likely weather, an ensemble of forecasts is produced, taking into account uncertainty in the initial conditions, and imperfections in the model formulation, and aiming to give an indication of the range of possible future states of the atmosphere. Buizza (2008) described the two main advantages of ensemble forecasts over deterministic forecasts: "first, the fact that an ensemble system predicts not only the most likely scenario but also the probability of occurrence of any event, and second the fact that an ensemble system can provide more consistent (i.e. less changeable) successive forecasts."

Dynamic ensemble forecasting of tropical cyclones (TCs) has an important role to play in capturing the situation dependent uncertainty in the forecasts for existing storms (e.g., Dupont et al., 2011; Yamaguchi et al., 2012; Du et al., 2016; Leonardo and Colle, 2017; Zhang and Yu, 2017), and in providing probabilistic information about TC genesis (e.g., Yamaguchi et al., 2015; Yamaguchi and Koide, 2017). Many global NWP centers produce TC track forecasts from their ensemble forecast models, develop TC forecast products, verify the forecasts, and share these via The International Grand Global Ensemble (TIGGE) cyclone exchange program (Swinbank et al., 2016). There is, however, great potential to increase the use of this probabilistic ensemble forecast information in operational TC forecasts and warnings.

The World Meteorological Organisation (WMO)'s 10-year high impact weather research project, HIWeather, aims to enable increased global resilience to severe weather by improving forecasts of severe weather and its impacts, and the communication of information to users, especially emergency managers (Jones et al., 2014). As a project within HIWeather, and in association with the World Weather Research Programme (WWRP) Predictability, Dynamics and Ensemble Forecasting (PDEF) working group, a group of international researchers started a collaboration aimed at enhancing collaboration amongst the research and operational community on the topic of ensemble forecasting and verification of TCs. The first activity was to design and circulate a questionnaire to all operational TC forecasting centers around the world asking about their use of dynamical ensemble forecast information.

While it is acknowledged that the use of multiple deterministic forecasts in consensus forecasting is currently a valuable cornerstone for track and intensity forecasting (e.g., Goerss, 2000; Goerss, 2007; Sampson et al., 2008; Goerss and Sampson, 2014; Yamaguchi et al., 2017), in order to create a manageable and clear scope for the questionnaire, neither consensus, statistical, nor analog-

based ensembles, were included in the definition of ensembles. The questionnaire defined "ensemble" to be a dynamically-based ensemble run with multiple members, using a combination of initial condition and model physics perturbations to represent the initial condition and model uncertainty, in order to focus on the current and potential use of dynamic ensembles in operational TC forecasting.

The questionnaire aimed to report a baseline on the current use of ensembles at global operational TC forecast centers to the 9th International Workshop on Tropical Cyclones (IWTC-9), and help shape future research and development in order to maximise the pull through of the benefits of ensemble forecasts into operational TC forecasts and warnings. The survey took the form of a Google Form, and was sent to key contacts at operational centers worldwide using an email distribution list provided by WMO, including Regional Specialised Met Centers (RSMCs), Tropical Cyclone Warning Centers (TCWCs), and the Typhoon Committee Secretariat, for onward circulation to their members. The forecasting centers were requested to circulate the survey to all forecasters at their center, and several follow-up reminders were also sent to ensure as wide a participation as possible. The more detailed objectives of the survey were as follows:

a) To document the current availability of ensemble forecasts in tropical cyclone forecast centers, and their use by operational forecasters.
b) To ascertain how uncertainty is represented in the operational warnings for track, intensity, genesis and hazards from their center, and whether or not this uncertainty information is taken from dynamical ensemble forecasts.
c) To obtain examples where probabilistic forecasts have been successfully integrated into operations, but also occasions where hurdles have prevented them from being fully utilised.
d) To collate forecaster feedback on where they would like to see future research and development focussed to enable them to make wider use of ensemble forecasts.

The list of questions in the survey is contained in Appendix 3-1 at the end of this chapter. The results of the questionnaire were presented at IWTC-9 in Hawaii in December 2018, as the main component of sub topic 6.3 titled “Current and Potential Use of Ensemble Forecasts in Operational TC Forecasting” (WMO, 2018). This paper contains a more detailed write-up of the results of the questionnaire. Section 3.2 contains the results of the questionnaire, organised into sub-sections according to the four main objectives as stated above; and Section 3.3 gives conclusions and recommendations going forward.

3.2 Results of the questionnaire

3.2.1 *Current availability and use of ensemble forecasts*

It was important to ensure that the questionnaire results took into account the views and experiences of forecasters in all areas of the world, and from a variety of types of forecast centers. A total of 60 questionnaire responses were received from 25 different centers around the world, including RSMCs, TCWCs, national meteorological services, and the Joint Typhoon Warning Center (see Figure 3-1). Please note that no question was compulsory so there are cases where the total number of respondents does not add up to 60. Figure 3-2 shows the TC basins that were selected to be of interest to the respondents, and indicates that there is a good balance globally, with at least 14 respondents having forecast interest in each of the global basins (with many respondents having forecast interest in more than one basin). The majority of respondents were operational forecasters, but there were also inputs from managers, researchers, and others such as technical developers.

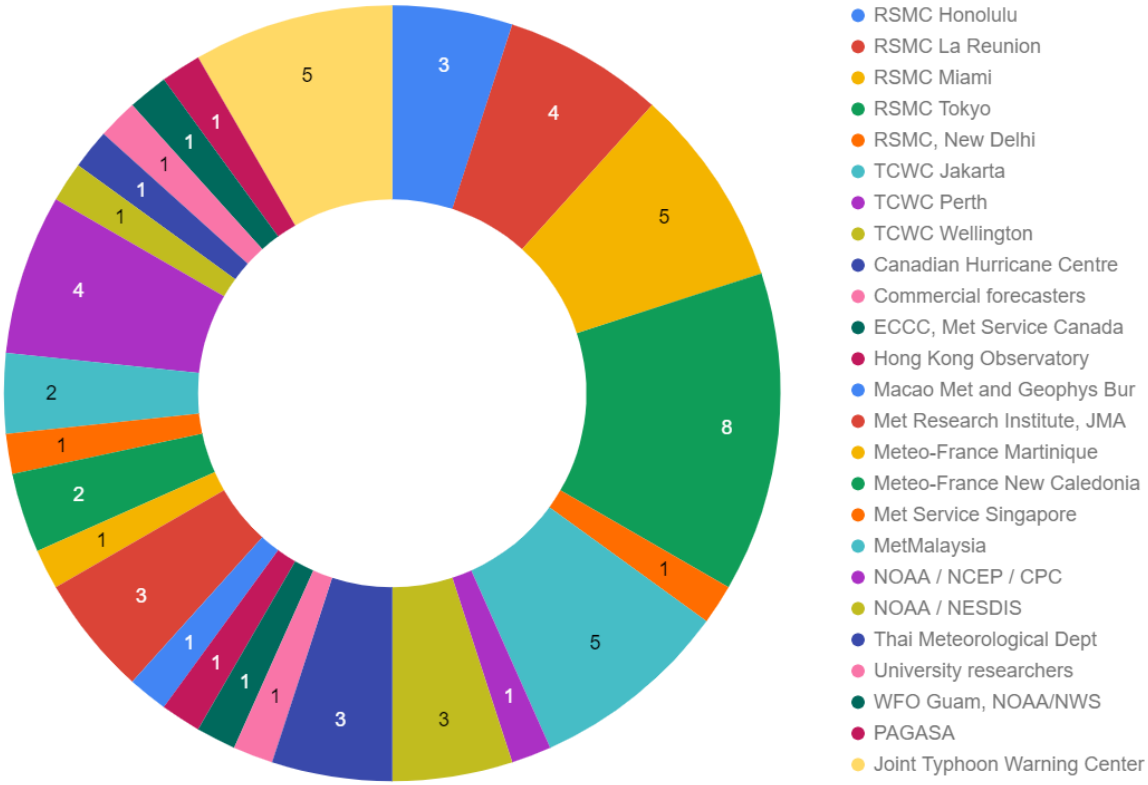


Figure 3-1. Organisation of respondents to the questionnaire (including RSMC/TCWC where applicable).

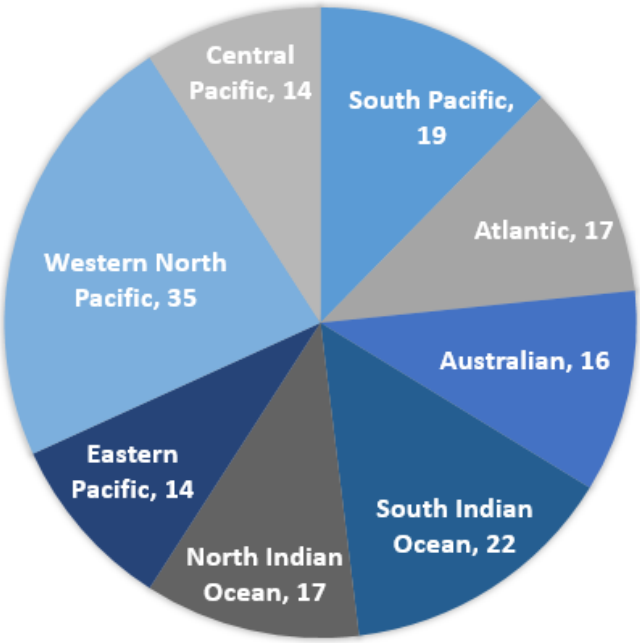


Figure 3-2. The tropical cyclone basins that were of interest to the respondents.

Dynamic ensemble forecasts were used by nearly all respondents (95%). This is maybe not a surprise, as although all forecasters were encouraged to fill in the questionnaire regardless of their use of ensembles, it is to be expected that there would be a higher degree of motivation to complete the survey if they were interested in ensemble forecasting. Nonetheless it shows that a large proportion of operational TC forecasters use ensemble forecasts. Those who did not were from National Met Services in developing countries, where ensemble forecasts are not readily accessible. Figure 3-3 shows the breakdown of how important respondents felt the use of ensemble forecasts was in each of the main areas of TC forecasting. Ensemble forecasts are seen to be particularly important in track and genesis forecasting. For hazard forecasting and impact-based forecasting, although a handful of respondents did not use ensemble forecasts, the majority did, but the importance of ensemble forecasts was less than in track and genesis forecasts. In intensity forecasting, the categories were very evenly split, with a significant proportion of respondents not using ensemble forecasts. The areas where ensemble forecasts were used least were in forecasting TC size/structure, and in extra-tropical transition forecasting (although this result may be influenced by not all of the forecasters being involved in predicting this process).

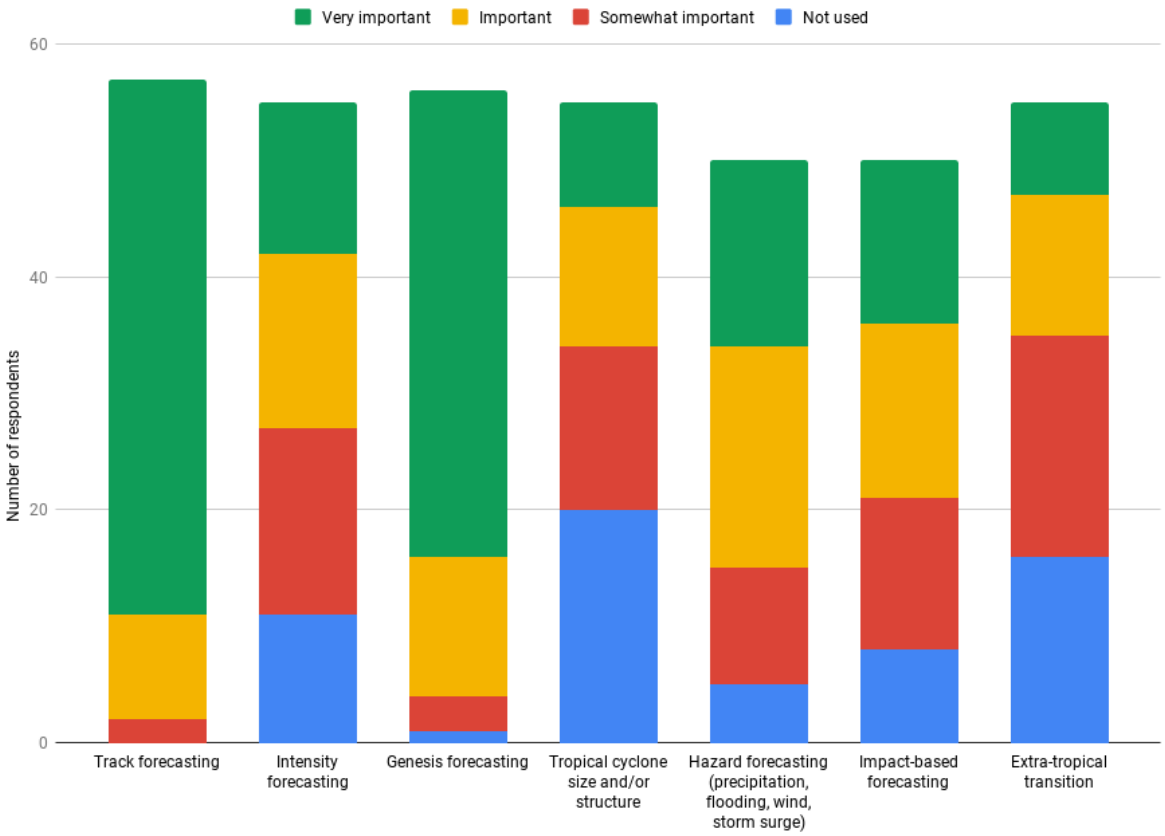


Figure 3-3. The questionnaire asked "How important would you say ensemble forecasts are in your forecasting of tropical cyclones?", in each of the above categories.

The ensemble forecast information was accessed in a variety of different ways. Over half of respondents were able to use ensemble-based products created at their center, but even larger numbers looked at online products from other centers or products sent by global NWP centers, either in combination with those produced in-center, or, in many cases, being reliant on these external sources of ensemble forecast information. Ensemble data was also accessed from the WMO Global Telecommunications System (GTS) and from the TIGGE archive. The most frequently used ensemble models worldwide are the European Centre for Medium-Range Weather Forecasts Ensemble (ECMWF ENS) and the National Centers for Environmental Prediction Global Ensemble Forecast System (NCEP GEFS) ensembles, followed by the UK Met Office Global and Regional Ensemble Prediction System (MOGREPS-G), Japan Meteorological Agency's global ensemble and the Canadian global ensemble (see Figure 3-4). Again these models are used primarily for their track and genesis forecasts, although a significant number do also use NCEP GEFS and ECMWF ENS for intensity and hazard forecasting. A range of other ensembles are used by a smaller number of respondents.

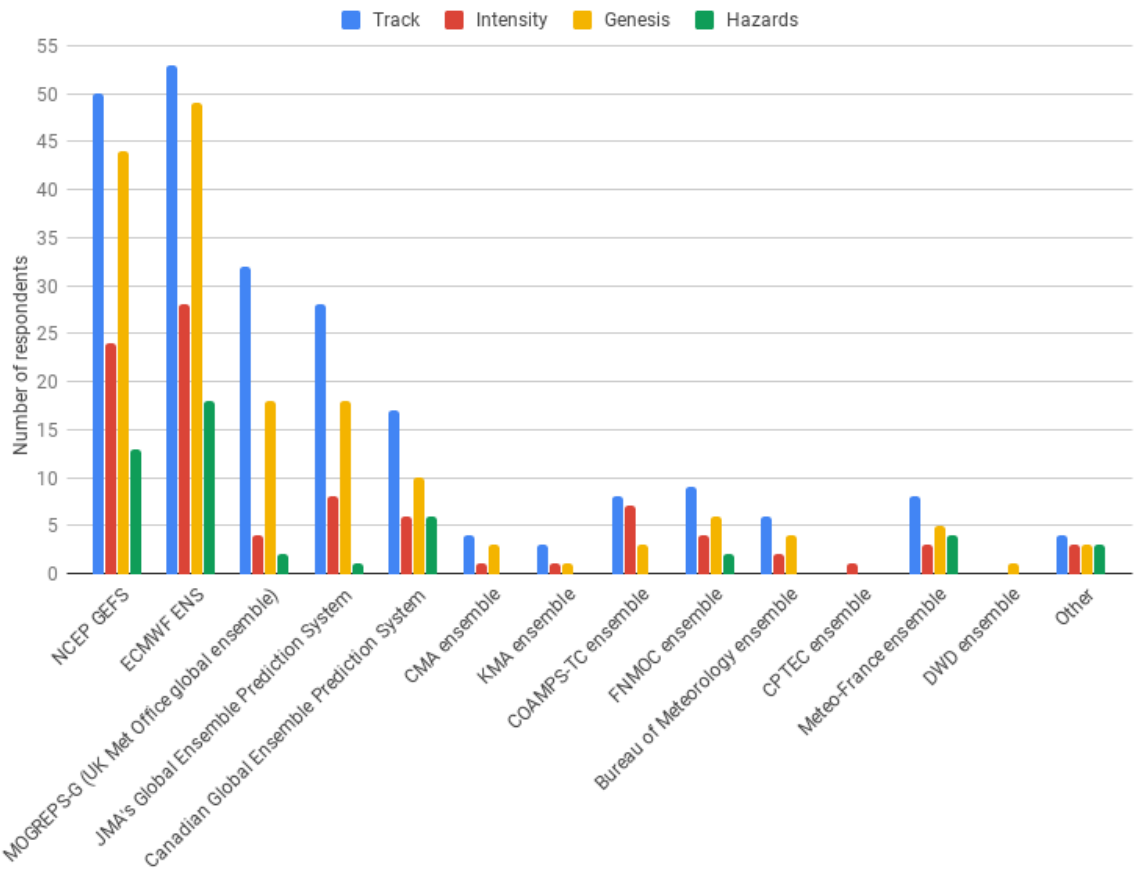


Figure 3-4. "For each of the main areas of TC forecasting, which ensemble forecast models do you use?"

Over 90% of respondents used multi-model ensemble forecasts (multiple dynamical ensembles), with around half using a full probabilistic multi-model ensemble combination and half mainly by comparing one ensemble with another. Calibrated ensemble forecasts were less-well used, with 40% of respondents using them for track forecasting, and between 20 and 25% for intensity and genesis forecasting. Figure 3-5 shows that the product type used most regularly was the ensemble mean track, followed by the tracks from the ensemble members (plumes), which were used more regularly than the strike probabilities created from them. However, the probabilistic information in the genesis products is well-used. It was noticeable that many respondents were keen to have future access to additional ensemble-based products such as landfall probability, storm surge, river discharge, flood risk and impact-based products.

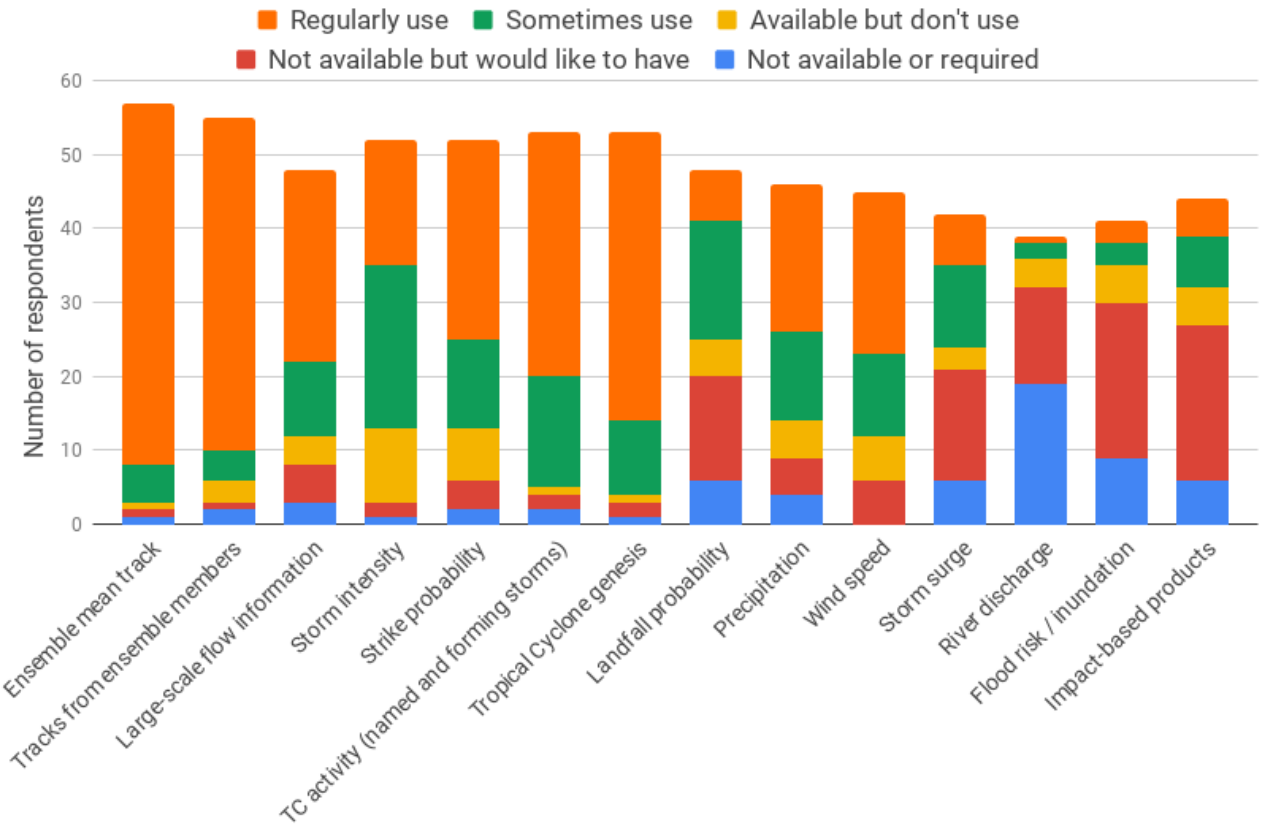


Figure 3-5. "Which ensemble-based or probabilistic products do you use?"

When asked if there was a particular product or ensemble that they would like to be able to see the answers were very varied, including: i) an operational ensemble-based cone of probability; ii) more useful intensity guidance including probability of intensity change and bias corrected intensity forecasts from multiple ensembles; iii) hazard forecasts e.g. high-impact winds and precipitation, storm surge, river discharge, flash flood risk/inundation; iv) impact-based products; v) box-whisker plots of TC tracks; and vi) ensemble clustering with probabilities.

Figures 3-6 and 3-7 show the differences in how forecasters use ensemble forecasts in two areas of TC forecasting: tracks and intensity. Only four respondents said that they rarely used ensemble forecasts for track forecasting, but this rose to half of respondents for intensity. However, well over half of people for track, and a third for intensity, said that they mainly used ensemble forecasts to compare the ensemble mean forecasts to the deterministic forecasts. This was more than those who used the full probabilistic forecast information provided by the ensemble, indicating that at present the ensemble information may be being used in a deterministic-focussed environment, where it is harder to make full use of the probabilistic forecast information provided by the ensemble forecasts. For track, three quarters of respondents found the spread useful to provide situation-dependent uncertainty, falling to one third for intensity. Hardly any respondents ticked

that they "don't find the ensemble forecasts very helpful when the spread is very large", indicating that they are finding cases with a large spread to be correctly highlighting the increased uncertainty in that forecast.

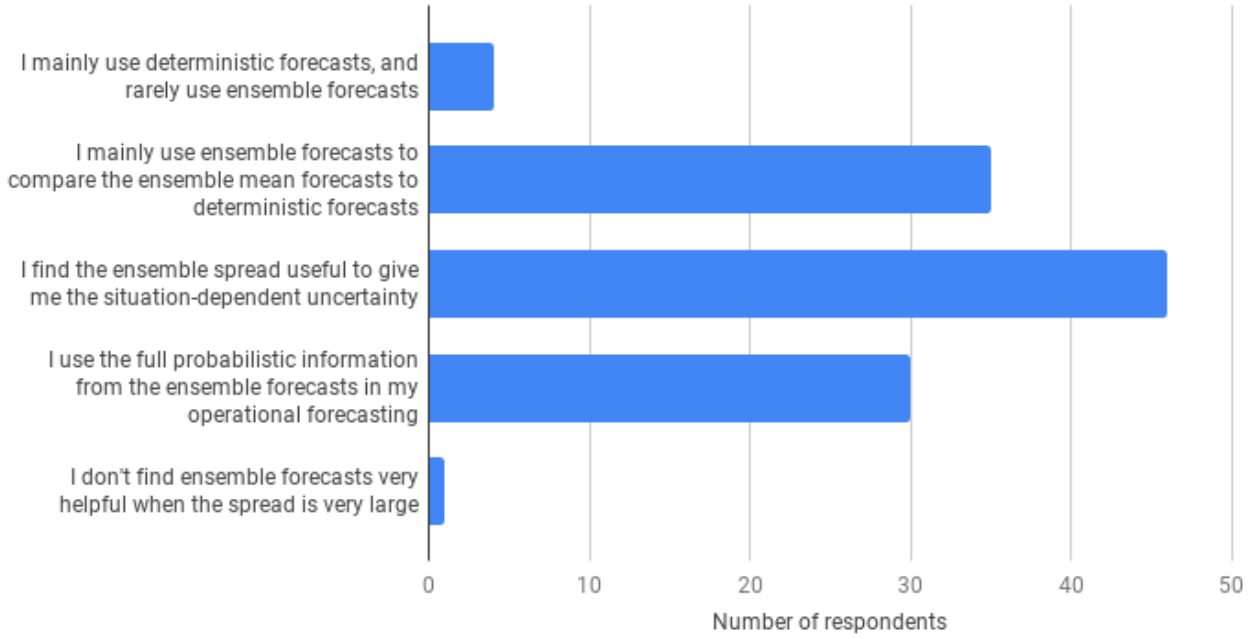


Figure 3-6. "Which of the following statements describe your use of ensemble forecast information in track forecasting?"

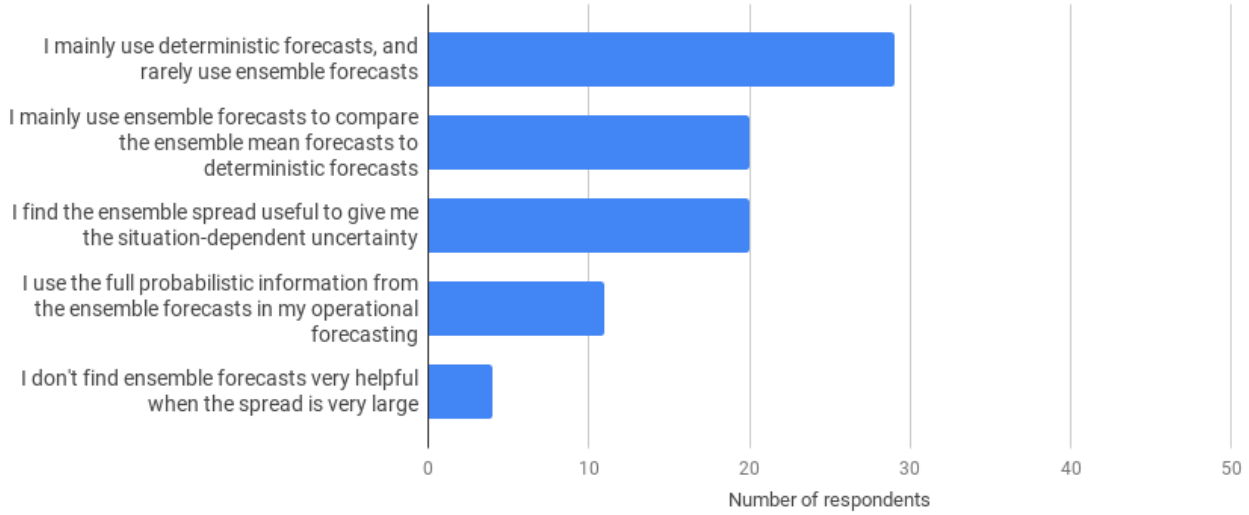


Figure 3-7. "Which of the following statements describe your use of ensemble forecast information in intensity forecasting?"

3.2.2 *How is uncertainty represented in operational warnings?*

Tables 3-1, 3-2, 3-3, and 3-4 detail how forecast uncertainty is represented in operational forecasts and warnings issued from each separate center (where this information was available from respondents). Table 3-1 shows the responses from each center for their operational track forecasts. There is no standard forecast length or frequency, with forecast lengths of 3, 5 or 7 days, and

frequencies of 3, 6, 12 or 24 hours. Uncertainty is represented in all forecasts, but this can be via a circle, a cone, or an error swath. The size of each circle/cone/error swath is generally set so that a storm would be expected to fall within it in for a given percentage of cases, but this percentage varies too from center to center, from 66% to 80%. Out of the RSMCs, only two include ensemble spread in this calculation (RSMC La Réunion and Tokyo). More commonly used are historical forecast error statistics. Out of the other warning centers only TCWC Wellington and Météo France New Caledonia include ensemble spread when calculating the uncertainty in their operational forecasts. Thus there appears to be a clear discrepancy between the number of operational TC forecasters who use and value ensemble forecast information, as shown in the previous section, and the pull-through into operational forecast products of the probabilistic guidance that ensembles can provide.

Table 3-2 shows the responses from each center about their operational intensity forecasts. Forecast lengths and frequencies again vary between 3, 5, and 7 days, at 3, 6, 12 or 24 hourly frequency. Only RSMC Honolulu and RSMC Miami said that uncertainty was included, with this being in the form of wind speed probabilities from a statistical model in both cases. Dynamical ensemble forecasts are however often included when calculating the uncertainty in operational TC genesis forecasts (Table 3-3), often in combination with consensus forecasting, climatological information, and forecaster judgement. Forecast lengths vary between 1 and 7 days. RSMC Tokyo issue with the highest frequency (3 hours), with many forecast centers issuing their genesis forecasts once per day. The most used way of representing uncertainty is to place forecasts into probability categories such as high/medium/low.

Table 3-4 shows that the expression of uncertainty and the use of ensembles in operational hazard forecasts varies considerably, with products often taking the form of named or color-coded categories assigned based on probabilities (statistical or dynamical).

Center name	Forecast lengths / frequency	How is uncertainty represented (and what proportion of cases are expected to stay within this)?	How is it calculated?
RSMC Tokyo	5 days / 8 times a day	Circle (70%)	Historical forecast error statistics (up to 72 hours) and ensemble spread (after 96 hours, Yamaguchi et al. 2009)
RSMC Honolulu	5 days / 6 hourly	Cone (67%)	Historical forecast error (5 year)
RSMC La Reunion	5 days / 6 hourly	Cone (75%)	Based on both ensemble spread and historical forecast error, (Dupont et al. 2011)
RSMC Miami	5 days publicly, 7 days internally / 6 hourly (more frequent when necessary)	Cone (67%)	Previous 5 years of track error for that basin
RSMC New Delhi	5 days / 6 hourly (and 3 hourly updates for intervening)	Cone (72%)	Historical forecast error statistics
TCWC Jakarta	3 days / 6 or 12 hourly	Cone (80%)	Consensus spread. Patterns obtained from wind model forecasts, and sometimes the form is modified manually to make the shape smoother.
TCWC Perth	7 days / 1, 3, or 6 hourly	Cone	Situation dependent, usually either consensus spread or climatological uncertainty but can be manually manipulated
TCWC Wellington	24 hours, but up to 5 days if a threat to NZ / 6 hourly	Cone (70%)	Calculated based on consensus spread, ensemble spread and climatological uncertainty
Joint Typhoon Warning Center (JWTC)	5 days / 6 hourly	Error swath	Calculated by adding the JTWC 5-year running mean forecast track error to the forecast 34-knot wind radii at each forecast time
Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA)	5 days / 6 hourly	Cone / circle (70%)	Historical forecast error statistics
Weather Forecast Office (WFO) Guam	5 days / 6 hourly	Error swath (66%)	Standard deviation of track error
Thai Meteorological Department	3 days / 3, 6, 12 hourly dependent on situation	Cone	Historical forecast error statistics
Malaysian Meteorological Department (MetMalaysia)	7 days / 3 hourly if TC with in our region	Circle (80%)	Not given
Météo-France New Caledonia	3 days / 6 hourly	Cone (75%)	Choice between climatological uncertainty and ensemble-based uncertainty (ECMWF). We mostly use climatological uncertainty for operational reasons (faster).
Météo-France Martinique	5 days / 6 hourly	Cone (66%)	Climatological uncertainty
Hong Kong Observatory	5 days / 24 hourly	Cone/Circle (70%)	Historical forecast error statistics
Canadian Hurricane Center	5 days / 6 hourly	Cone (70%)	Historical forecast error

Table 3-1. Operational track forecasts from each tropical cyclone forecasting center.

Center name	Forecast lengths / frequency	Is uncertainty represented, and if so how?
RSMC Tokyo	3 days / 8 times a day (extended to 5 days in March 2019)	No
RSMC Honolulu	5 days / 6 hourly	Yes, wind speed probabilities for 34/50/65 knots – Monte Carlo simulation
RSMC La Reunion	5 days / 6 hourly	No
RSMC Miami	5 to 7 days / 6 hourly	Yes, wind speed probabilities
RSMC New Delhi	5 days / 6 hourly (and 3 hourly updates for intervening)	No
TCWC Jakarta	3 days / 6 hourly	No
TCWC Perth	7 days / 1, 3 or 6 hourly	No
TCWC Wellington	24 hours but up to 5 days if a threat to NZ / 6 hourly	No
JTWC	5 days / 6 hourly	No (only qualitatively in discussions)
PAGASA	Only forecast the TC by category / 6 hourly	No
WFO Guam	5 days / 6 hourly	No
Thai Meteorological Department	3 days / 3, 6, 12 hourly depending on situation	No
MetMalaysia	7 days / 6 hourly	No
Météo-France New Caledonia	3 days / 6 hourly	No
Météo-France Martinique	5 days / 6 hourly	No
Hong Kong Observatory	5 days / 24 hourly	No
Canadian Hurricane Center	5 days / 6 hourly	No

Table 3-2. Operational intensity forecasts from each tropical cyclone forecasting center.

Center name	Forecast lengths / frequency	How is uncertainty represented?	How is the uncertainty calculated?
RSMC Tokyo	1 day / 8 times a day	Not represented	n/a
RSMC Honolulu	5 days / 6 hourly	High/medium/low categories with assigned probabilities	Subjectively, forecasters use climatology, multi-model consensus, ensembles, and satellite observation trends
RSMC La Reunion	5 days / daily	Probability categories (very low, low, moderate, high, very high)	Ensemble based and forecaster judgement
RSMC Miami	5 days / 6 hourly	Percentages to nearest 10% that are grouped into high/medium/low categories	Combination of ensembles, multi-model ensemble, and forecaster experience
RSMC New Delhi	5 days / daily	Nil, Low, Fair, Moderate, High corresponding to 00, 1-25, 26-50, 51-75, 76-100% of probability	Based on consensus derived from about 10 deterministic and probabilistic model guidance
TCWC Jakarta	3 days / 6 hourly	high/medium/low	Climatological, consensus
TCWC Perth	3-7 days / daily	High/moderate/low/very low	Assigned by forecaster
TCWC Wellington	5 days / daily	Low potential/moderate potential/ high potential	Uncertainty is based on ensemble (mostly ECMWF and MOGREPS tropical storm strike/ genesis probabilities) and forecaster diagnosis and prognosis of the environmental conditions susceptible to TC genesis
JTWC	14 days / 12 hourly	Low (<40%), Medium (40-60%), High (>60%) classification	Deterministic and ensemble model forecasts, MJO and statistical-dynamical tools
PAGASA	48 hours / daily	Using terms like "less likely", "likely" to develop	Ensemble based and other trends in TC parameters
WFO Guam	5 day / 12 hourly	Discussion, and use of JTWC high, medium and low	NWP, climatology and experience with synoptic data (including upper air forecasts)
Thai Meteorological Department	5 days /6 or 12 hourly depending on situation	Probability	Ensemble-based
Météo-France New Caledonia	7 days / daily	High/medium/low	Ensemble-based and multi-model
Météo-France Martinique	5 days (6 hourly), one week (12 hourly) and one month (Mon/Thu).	5 days (high/medium/low) Longer range (percentiles)	5 days: from NHC guidance Longer range: from ECMWF
Canadian Hurricane Center	Follow National Hurricane Center (NHC) guidance.	n/a	n/a

Table 3-3. Operational genesis forecasts from each tropical cyclone forecasting center.

Center name	In hazard forecasts e.g. precipitation, winds, storm surge, flooding, how is uncertainty represented?	How is the uncertainty calculated?
RSMC Tokyo	Hazard forecasts not currently included in our service	n/a
RSMC Honolulu	Wind speed probability graphics for 34/50/65 knots Adjective category and color coded hazard threat index: - Little to None - Elevated (changed from Low) - Moderate - High - Extreme	Probabilistic according to the hazard e.g. wind speed probabilities
RSMC La Reunion	No probabilistic data available for the public	n/a
RSMC Miami	In categories (representing probabilities), that vary depending on the hazard.	Generally ensemble based, either statistically or direct from ensemble output
RSMC New Delhi	Forecast for heavy precipitation is given in range (7-11 cm, 12-20 cm and more than 20 cm in 24 hrs) and likelihood of occurrence (likely: 26-50%, very likely: 51-75% and most likely: 76-100%). The winds, storm surges and flooding are also mentioned in ranges with above probability of occurrence.	Uncertainty is calculated based on likelihood of occurrence and confidence level of forecast based on various deterministic and ensemble models. Accordingly the colour codes of green, yellow, orange and red are used in increasing order of severity based on confidence level and likelihood of occurrence.
TCWC Perth	Probabilistic language (likely, possibly etc)	Assigned by forecaster
TCWC Wellington	High/medium/low risk in severe weather outlooks	Subjective assessment based on analysis of system and environment, deterministic forecasts, and ensemble outlooks.
JTWC	n/a	n/a
PAGASA	In description e.g. we say that the hazard is possible to happen	Based on daily observations and deterministic forecast from other models
WFO Guam	Descriptively	Forecast accuracy and type of system
Thai Meteorological Department	Show on map	Risk matrix
Météo-France New Caledonia	There is no graphical representation of hazard forecasts. They are mentioned in bulletins with key words ("probably", "risk", "certainly" etc).	Ensemble-based and multi-model
Météo-France Martinique	Not available for the public	ECMWF EPS and ARPEGE PEARP
Hong Kong Observatory	Uncertainty in the change of tropical cyclone warning signal, which is related to local wind forecast, is represented in five categories of probability for government departments	Based on the probabilistic relationship between local winds and position/intensity of tropical cyclone
Canadian Hurricane Center	Listing the worst case scenario and the sensitivity of forecast errors. Qualitatively in free form text discussion.	Empirically, plus local effects and criteria (i.e. known flooding levels for storm surge)

Table 3-4. Operational hazard forecasts from each tropical cyclone forecasting center.

3.2.3 *Integration of probabilistic forecasts into operations*

Respondents were asked for examples where probabilistic forecasts have been successfully integrated into operations, and also for occasions where hurdles have prevented them from being fully utilized. Numerous excellent examples were given of where probabilistic forecasts have been pulled through into operational forecasts and warnings, including the ensemble-based RSMC La Réunion operational track uncertainty cone (Dupont et al., 2011), and the multi-model ensemble TC track and activity forecast products made available from RSMC Tokyo to Typhoon Committee members based on the outcomes of the North Western Pacific Tropical Cyclone Ensemble Forecast Project (NWP-TCEFP, Yamaguchi, et al., 2012; Yamaguchi, et al., 2014; Yamaguchi et al., 2015). Other examples were the successful integration of ensemble probabilities in the production of

Tropical Cyclone Potential Bulletins at TCWC Wellington, and the calibrated genesis potential guidance from the CFS, GEFS, and ECMWF ensembles that are used directly in the creation of the Global Tropical Hazards and Benefits outlook created by the NWS Climate Prediction Center.

Just over half of respondents stated that they now placed more emphasis on using dynamical ensemble forecast information rather than using the traditional consensus approach using multiple deterministic forecasts. Of the consensus-favouring group, several said this was mainly due to habit, tradition, or because they had neither the desired access to, understanding of, nor the tools required to use the dynamical ensemble forecasts. Others felt the consensus approach was more skilful, particularly at shorter lead times and for intensity and structure forecasts, due to the higher resolution and the different model physics, biases, and parameterizations when combining multiple deterministic models. Several people felt that they would have liked to have answered that both were equally important, with the consensus forecasts better suited to the required deterministic components of the products and warnings, and ensembles being more useful for probabilistic forecasts and communicating uncertainty, while others requested research to compare the skill from each approach, and develop methods that combine and maximize the benefit from both consensus and ensemble forecasts.

The main hurdles experienced that had prevented ensemble forecasts from being fully utilized fell into four broad categories:

i) *Availability of ensemble data* e.g. lack of access to ensemble data at their center, lack of inclusion of ensembles in their key operational or visualization tools, and late availability time in an operational context (particularly from ECMWF ENS). Several respondents named TIGGE CXML as a great resource but would like to see it extended so that all centers provide forming information, and additional information such as intensity, structure and precipitation, which would increase research opportunities and lead to more research to operations (R2O) achievements in the future. Another respondent would like to see ensemble forecast information consistently available in Automated Tropical Cyclone Forecasting System (ATCF) format.

ii) *Performance of the ensemble forecasts* e.g. concerns that the resolution is not useful for intensity prediction, about unrealistic ensemble spread (respondents commented both on excessive spread with unrealistic solutions and on the forecasts often being under-spread), and that ensembles can struggle to represent the vortex and get the structure that is essential to assess the hazard risks.

iii) *Lack of familiarity with the interpretation of ensemble and probabilistic forecasts* by forecasters, and problems with synthesizing the enormous amount of data in a time-restricted operational environment.

iv) *Challenge of communicating uncertainty and probabilistic forecasts to customers and the public*, who are used to deterministic weather bulletins and warnings.

3.2.4 *Priorities for future research and development*

The final section of the questionnaire asked the respondents where they would like to see future research and development focus to maximise the utility of and benefit from ensemble forecasts in operational TC forecasting. Figure 3-8 shows the scores from 1 (not important) to 5 (very important) given by the respondents when asked which of the statements would help them make better use of ensemble forecast information in the future. All of the ideas were popular, with the yellow, green and orange bars (representing the higher end of the scale) dominating on all of the statements. However, the top statements having the highest number of respondents selecting a "5" were: i) Improvement in the skill of the ensemble forecasts (intensity); ii) Greater access to ensemble forecast data; iii) Improvement in the skill of the ensemble forecasts (track); iv) Better understanding of best practice in ensemble-based forecasting of TCs; and v) The development of hazard-based ensemble products.

Hazard-based and impact-based products also featured strongly when asked for specific details of a particular product that they would like to be able to see. Other requests were an operational ensemble-based cone of probability, more useful intensity guidance including probability of intensity change and bias corrected intensity forecasts from multiple ensembles, and ensemble clustering with probabilities.

The most popular verification metric amongst the respondents for evaluating the skill and value in ensemble forecasts of TCs was ensemble spread, followed by the root mean square (RMS) error of the ensemble mean, reliability diagrams, and the % of forecasts falling within an ensemble defined ellipse/cone (see Figure 3-9). Several ensemble-based verification metrics such as Relative Operating Characteristic (ROC) diagrams, Brier Skill Score, Ranked Probability Skill Score, and in particular economic value plots, were not as well used or understood, and therefore ensemble-based verification may need to be communicated more clearly, or adapted to provide more user-relevant information. The respondents gave several ideas for more user-oriented verification information that they would like to see that are more relevant to their day-to-day forecasting experience. For example, only including members that had a similar track to the actual event in intensity verification, creating and verifying metrics that translate ensemble spread into categorical

estimates of forecast confidence in comparison to the widely used historical forecast error statistics, and regularly updated or real-time performance metrics comparing the performance and skill for all types of models (deterministic, consensus and ensemble models) for the current storm, basin, and season-to-date.

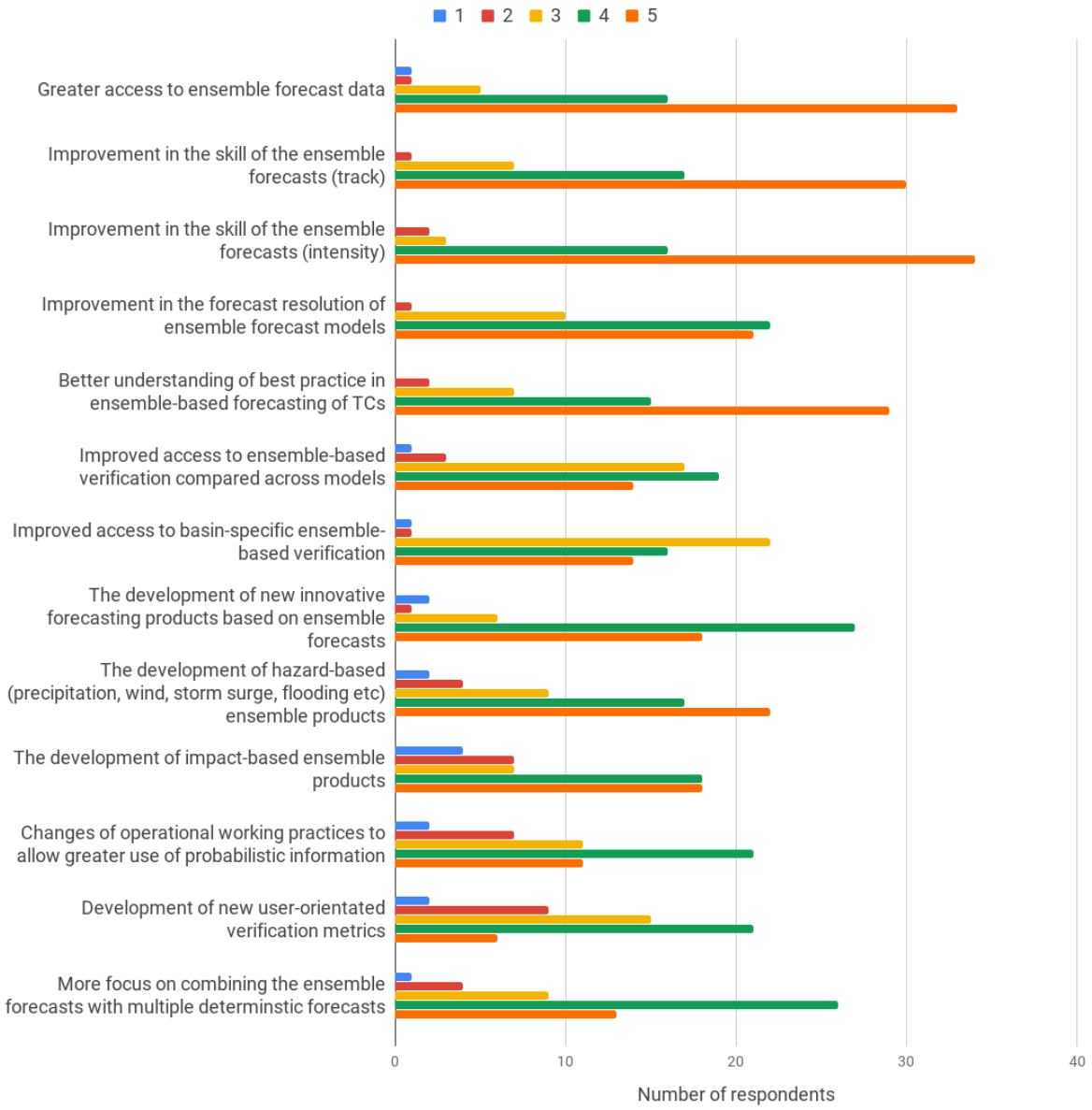


Figure 3-8. The questionnaire asked which of the above statements would help them make better use of ensemble forecast information in the future, rating each idea from 1 (not important) to 5 (very important).

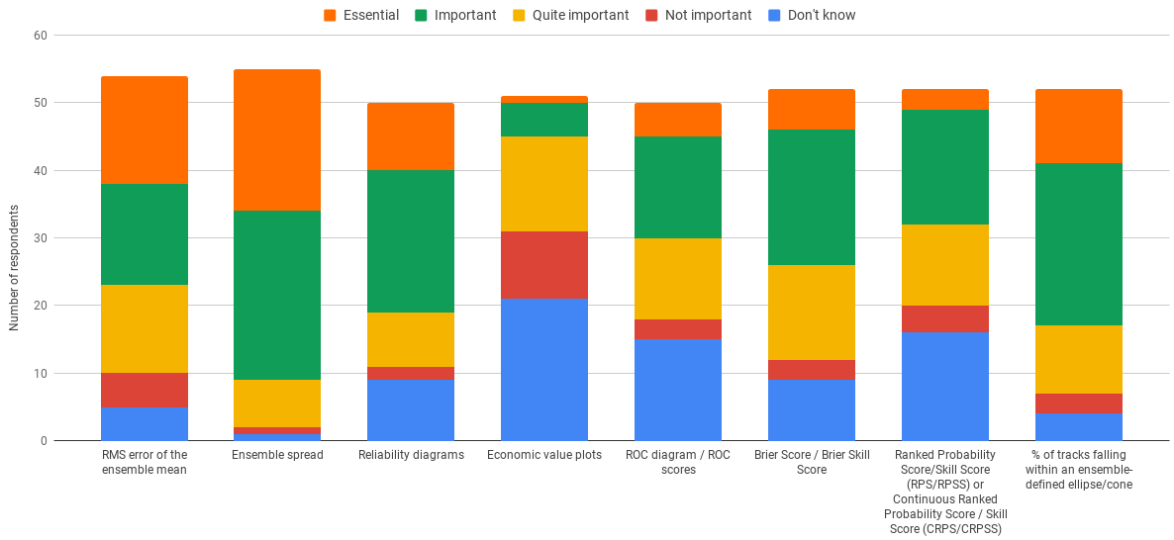


Figure 3-9. "Please state how important you feel each of the following verification metrics are when evaluating the skill and value in ensemble forecasts."

At the end of the questionnaire respondents were asked what specific developments they would like to see. The key themes coming through these responses are detailed below:

- Several responses urged a focus on how to ***communicate uncertainty to the general public***, and explain how the public can use probabilistic information, with a need to move away from the very deterministic-orientated information given in weather charts and bulletins. One respondent emphasised that a key challenge for the future was to implement changes to operational products and workflows that better utilise ensemble forecasts while balancing the communication of a clear and concise warning product.
- Others felt the ***development of new tools, products and visualizations*** to allow forecasters to rapidly and efficiently view and interpret ensemble data should be the main goal in the coming years, given the limited time available in operational context. One respondent stressed the importance of integrating ensemble data into the forecast platform, rather than relying on external sources such as internet products.
- ***Collaboration between all NWP centers and forecasting centers*** was requested, to share ensemble data in standard formats, products, verification packages and results, and expertise on interpreting ensemble forecast data and including ensemble information in forecast and warning products.
- ***Training of forecasters on the utilisation of ensembles*** was mentioned by several respondents as essential along with the need for a ***change in operational working practice towards using probabilistic information***. This need is not unique to TC forecasting, as illustrated by a recent survey of meteorologists' understanding of forecast uncertainty and probabilistic forecast guidance at NOAA's Hazardous Weather Testbed (Wilson et al., 2019), which found substantial

variability in the extent to which participants could think beyond a deterministic mindset and successfully synthesize and apply uncertainty guidance information.

- Several users requested a focus on the development of **ensemble-based hazard-based forecasts** e.g. storm surge, coastal inundation, inland flash flood and river flooding, especially where the risk is focussed remotely (away from the center of the cone). This should include research into evaluating and improving the ensemble's ability to resolve the details of the vortex in order to allow a better assessment of the predictability of the associated hazards.
- The **development of impact-based ensemble products** was seen as crucial by a number of respondents, with access to and/or collection of socio-economic data sources being one key issue here.
- **Improvement in the skill of ensemble forecasts of tropical cyclone intensity and structure** was seen as a key priority for some. Improved availability of bias-corrected ensemble forecasts, or methods to do so, was also requested.
- **User-oriented and basin-specific verification comparing ensemble guidance across models** was seen as important to help forecasters gain confidence in using ensemble forecast information. One respondent said it would help if training were available on the performance and skill of all models (deterministic, consensus, and ensemble models) in their basin. Another emphasized the importance of research into how to optimally combine the guidance from all available models (deterministic and ensemble, global and regional).

The results of the questionnaire should help to provide guidance to NWP modelling centers, researchers working in this area, and managers in the operational forecasting centers, on where further development should focus to maximize the potential benefit of ensemble forecasting in operational TC forecasting.

3.3 Conclusions and recommendations

The questionnaire results show that dynamic ensemble forecasts are currently used by nearly all operational TC forecasters and at TC forecast centers around the world, although there is a need to ensure that ensemble forecast information is more widely available to forecasters in developing countries. The survey showed that ensemble forecasts are seen to be particularly important in track and genesis forecasting, but of only limited use in intensity and structure forecasting due to the difficulty of adequately resolving and forecasting these in current global dynamical ensemble models. The responses to several questions point to the ensembles being introduced into a very deterministic-focussed environment, where the ensemble member tracks, or even just the ensemble mean track forecasts, are viewed alongside deterministic forecasts, rather than

forecasters being able to utilise the full probabilistic uncertainty information that is available from ensemble forecasts. There are several examples of excellent pull-through of ensemble forecasts from research into the operational TC forecasting process. However, many forecast centers still use historical error statistics rather than dynamic situation-based uncertainty in determining the width of the uncertainty cone/circle/swath in their operational track warning products. The survey therefore reveals a notable difference between the high proportion of operational TC forecasters who use and value ensemble forecast information, and the slower pull-through into operational warnings and products of the probabilistic guidance and uncertainty information that ensembles can provide.

The survey highlighted several areas of research and development that would help TC forecasters to make increased use of ensemble forecast information in the future: a) an improvement in the skill of the ensemble forecasts (in particular for intensity); b) improved access to ensemble forecast data, tools, visualizations and user-orientated verification; c) better understanding of best practice in ensemble-based forecasting of TCs; d) the development of hazard-based and impact-based ensemble products; and e) a change in operational working practices towards using probabilistic information, and providing and communicating dynamic uncertainty information in operational forecasts and warnings.

The specific recommendations put forward and accepted by IWTC-9 based on the results of the questionnaire were as follows:

- Researchers are encouraged to promote research activities to evaluate and improve probabilistic TC intensity and structure forecasts (e.g. Magnusson et al., 2019).
- Researchers are encouraged to promote research activities to demonstrate the usefulness of dynamical ensemble forecasts for hazard-based products such as for strong wind, precipitation, flooding and storm surge, and the use of ensembles in impact-based forecasting for TCs.
- Greater access to forecast data including deterministic/ensemble forecast data and global/regional forecast data would be beneficial. Research is encouraged to evaluate how best to combine the information from these models, in order to maximize forecast skill and reliability.
- The research community should recognize the value and importance of access to forecast data such as TIGGE and further promote the use of such data, including wider sharing of code for product generation and verification to facilitate the operational use of ensemble forecasts.
- There is a need for greater collaboration both amongst the operational forecast centers, and between operational forecast centers, NWP modelling centers, and researchers, in order to co-ordinate the development of new ensemble-based tools and methods, and develop and share best practice on the use of ensemble forecasts in operational TC forecasting.

- The operational forecasting community are encouraged to facilitate a change in operational working practice towards using ensemble-based uncertainty and probabilistic information in forecasts and warnings, including a focus on how best to communicate uncertainty to the general public.

Appendix 3-1: Survey questions

Introduction

- Organisation (including RSMC/TCWC where applicable)
- Basin of interest [North Indian Ocean, Western North Pacific, Central Pacific, Eastern Pacific, Atlantic, South Indian Ocean, Australian, South Pacific]
- Job role [Operational forecaster, Manager, Researcher, Other]
- Do you currently use ensemble forecast information [Yes, No]

Current use of ensemble forecast information

- How important would you say ensemble forecasts are in your forecasting of tropical cyclones (Not used, Somewhat important, Important, Very important)? [Impact based forecasting, Track forecasting, Hazard forecasting (precipitation, flooding, wind, storm surge), Extra-tropical transition, Intensity forecasting, Tropical cyclone size and/ or structure, Genesis forecasting]
- For each of the main areas of tropical cyclone forecasting (Track, Intensity, Genesis, Hazards), which ensemble forecast models do you use? [NCEP GEFS, ECMWF ENS, MOGREPS-G (UK Met Office global ensemble), JMA's Global Ensemble Prediction System, Canadian Global Ensemble Prediction System, CMA ensemble, KMA ensemble, COAMPS-TC ensemble, FNMOC ensemble, Bureau of Meteorology ensemble, CPTEC ensemble, Meteo-France ensemble, DWD ensemble, Other]
- Do you use multi-model ensemble forecast data (i.e. combination of multiple dynamical ensembles)? [Yes, in full probabilistic multi-model ensemble combination, Yes, but only by comparing one ensemble with another, No]
- Please tick to indicate if you use calibrated ensemble forecasts and for which type of forecast [Track, Intensity, Genesis]
- How do you access ensemble forecast information? [From products sent by global NWP centres, From products created in your centre, From online products from other centres, From the TIGGE archive, From WMO Global Telecommunications System (GTS), Other]
- Which ensemble-based or probabilistic products do you use (Not available or required, Not available but would like to have, Available but don't use, Sometimes use, Regularly use)? [Ensemble mean track, Tracks from ensemble members, Large-scale flow information, Storm

intensity, Strike probability, Tropical cyclone activity (named and forming storms), Tropical Cyclone genesis, Landfall probability, Precipitation, Wind speed, Storm surge, River discharge, Flood risk / inundation, Impact-based products]

- If there is a product that you find particularly helpful please give the product and model details below.
- If there is a particular product or ensemble that you would like to be able to see please give the details below.
- Which of the following statements describe your use of ensemble forecast information in track forecasting? [I mainly use deterministic forecasts, and rarely use ensemble forecasts, I mainly use ensemble forecasts to compare the ensemble mean forecasts to deterministic forecasts, I find the ensemble spread useful to give me the situation-dependent uncertainty, I use the full probabilistic information from the ensemble forecasts in my operational forecasting, I don't find ensemble forecasts very helpful when the spread is very large]
- Which of the following statements describe your use of ensemble forecast information in intensity forecasting? [I mainly use deterministic forecasts, and rarely use ensemble forecasts, I mainly use ensemble forecasts to compare the ensemble mean forecasts to deterministic forecasts, I find the ensemble spread useful to give me the situation-dependent uncertainty, I use the full probabilistic information from the ensemble forecasts in my operational forecasting, I don't find ensemble forecasts very helpful when the spread is very large]
- Please state how important you feel each of the following verification metrics are when evaluating the skill and value in ensemble forecasts (Don't know, Not important, Quite important, Important, Essential) [RMS error of the ensemble mean, Ensemble spread, Reliability diagrams, Economic value plots, ROC diagram / ROC scores, Brier Score / Brier Skill Score, Ranked Probability Score/Skill Score (RPS/RPSS) or Continuous Ranked Probability Score / Skill Score (CRPS/CRPSS), % of tracks falling within an ensemble-defined ellipse/cone, Other (please specify)]
- If you have an idea for a particular "user-orientated" verification metric from ensemble forecasts please detail below.

Communication of forecast uncertainty in operational forecasts and warnings

- Track: In operational track forecasts from your organisation, what forecast lengths are issued?
- Track: With what frequency are forecasts issued?
- Track: How is uncertainty represented? [Cone, Circle, Error swath, Don't know, Other]
- Track: How is the cone/circle/swath defined?

- Track: Please detail how the size of cone/circle/swath is calculated for each storm? e.g. historical forecast error statistics, climatological uncertainty, consensus spread, ensemble spread, ensemble-based uncertainty etc.
- Intensity: In operational intensity forecasts from your organisation, what forecast lengths are issued?
- Intensity: With what frequency are forecasts issued?
- Intensity: Is uncertainty represented? [Yes, No]. If yes, how is uncertainty in intensity forecasts calculated?
- Genesis: In genesis forecasts at your organisation, what forecast lengths are used?
- Genesis: With what frequency are genesis forecasts issued?
- Genesis: How is uncertainty represented in the genesis forecasts? e.g. high/medium/low or probability categories etc.
- Genesis: How is the uncertainty calculated in the genesis forecasts? e.g. climatological, ensemble-based, other methods etc.
- Hazard: In hazard forecasts from your organisation (e.g. precipitation, winds, storm surge, flooding), how is uncertainty represented?
- Hazard: In hazard forecasts from your organisation, how is uncertainty calculated?

Examples of the use of ensemble forecasts

- Please give an example of where probabilistic ensemble forecasts have been successfully integrated into the operational forecasting process.
- Please give an example of where hurdles have prevented probabilistic ensemble forecasts being successfully integrated into the operational forecasting process.
- Do you place more emphasis on the traditional consensus approach to tropical cyclone forecasting (using multiple deterministic forecasts), rather than the use of dynamical ensemble forecast information? [Yes, No]. If you answered yes, please explain why below.

Future use of ensemble forecast information

- Which of the following would help you to make better use of ensemble forecast information in the future (Please rate each idea from 1 (not important) to 5 (very important))? [Greater access to ensemble forecast data, Improvement in the skill of the ensemble forecasts (track), Improvement in the skill of the ensemble forecasts (intensity), Improvement in the forecast resolution of ensemble forecast models, Better understanding of best practice in ensemble-based forecasting of TCs, Improved access to ensemble-based verification compared across models, Improved access to basin-specific ensemble-based verification, The development of new innovative forecasting products based on ensemble forecasts, The development of hazard-

based (precipitation, wind, storm surge, flooding etc.) ensemble products, The development of impact based ensemble products, Changes of operational working practices to allow greater use of probabilistic information, Development of new user-orientated verification metrics, More focus on combining the ensemble forecasts with multiple deterministic forecasts]

- Please give further details of your thoughts when you were giving your answers above e.g. what specific developments would you like to see?
- Please use this space to put any additional feedback on any other topic relating to the ensemble forecasting and verification of tropical cyclones.

Chapter 4 A global evaluation of multi-model ensemble tropical cyclone track probability forecasts

The second objective of the PhD was to evaluate the benefits of multi-model ensemble tropical cyclone track probability forecasts. The paper covering this objective has been published in the journal Quarterly Journal of the Royal Meteorological Society (QJRMS) with the following reference:

Titley, H.A., Bowyer, R.L. and Cloke, H.L. (2020) A global evaluation of multi-model ensemble tropical cyclone track probability forecasts. Quarterly Journal of the Royal Meteorological Society, 146 (726): 531–545. doi:10.1002/qj.3712.

The contributions of the authors of this paper are as follows: H.T. designed the experiment, produced the tropical cyclone datasets, and led the analysis and writing of the manuscript. R.B. ran the verification software and assisted with the analysis and manuscript writing. H.C. assisted with writing the manuscript. Overall, 80% of the writing was undertaken by H.T..

The published article can be found in Appendix A2.

Abstract. At the Met Office, dynamic ensemble forecasts from the Met Office Global and Regional Ensemble Prediction System (MOGREPS-G), the European Centre for Medium-Range Weather Forecasts Ensemble (ECMWF ENS) and National Centers for Environmental Prediction Global Ensemble Forecast System (NCEP GEFS) global ensemble forecast models are post-processed to identify and track tropical cyclones (TCs). The ensemble members from each model are also combined into a 108-member multi-model ensemble. Track probability forecasts are produced for named TCs showing the probability of a location being within 120 km of a named TC at any point in the next 7 days, and also broken down into each 24-hour forecast period. This study presents the verification of these named-storm track probabilities over a two-year period across all global TC basins, and compares the results from basin to basin. The combined multi-model ensemble is found to increase the skill and value of the track probability forecasts over the best-performing individual ensemble (ECMWF ENS), for both overall 7-day track probability forecasts and 24-hour track probabilities. Basin-based and storm-based verification illustrates that the best performing individual ensemble can change from basin to basin and from storm to storm, but that the multi-model ensemble adds skill in every basin, and is also able to match the best performing individual ensemble in terms of overall probabilistic forecast skill in several high-profile case-studies. This study helps to illustrate the potential value and skill to be gained if operational TC forecasting can continue to migrate away from a deterministic-focused forecasting environment to one where the

probabilistic situation-based uncertainty information provided by the dynamic multi-model ensembles can be incorporated into operational forecasts and warnings.

4.1 Introduction

In dynamic ensemble forecasting, instead of making a single forecast of the most likely weather conditions, a forecast model is run multiple times to produce an ensemble of forecasts. These ensemble forecasts take into account uncertainty in the initial conditions and imperfections in the model formulation, and aim to give an indication of the range of possible future states of the atmosphere. For over 25 years, these dynamical ensemble model forecasts have been routinely produced by several global numerical weather prediction modelling centres, including the European Centre for Medium-Range Weather Forecasts Ensemble (ECMWF ENS: Palmer, 2019), the National Centers for Environmental Prediction Global Ensemble Forecast System (NCEP GEFS: Toth and Kalnay, 1997) and the Met Office Global and Regional Ensemble Prediction System (MOGREPS-G: Bowler et al., 2008).

These ensemble forecast models have an important role to play in TC forecasting, through their ability to highlight the situation-dependent uncertainty and provide probabilistic forecast information to help inform decision makers. Consequently, many global modelling centres produce TC track forecasts from their ensemble forecast models, develop ensemble TC forecast products, verify these forecasts, and share the forecast track data via the TIGGE (The International Grand Global Ensemble) cyclone exchange programme (Swinbank et al., 2016). Several studies have shown the benefits of dynamic ensemble forecasting for both existing TCs (e.g. Dupont et al., 2011; Yamaguchi, et al., 2012; Du et al., 2016; Leonardo and Colle, 2017; Zhang and Yu, 2017), and for providing probabilistic information about TC genesis (e.g. Yamaguchi et al., 2015; Yamaguchi and Koide, 2017). However, a recent survey among operational TC forecasters (Titley et al., 2019) showed that although ensemble forecasts are used and valued by almost all forecasters, the deterministic-focussed forecasting environment has often limited the extent to which the full probabilistic information provided by ensembles has been pulled through into operational TC forecasts and warnings.

Objectively identifying the forecast track of each ensemble member is essential both for post-event model evaluation and for the generation of forecast guidance products in real time. Various tracking techniques are used by operational centres around the world (Vitart and Stockdale, 2001; Van der Grijn et al., 2005; Tallapragada et al., 2014) and in research (Hodges, 1995). At the Met Office, a bivariate (850 hPa relative vorticity and mean sea-level pressure) TC tracker known as MOTCTracker (Heming, 2017) is run in real time on the Met Office MOGREPS-G ensemble, and the tracks are

made available to Regional Specialised Met Centres (RSMCs) and via the research TIGGE cyclone CXML archive.

Although ensemble forecasts are a good way of assessing forecast uncertainty, they are limited to the uncertainty captured by a specific modelling system, and there is a tendency for single-model ensembles to be under-spread, with the observations too often falling outside of the range of solutions. A multi-model ensemble approach, where dynamic ensemble systems from multiple centres are combined together into a grand multi-centre ensemble (thereafter called multi-model ensemble), can help address this shortcoming and provide a more complete representation of the uncertainty in the model structure, also potentially reducing the errors. The rationale behind multi-model ensemble forecasting was summarised by Hagedorn et al. (2005) who stated that "the key to the success of the multi-model concept lies in combining independent and skilful models, each with its own strengths and weaknesses". Several studies have demonstrated that probabilistic forecast skill and reliability can be improved through the use of multi-model ensembles, including Park et al. (2008), Johnson and Swinbank (2009), Hagedorn et al. (2012), Hamill (2012) and Matsueda and Nakazawa (2015).

The application of a grand combined multi-model ensemble approach in TC forecasting is a natural extension of the "consensus" forecasting approach that has been a valuable cornerstone of TC track and intensity forecasting for many years, where traditionally three or more deterministic forecasts have been combined or averaged into a "consensus" forecast and also used to give a prediction of likely forecast error (e.g. Goerss, 2000; Goerss, 2007; Sampson et al., 2008; Goerss and Sampson, 2014; Yamaguchi et al., 2017). The process of combining together several ensemble forecast models into a multi-model ensemble combines the strengths of the consensus and ensemble approaches, by pulling through the full probabilistic forecast information from several dynamic ensembles and model formulations into multi-model ensemble probability forecasts.

At the Met Office, in order to produce real-time multi-model ensemble TC products, MOTCTracker is also run in real time on the direct input data from the ECMWF ENS and NCEP GEFS ensembles. The ensemble forecast tracks from the ensemble members from these models are combined with the Met Office MOGREPS-G ensemble tracks, to create a 108-member multi-model ensemble. The three corresponding deterministic models (the Met Office Unified Model, the ECMWF IFS, and the NCEP GFS model) are also tracked. A range of products, including track and intensity forecasts for both named and developing TCs, are produced and used by forecasters in the Met Office Global Guidance Unit. An example of the track and track probability products produced for named TCs from the multi-model ensemble is shown in Figure 4-1 for Typhoon Kong-rey. The products are also distributed to several operational TC forecasting centres, including the RSMCs in Miami and La

Réunion. The tracks and track probabilities from each of the ensembles and the multi-model ensemble are also displayed in the Met Office Global Hazard Map (Robbins and Titley, 2018), where they can be viewed alongside the associated wind and rain hazard information, and overlain on vulnerability and exposure fields.

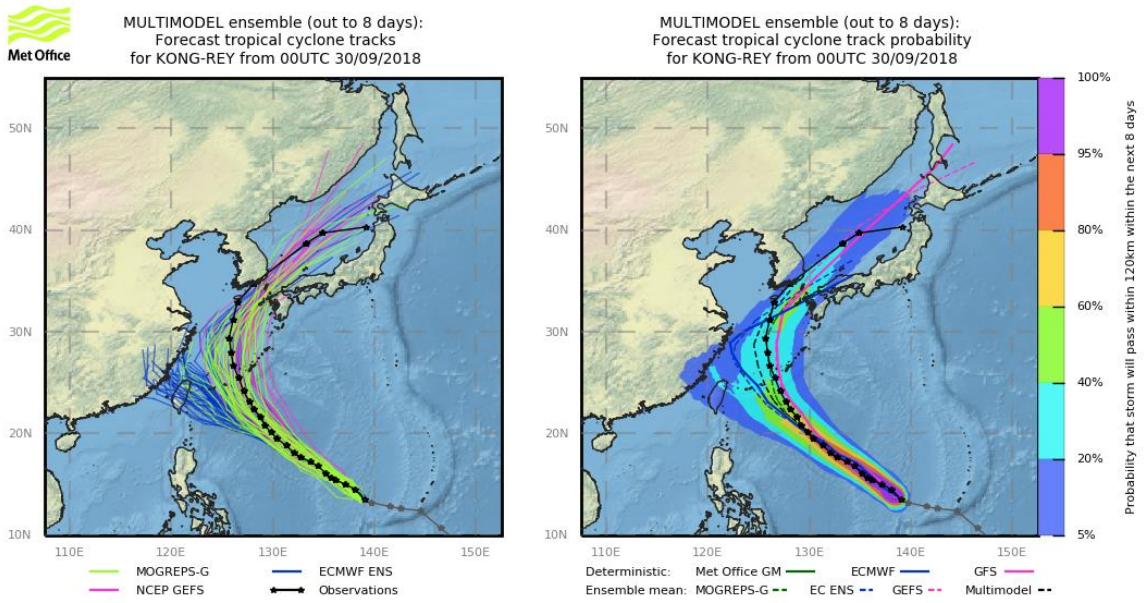


Figure 4-1. Multi-model ensemble forecasts for Typhoon Kong-rey from 00UTC 30/09/2018: Tracks coloured according to model (left), and multi-model ensemble track probability with deterministic (solid) and ensemble mean (dashed) tracks (right).

To fully assess the skill and value of the multi-model ensemble TC forecasts produced at the Met Office, a framework to produce objective verification of named TC track probability forecasts has been developed, the results of which are presented in this chapter. Previous studies evaluating named TC track probability forecasts using multi-model ensembles have focussed on only one or two Northern Hemisphere basins, (e.g. Majumdar and Finocchio, 2010 (North Atlantic and Northwest Pacific); Yamaguchi et al., 2012 (Northwest Pacific); Leonardo and Colle, 2017 (North Atlantic)). In this chapter the verification results are analysed for all named TCs in all TC basins around the globe during 2017 and 2018, and then split into each basin in order to compare the absolute and relative performance of the ensembles across basins. Storm-based verification results have also been calculated for some high-profile TCs in order to compare the results from storm to storm.

The key questions addressed by this chapter are:

- How do forecast performance and characteristics vary between the global ensembles, and with lead time?
- Is there benefit in terms of probabilistic forecast skill, reliability and value, from combining the three global ensembles into a multi-model ensemble?

- Does using the full probability forecast information via the multi-model ensemble add skill compared to a "consensus" forecast of the parent deterministic models?
- When comparing basin to basin, does forecast performance vary, both overall compared to a reference consensus forecast, and relatively between the different ensembles?
- When comparing storm to storm, how does forecast performance vary, and what does it tell us about the benefit of multi-model ensembles?

The verification framework and methodology are described in Section 4.2. Section 4.3 presents the results, split into three sections: (a) 7-day ("overall") track probability forecast results; (b) 24-hour track probability forecast results; and (c) storm-specific verification results for two high-profile North Atlantic hurricanes. Section 4.4 provides a discussion, focussed around the implications of the results for operational forecasters, decision-makers, and model developers, while Section 4.5 states some key conclusions and ideas for future work.

4.2 Verification methodology

The three ensembles tracked in real time at the Met Office using MOTCTracker (Heming, 2017), and included in the multi-model ensemble products, are MOGREPS-G (24 members until 11th July 2017, then 36 members thereafter), ECMWF ENS (51 members) and NCEP GEFS (21 members). The ensemble forecast tracks from the three ensembles are combined to create a 108-member multi-model ensemble (96 members until 11th July 2017), with each individual member from each ensemble equally weighted when creating the track probability forecasts. The track probability forecasts are calculated for named TCs from each of the three global ensembles, and the combined multi-model ensemble, both for the next 7 days (with no lead time component, to directly verify the forecast product shown in Figure 4-1 – henceforth described as the "overall" track probability), and for 24-hour forecast periods out to 7 days (in order to evaluate the change in performance over different lead times). Track probabilities are defined as the probability of the named TC passing within 120km of each grid point in the given time period using the commonly used "strike probability" definition first defined in (Van der Grijn et al., 2005). In this paper the term track probability is preferred, because a TC is a large weather system that can also "strike" and lead to impacts further away than 120km from the centre of the storm track, for example from storm surges, flooding rain and extreme winds including tornadoes. Thus the term "strike probability" could be misleading to the public as it could falsely imply that those areas outside of 120km from the track will not be impacted. Tropical cyclone genesis (storms that are forecast to form during the forecast) is not included in the verification in this paper. Observed track data is collected routinely at the Met Office from Global Telecommunications System (GTS) bulletins from the RSMCs, TC warning centres, and the Joint Typhoon Warning Centre (used if no track is available from the

RSMC), to accumulate observed TC positions for all named storms to verify the forecasts. The mean track or "consensus" of the three deterministic model tracks is calculated and used as the reference forecast to fully assess the forecast skill.

Since 2015 the verification has been produced operationally, biannually, at the end of the northern and southern hemisphere seasons, for TCs in all global basins over the previous 12 months, in order to evaluate the most recent configurations of each ensemble. This paper verifies a 2-year period (January 2017 to December 2018), in order to increase the sample of cases, whilst ensuring the model configurations are still relatively recent.

For each of the named TCs in the study period the following steps are carried out in the overall track probability verification process:

i. All observed positions for this storm are read in and the observations are included from the first time period that it has an observed intensity of greater than 34 knots to the last time that it exceeds this value. This ensures that only TCs that reach tropical storm strength or greater are verified, and that for the included storms, all observations in between these two times are included even if the intensity dips below this threshold in between.
ii. A storm-specific verification grid is created for the area covered by the storm track at a resolution of 0.5 degrees.
iii. For each of the 00/12 UTC forecast run times at which this TC was named, a track probability gridded data file is calculated for the observations, the reference consensus forecast, and each forecast model. For the observation data this is a file containing 1's (for grid points which fall within 120km of the observed track of the storm over the next seven days), and 0's (those that do not). To create the reference consensus forecast, the mean of the three deterministic forecast tracks is calculated, and a file created containing 1's (for grid points which fall within 120km of the consensus track of the storm over the next seven days), and 0's (those that do not). For the ensemble forecast models, each grid point contains a probability value between 0 and 1, calculated by counting how many members have forecast tracks for this storm that lie within 120km of that point at some point between T+0 and T+168, and dividing by the number of ensemble members.
iv. For the forecast model data (consensus and ensemble forecasts), the forecasts are only included up to the last time that a matching observation is available. Therefore this verification is focussed on the ability of the models to predict the future track position of the TC while it is at tropical storm strength or above, rather than on the ability of the models to weaken and dissipate the storm at the same time as shown in the observations, as in practice the point at

which the forecast tracks end is highly dependent on the choice of thresholds for dissipation in the storm tracker.

v. As many forecast runs do not have a full seven days of matching observed tracks to verify against, there are a considerably smaller overall sample size of tracks verifying at day 7 than at day 1. To prevent the overall track probability verification from being dominated by the verification of shorter lead time forecasts, which are not so relevant for society given that important decisions such as evacuations generally need to be made on forecasts of two or more days, only those forecast runs where there are matching observed tracks for T+48 or longer are included.

A similar process is then followed to verify for each 24-hour period from T+0-T+24 through to T+144-T+168. In this case however, the forecasts are verified at the basin-scale, again to replicate the forecast product, which is a basin-based animation of the track probability forecasts for every 24-hour period.

Gridded verification is then carried out comparing the track probability forecast data with the corresponding observed data. A range of probabilistic verification statistics are calculated to assess the skill, reliability and value of the ensemble forecasts. These include the Relative Operating Characteristic (ROC), reliability diagrams, relative economic value and Brier skill score (BSS), as described in Jolliffe and Stephenson (2012), and summarised below:

- The ROC plot assesses the skill of the forecast at discriminating between events and non-events. The points along the curve are the hit rates and false alarm rates for each probability bin. Perfect skill would produce a curve from bottom left to top left to top right, and no skill is indicated by the diagonal line from (0,0) to (1,1).
- Reliability diagrams display how well the predicted probabilities correspond to their observed frequencies. Perfect reliability would be a diagonal line from (0,0) to (1,1), a line above the diagonal indicates under-forecasting and below the diagonal shows over-forecasting, while a line which falls below the diagonal for high probabilities and above it for low probabilities exhibits over-confidence (Wilks, 2011).
- For a given user, their cost-loss ratio is the term given to the ratio of the cost of a preventative measure to the loss averted, and can be used to guide the probability threshold above which to take action. At each probability threshold there will be a 2x2 contingency table containing the number of hits, false alarms, misses and correct rejections. Assuming the user takes action whenever an event is forecast, then a cost (C) can be associated with the hits and false alarms and a loss (L) associated to the misses. The relative economic value can then be calculated for a range of cost-loss ratios and visualised using a relative economic value plot which displays the

relative improvement in economic value between the sample climatology and a perfect forecast for all cost-loss ratios (Richardson, 2000). The relative economic value is a useful additional user-focussed measure for comparing forecasts. In practice the user's cost-loss ratio may be difficult to determine and may have to be estimated, however if one forecast consistently has higher value that another forecast across all cost loss ratios then it clearly has greater value for any user.

- The BSS assesses the relative skill of the probabilistic forecast over that of a reference forecast in terms of predicting whether an event occurred. A score of 0 indicates no skill when compared to the reference forecast and a score of 1 would be a perfect score. In this paper the reference forecasts are the mean or consensus track from the three deterministic forecasts. This is a deliberately challenging reference forecast compared to climatology-based reference forecasts, and the implications of this are discussed in Section 4.4.

Verification statistics are produced for all TCs, and then are also split into six global basins to allow for an inter-region comparison, as shown in Figure 4-2.

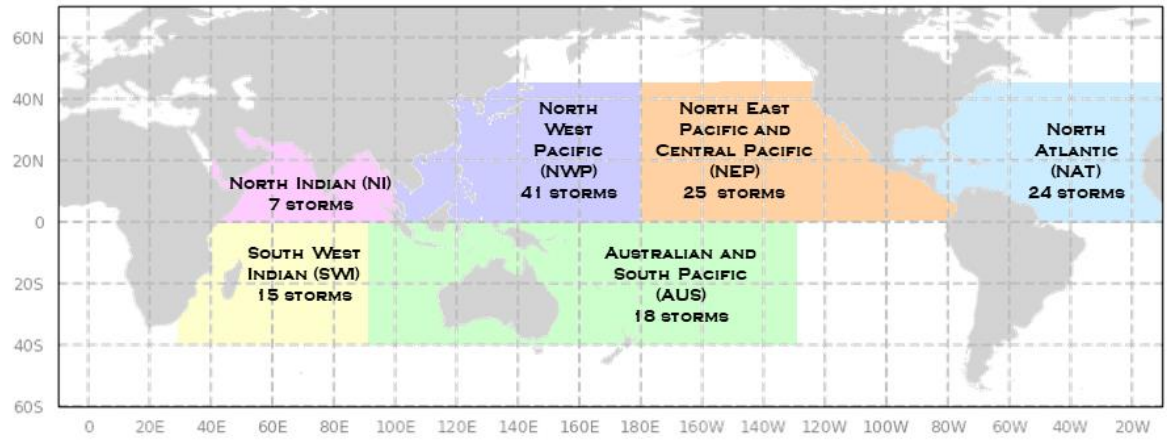


Figure 4-2. The six TC basins verified in this study, and the number of named TCs included in the verification in each basin in the study period (2017 and 2018).

4.3 Results

4.3.1 *Overall track probability forecasts for named tropical cyclones*

During the verification period of January 2017 to December 2018, 130 named TCs across the six basins had at least one forecast run where the verification criteria were met and all forecast data were available (24 in NAT, 25 in NEP, 41 in NWP, 7 in NI, 15 in SWI and 18 in AUS, as shown in Figure 4-2). Overall this two-year period had close to the average number of named TCs per year, with above-average activity in the North Atlantic basin, and below-average activity in the Southern Hemisphere.

The ROC plot in Figure 4-3 shows excellent skill for all models; the multi-model ensemble has the largest ROC area (0.985), followed by ECMWF ENS (0.967), MOGREPS-G (0.956) and NCEP GEFS (0.925). All of the models have very low false alarm rates for the majority of forecast probabilities, while the corresponding hit rates vary more significantly across the forecast probabilities. Figure 4-4 indicates that all models show good reliability, particularly ECMWF ENS and the combined multi-model ensemble. MOGREPS-G and NCEP GEFS are slightly over-confident, with the line falling above the diagonal for low probabilities and below it for higher probabilities. Over-confident reliability diagrams are a common characteristic of ensemble forecasts and are characteristic of under-spread in the ensemble. So when a high track probability is forecast the ensemble member forecast tracks are sometimes too closely clustered and the TC track is less likely than forecast, but when a low probability is forecast the ensemble members are again too closely clustered and so the probability should sometimes be a bit higher.

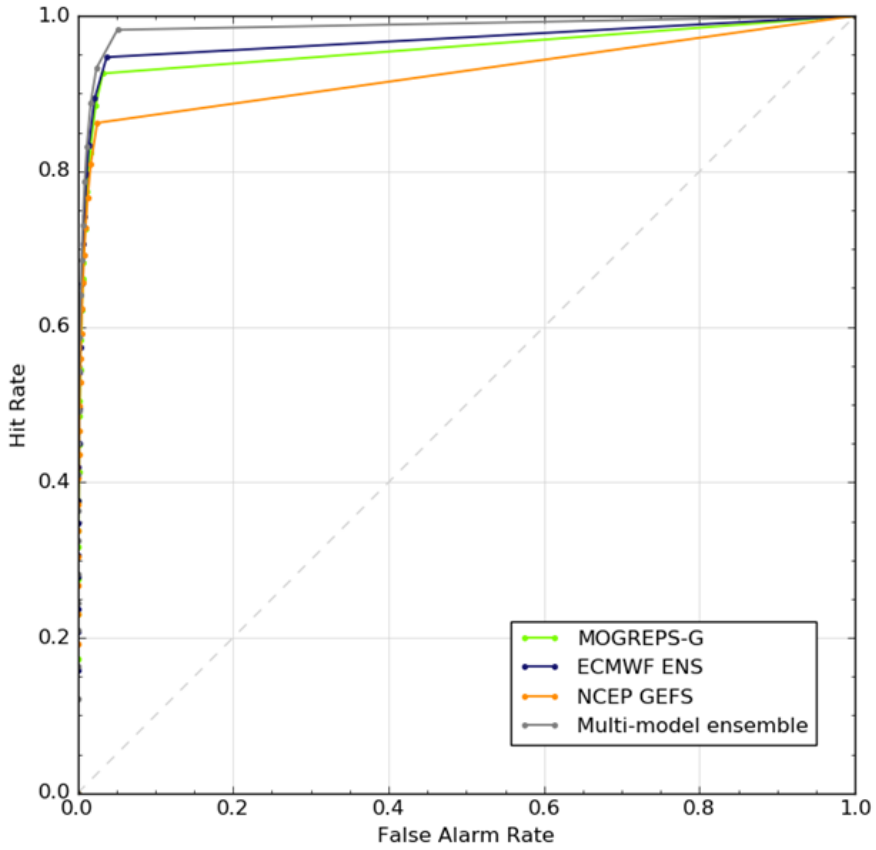


Figure 4-3. ROC plot showing the hit rate and false alarm rate at each probability threshold, for the overall track probability forecasts from the three individual ensembles (MOGREPS-G, ECMWF ENS and NCEP GEFS) and the combined multi-model ensemble.

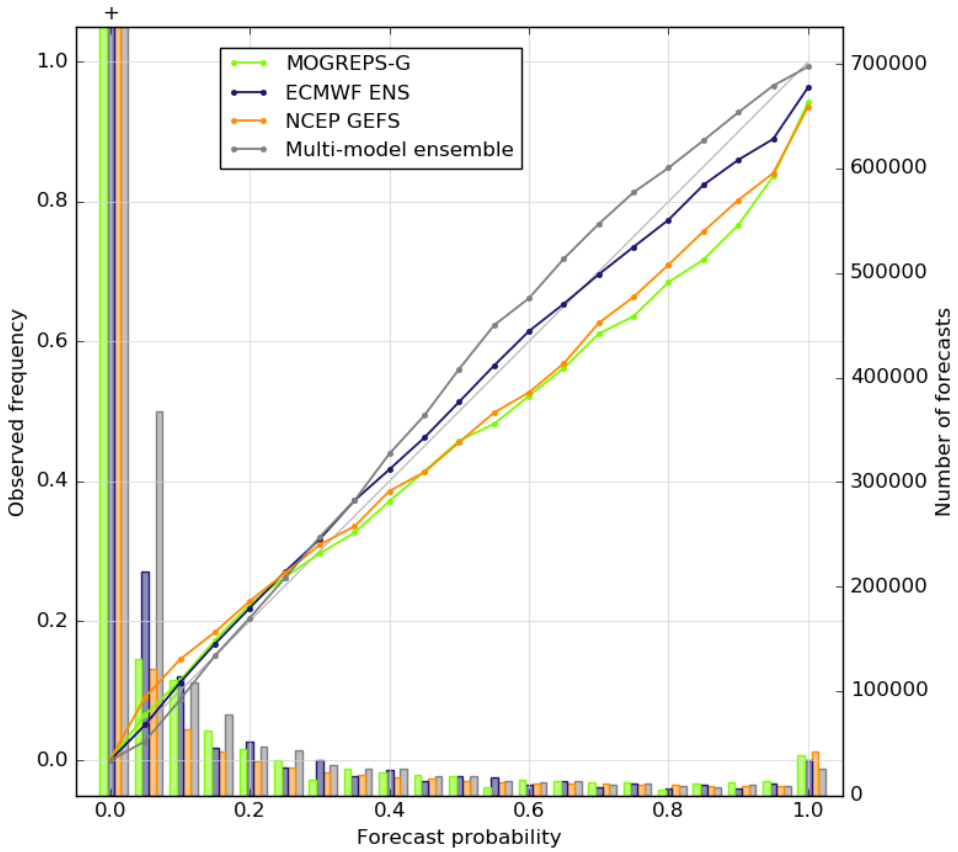


Figure 4-4. Reliability diagram for the overall track probability forecasts for the three individual ensembles (MOGREPS-G, ECMWF ENS and NCEP GEFS) and the combined multi-model ensemble. The number of forecasts (grid points) in each bin is also indicated by bars along the x-axis.

Figure 4-5 demonstrates that the combined multi-model ensemble has the greatest relative economic value for all cost-loss ratios, with the multi-model ensemble curve fully encompassing the curves of the three individual models. All models show the greatest relative economic value for low cost-loss ratios (0-0.2) where the loss associated with an event is significantly greater than the cost of acting to mitigate against it when forecasted. This is a promising result as although the cost-loss ratios vary for different forecast users, TCs tend to be associated with low cost-loss ratios due to the severity of their associated hazards and potential impacts.

The BSS, calculated using the consensus of the deterministic models as the reference forecast, is displayed in Figure 4-6, both for all named TCs, and split into their relevant basin. Although the absolute and relative performance of the individual ensembles varies from basin to basin, the combined multi-model ensemble has the largest BSS across all basins, showing that the multi-model ensemble adds forecast skill in every basin. The strongest performing individual ensemble is ECMWF ENS in all basins except the South West Indian where NCEP GEFS has the largest BSS of the individual ensembles. The relative performance of MOGREPS-G and NCEP GEFS varies between basins.

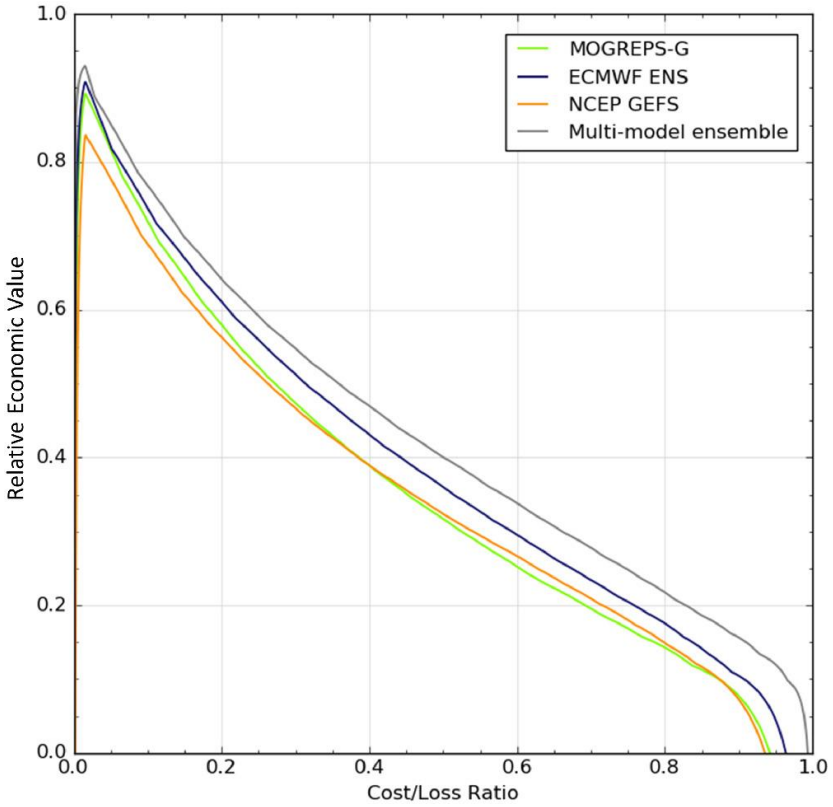


Figure 4-5. Relative economic value plot displaying the relative economic value at each cost-loss ratio for the overall track probability forecasts from the three individual ensembles (MOGREPS-G, ECMWF ENS and NCEP GEFS) and the combined multi-model ensemble.

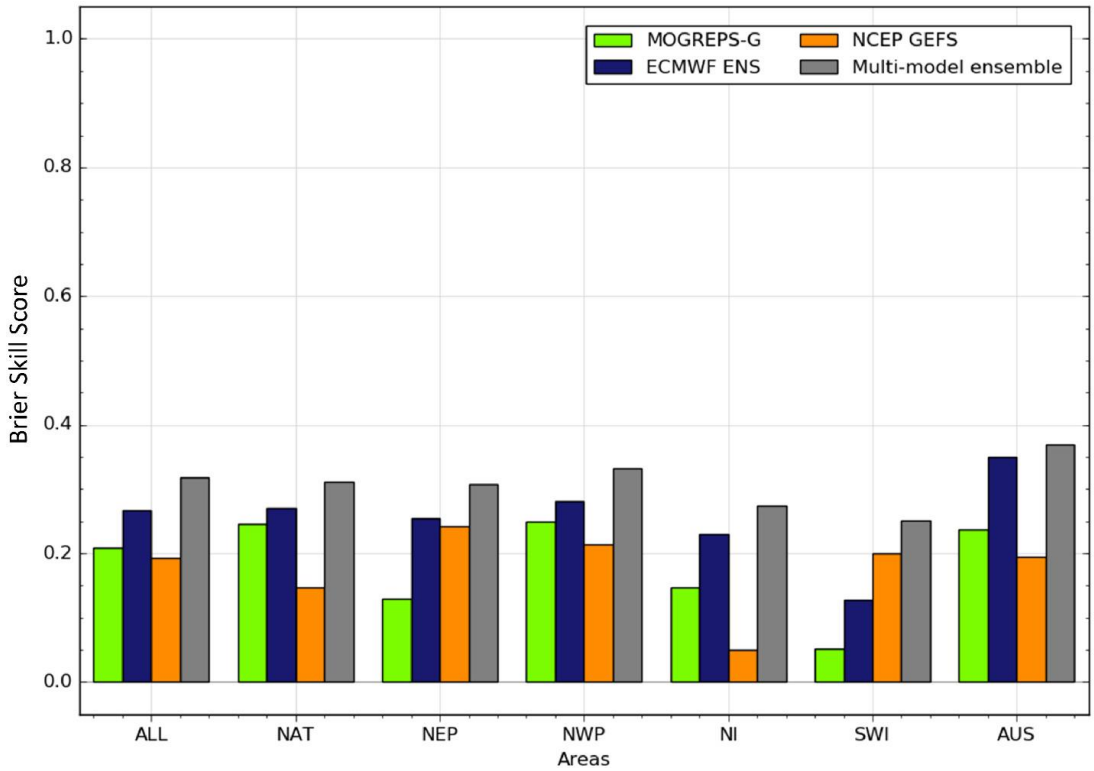


Figure 4-6. BSS for the overall track probability forecasts for all named TCs (left) and for named TCs in each of the individual six basins, for each individual ensemble and the multi-model ensemble. The reference forecast used in the skill score calculation is the consensus (mean) track of the three deterministic forecasts.

4.3.2 *Track probability forecasts for named tropical cyclones over each 24-hour forecast period*

To illustrate the changes in forecast characteristics and performance against lead time, the reliability diagrams and relative economic value plots shown in Figures 4-7 and 4-8 display the 24-hour track probability centred on six lead times from T+24 to T+144. The reliability diagram (Figure 4-7) shows slight over-confidence for all models at T+24 (Figure 4-7a). At the higher probability bins (> 0.5), the advantage of the multi-model ensemble is evident as it maintains good reliability compared to the three individual ensembles, which all show a similar degree of over-confidence. As lead time increases, reliability begins to decrease in the NCEP GEFS and MOGREPS-G ensembles, while ECMWF ENS, and especially the multi-model ensemble, maintain very good reliability particularly at probabilities greater than 0.4 at T+72 and T+96 (Figure 4-7c and Figure 4-7d). MOGREPS-G and NCEP GEFS are showing significant over-confidence by T+120 (Figure 4-7e), indicating that these models are significantly under-dispersive in their five-day track forecasts. At T+120, the multi-model ensemble remains the most reliable at probabilities less than 0.8. This is an important result, as it shows that MOGREPS-G and NCEP GEFS, despite being significantly under-dispersive by this point, are still adding benefit to the ECWMF ENS in the multi-model ensemble at most forecast probabilities. However at this lead time and in particular at T+144 (Figure 4-7f) there are far fewer cases in the higher probability bins, leading to more noisy results.

In Figure 4-8, nearly all cost-loss ratios (0–0.9) have relative economic value at T+24 (Figure 4-8a). The peak value is at low cost-loss ratios, with approximately 0.9 relative economic value for a cost-loss ratio of 0–0.1. As lead time increases there is a gradual drop in value, particularly in the higher cost-loss ratios, but even by T+144 significant value still remains at 0–0.1 cost-loss ratios (Figure 4-8f). At all lead times the multi-model ensemble value curve fully encompasses those for the individual ensembles. Out to T+72 (Figure 4-8c) the value of the multi-model ensemble is significantly greater than any of the individual ensembles, but by T+120 (Figure 4-8e) the value for the ECMWF ENS is noticeably greater than the other two ensembles, and closer to the value shown by the multi-model ensemble.

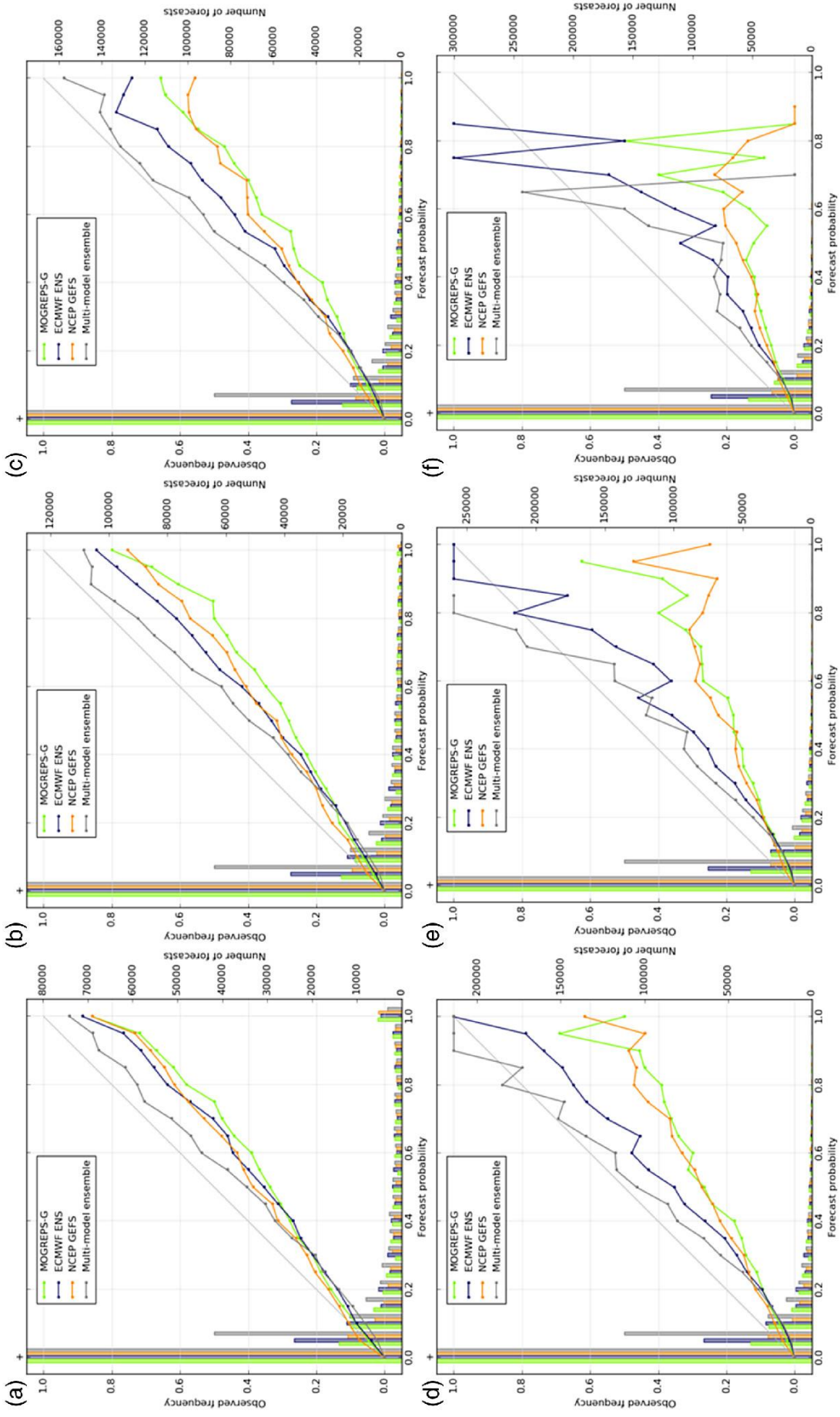


Figure 4-7. Reliability diagrams for 24-hour track probability forecasts for all named storms for each individual ensemble and the multi-model ensemble, for 24-hour periods centred on (a) T+24, (b) T+48, (c) T+72, (d) T+96, (e) T+120 and (f) T+144

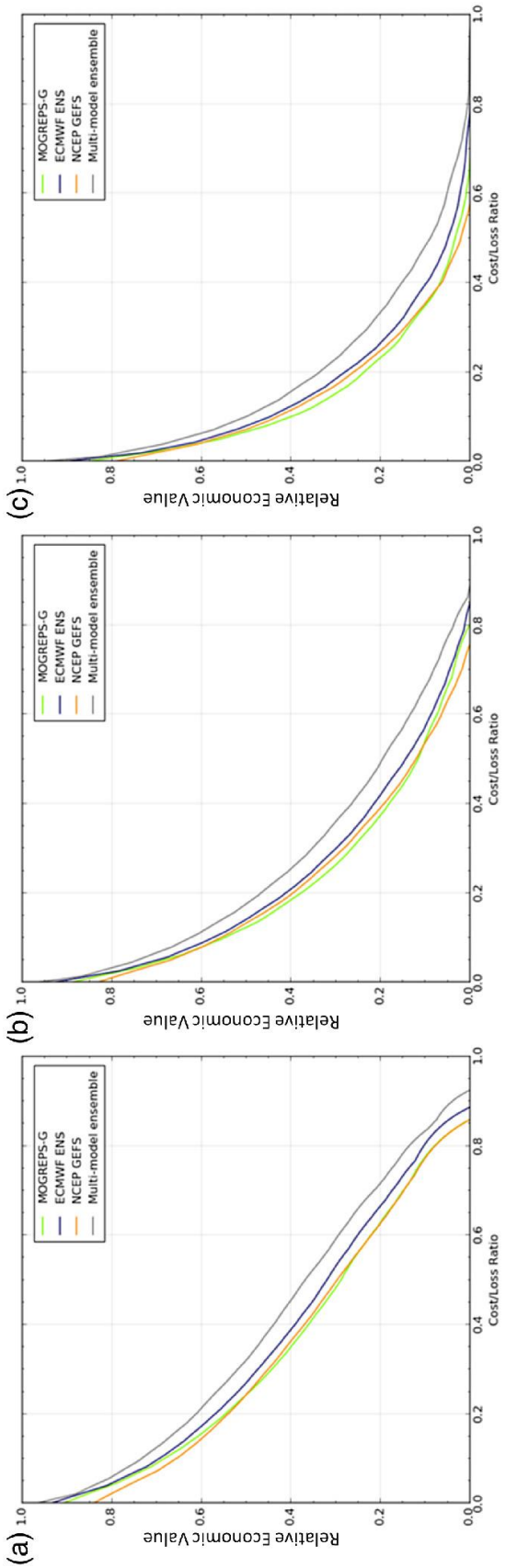


Figure 4-8. Relative economic value curves for 24-hour track probability forecasts for all named storms for each individual ensemble and the multi-model ensemble, for 24-hour periods centred on (a) T+24, (b) T+48, (c) T+72, (d) T+96, (e) T+120 and (f) T+144

The area under the ROC curve (AUC) values are shown against lead time in Figure 4-9. The AUC decreases with lead time for all models. The multi-model ensemble maintains an AUC of over 0.93 at all lead times, whereas the AUC score for the individual ensemble decreases more significantly with lead time (in particular for NCEP GEFS). The Brier skill score (BSS) results in Figure 4-10 show that the skill of the 24-hour track probability forecasts varies with lead time in each basin, with the skill compared to the deterministic consensus forecast rising with increasing lead time. This is as expected, and shows that the benefit of probabilistic forecasts over a consensus of deterministic forecasts becomes greater as the forecast lead time increases. When all TCs are included, irrespective of the basin (the light green line in Figure 4-10), the multi-model ensemble probabilistic forecasts show positive skill over the multi-model consensus from T+60 onwards. However, this hides big differences between the basins. In the North Atlantic and North East Pacific basins the BSS is positive throughout and rises to a value of 0.3 by T+156. In the Northwest Pacific basin the BSS is positive from T+48 onwards. However, in the North Indian Ocean, the Southwest Indian Ocean and the Australian basins, the BSS stays negative throughout, showing that in these basins the ensembles are not adding skill to the consensus reference forecast for the 24-hour track probability forecasts in this study period. The negative BSS for these basins in their 24-hour track probability forecasts is in contrast to the positive BSS in their overall 7-day track probability forecasts that was shown in Figure 4-6, and highlights that the 7-day verification is more forgiving of along-track errors. It is also worth noting that these are also the three basins where the sample size of storms is the smallest. A case-by-case assessment of the TCs in the worst-performing SWI basin reveals that the deterministic errors were relatively low compared to previous seasons, making the consensus reference forecast hard to beat for this sample of storms.

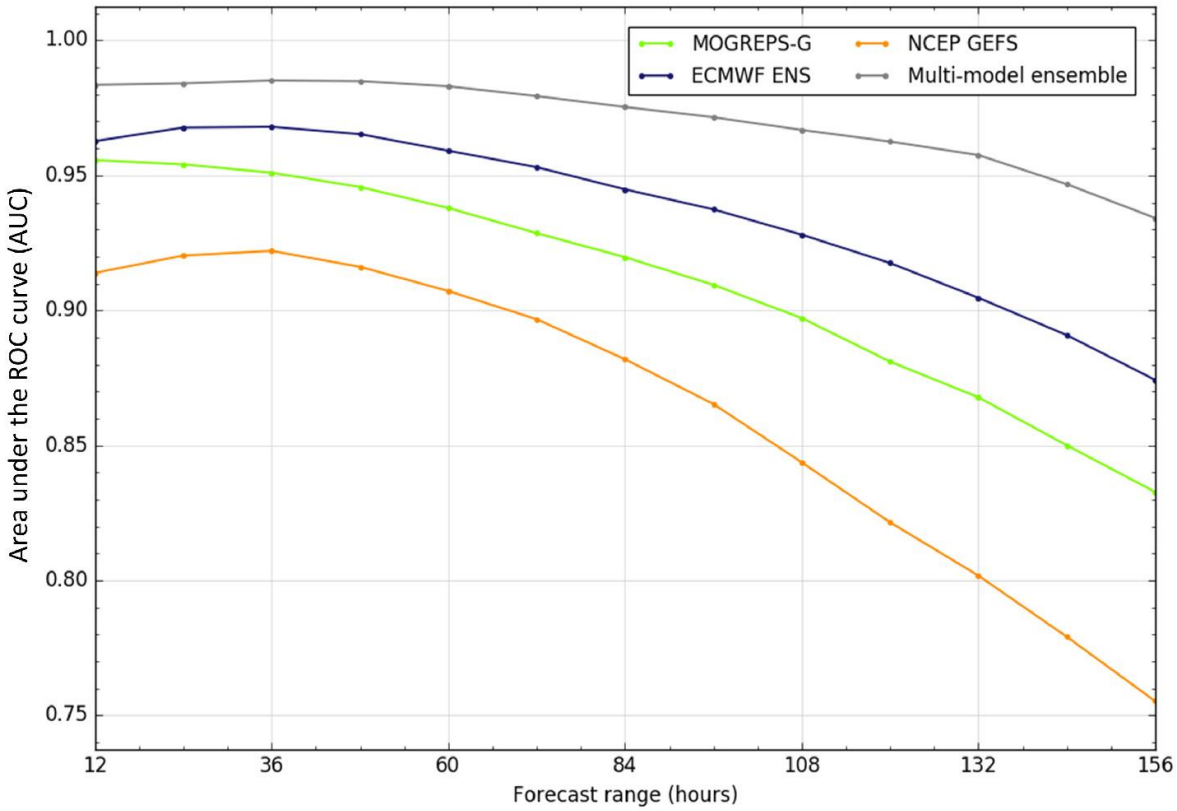


Figure 4-9. The AUC for each individual ensemble and the multi-model ensemble, for the 24-hour track probability forecasts centred on each forecast range.

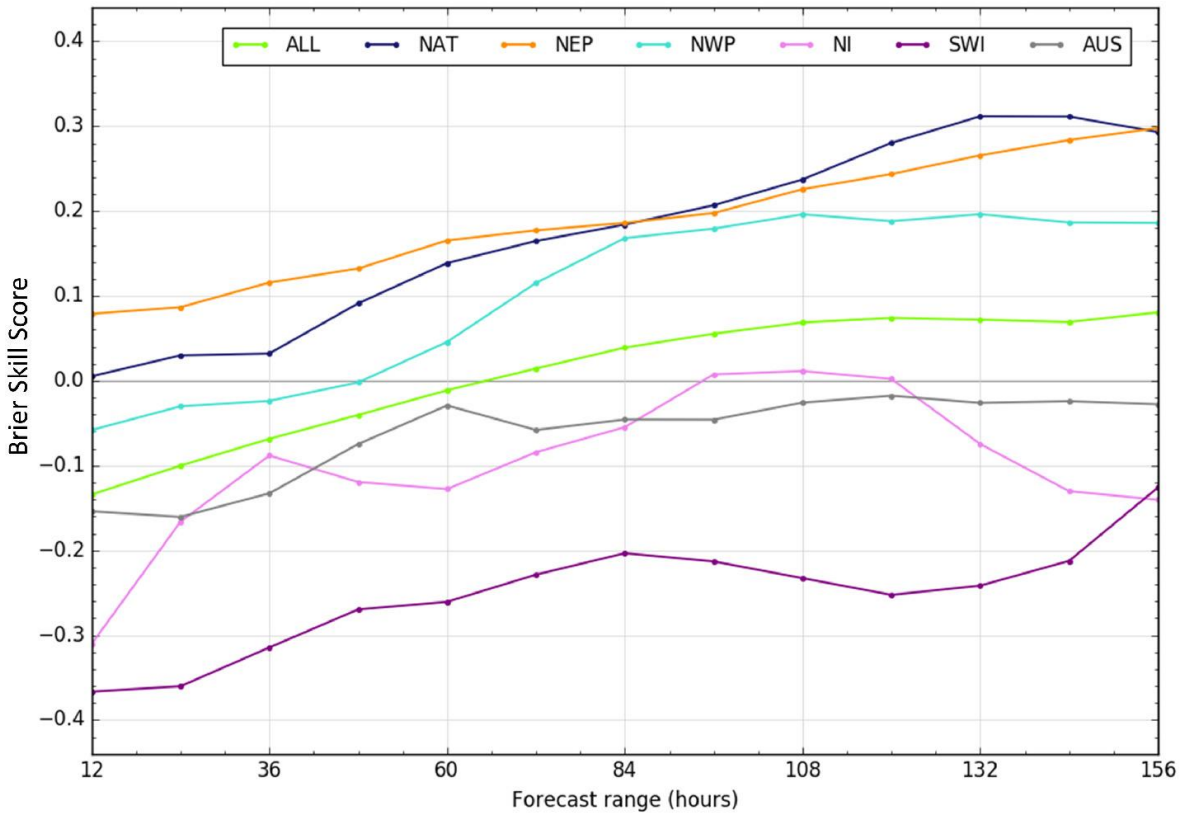


Figure 4-10. BSS for the multi-model ensemble, for all storms (in green) and split into the six TC basins, for the 24-hour track probability forecasts centred around each forecast range.

4.3.3 *Verification for two high-profile tropical cyclones*

Figure 4-11 compares the BSS for the storm-based verification of two high-profile TCs in the North Atlantic basin: Hurricane Matthew in 2016 and Hurricane Irma in 2017. For each storm a different individual ensemble displays the highest skill (MOGREPS-G for Matthew and ECMWF ENS for Irma), illustrating that even within the same basin, the strongest performing individual ensemble varies from storm to storm, rather than one ensemble always being the most skilful in a given basin. In both cases, the multi-model ensemble shows comparable forecast skill to the strongest performing model. Figure 4-12 shows one of the forecasts included in the storm-based verification in each case, illustrating the strong performance of MOGREPS-G for Matthew and ECMWF ENS for Irma. For Hurricane Matthew, the MOGREPS-G ensemble was the first model to give a strong signal for the storm to track just off the Florida coast with an eventual landfall in South Carolina. For Hurricane Irma, the ECMWF ensemble was the only ensemble to contain the observed track in the ensemble track spread in the early forecast runs. At the time of the forecast it is not known which of the individual ensemble forecast models will have the greatest forecast skill for that particular storm, and so the key result here is that in both cases the multi-model ensemble probability forecasts were able to provide equivalent forecast skill to the best performing individual ensemble forecasts.

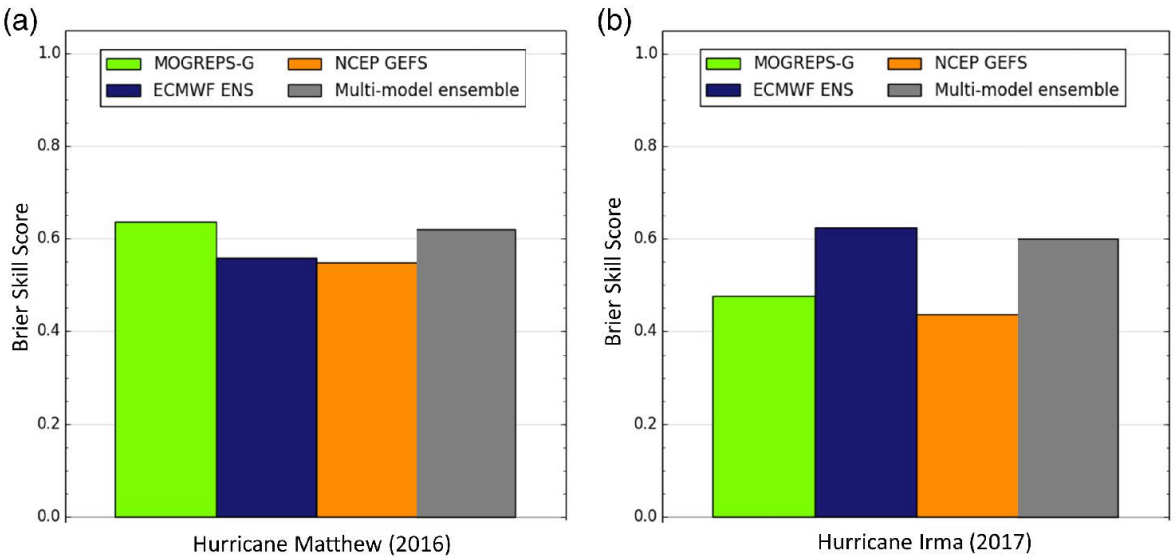


Figure 4-11. Example storm-specific verification for all forecast runs of (a) Hurricane Matthew (2016) and (b) Hurricane Irma (2017): Brier skill score for MOGREPS-G, ECMWF ENS, NCEP GEFS and multi-model ensemble forecasts (reference forecast in this verification is the sample climatology).

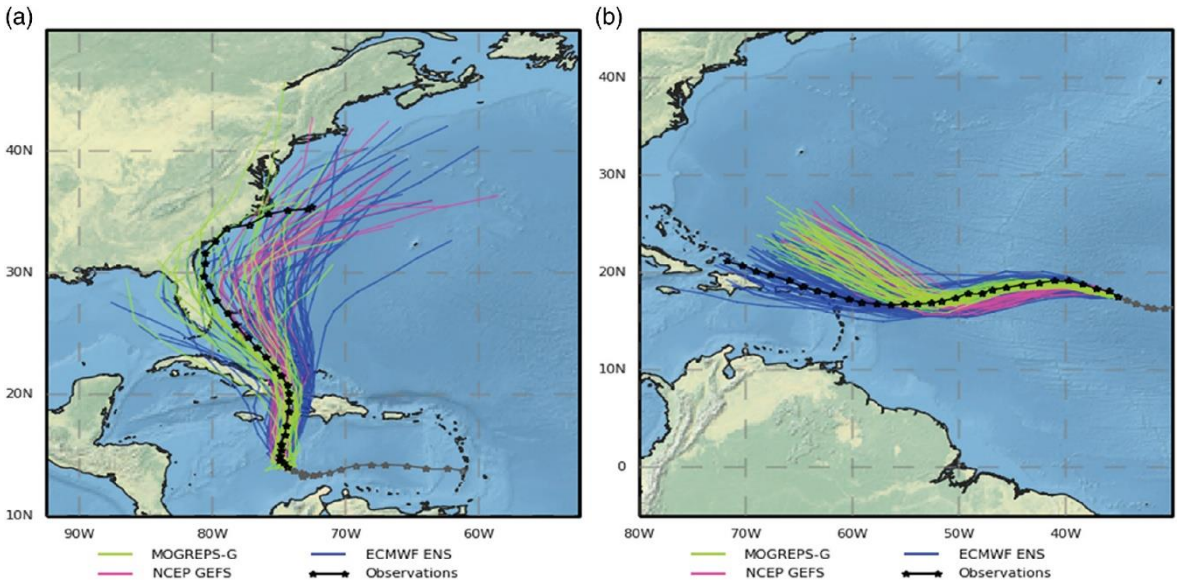


Figure 4-12. Ensemble member tracks from one of the forecast runs included in the verification of each case: (a) 1200 UTC 2 October 2016 forecast for Hurricane Matthew; (b) 0000 UTC 1 September 2017 forecast for Hurricane Irma. Tracks are coloured according to model (green = MOGREPS-G, blue = ECMWF ENS and pink = NCEP GEFS).

4.4 Discussion

In Section 4-1, five key questions were laid out to be addressed by the probabilistic evaluation of TC track probability forecasts in this study. This discussion section is organised to answer these questions, and draw out the key implications from these:

- *How do forecast performance and characteristics vary between the global ensembles, and with lead time?*

All ensembles exhibit good reliability and value in the named TC track probability forecasts, particularly at low cost-ratios. This shows how the probability forecasts have huge potential to be useful to decision makers and downstream users of TC forecasts, who will often have low cost-loss ratios due to high potential losses to property and personal safety, compared to the cost of mitigative efforts such as putting up shutters or evacuating. It highlights the importance of initiatives to increase the pull-through of probabilistic situation-based uncertainty information into operational warnings, such as the collaboration through the WMO HIWeather project that is discussed in Titley et al. (2019).

The best performing of the three individual ensembles included in the study, in terms of the verification statistics presented for TC track probability forecasts, is the ECMWF ENS, followed by MOGREPS-G and NCEP GEFS, which are both over-confident in their track probability forecasts, indicating the known tendency of these models to be under-dispersive. There are many differences between the three ensemble forecast systems, including differences in data assimilation influencing

both the initial conditions of the storm itself and the wider environmental steering flow, the model formulation, and the ensemble perturbation strategies. The relative contribution of perturbations from an ensemble of data assimilations, singular vectors, and stochastic model perturbations to ECMWF ENS track spread was presented in (Lang et al., 2012). Benefits of the ECMWF ENS perturbation strategy include the ability to target singular vectors on TCs, and an enhanced ability to be able to tune perturbations to give improved spread and reliability, compared with the perturbation schemes in the other ensembles.

In the verification of the track probability forecasts for each 24-hour period, all models were shown to be over-confident in their track probability forecasts, but this became much more pronounced at longer lead times for MOGREPS-G and NCEP GEFS, where the higher probabilities were forecast far too often compared to the observed frequency. These results are important to model developers as they show the importance of increasing the ensemble spread in the MOGREPS-G and NCEP GEFS ensembles. In MOGREPS-G, a major upgrade was scheduled to go live in autumn 2019 to try to address this issue. The ensemble perturbation system will be changed from Ensemble Transform Kalman Filter (ETKF) to an ensemble of data assimilations (En-4D-En-Var: Bowler et al., 2017). In the new system, data assimilation is performed for each member, creating increments relative to its own background trajectory. A partial re-centring around the deterministic analysis gives an additional increase in skill and reduces jumpiness. Comparative trials of the new En-4D-En-Var ensemble have shown faster spread growth across many variables including 850 hPa wind speed in the tropics, with a much better match to observed errors. The ensemble trials have also been processed through the TC tracking and post-processing system, and show a significant improvement in track spread at all lead times. Meanwhile at NCEP a new version of the deterministic Global Forecast System (GFS) with a new Finite-Volume Cubed-Sphere Dynamical Core (FV3) went operational in June 2019. The potential of this new model to improve hurricane forecast performance was described by Chen et al. (2019), who showed the improved performance in re-runs of the active 2017 Atlantic hurricane season. The FV3 dynamical core will be implemented into the GEFS ensemble in 2020, in an upgrade that will also see improved stochastic physics and an increase in ensemble members. The ECMWF ENS also continues to be upgraded, with IFS Cycle 46r1 implemented in June 2019, including a data assimilation upgrade which improves the initial conditions of the ensemble forecasts. Additional work to improve the TC intensity forecasts in the ECMWF ENS is described in Magnusson et al. (2019). The track probability verification described in this chapter will continue to run every 6 months to verify the three ensemble models and their combined multi-model ensemble to investigate the impact of these model upgrades on the skill of TC track probability forecasts.

- *Is there benefit in terms of probabilistic forecast skill, reliability and value, from combining the three global ensembles into a multi-model ensemble?*

As established in the introduction, the rationale of multi-model ensemble forecasting lies in combining independent and skilful models, each with its own strengths and weaknesses. The evaluation in this chapter clearly shows that additional forecast skill and value can be gained from combining the members from the three individual ensembles included in this study into a multi-model ensemble. The three ensembles have different data assimilation strategies, model formulations and ensemble perturbation schemes, that when combined together are shown to collectively provide more realistic estimates of TC track probabilities. In the overall track probability verification, the relative economic value curve of the multi-model ensemble fully encompasses that of the individual ensembles, and in the verification for each 24-hour period it is particularly noteworthy that the AUC decreases more significantly with lead time in the individual ensembles compared to the multi-model ensemble. This illustrates that the amount of potential forecast skill to be gained from using a multi-model ensemble to derive your track probabilities increases significantly with forecast lead time. This finding is important for operational TC forecasters, forecast centre managers, and numerical weather prediction centres, showing the importance of improving access to multiple ensemble forecast model forecasts in order to allow multi-model ensemble information to be used in the operational TC forecasting process.

- *Does using the full probability forecast information via the multi-model ensemble add skill compared to a "consensus" forecast of the parent deterministic models?*

The choice of reference forecast is crucial in the BSS calculation, as the BSS measures the improvement of the probabilistic forecast relative to a reference forecast. This forecast reference is often calculated from the sample climatology, or, in the case of TC forecasts, from a combined climatology and persistence forecast (CLIPER: Knaff et al., 2003). However, in order to be more relevant for the current common forecasting practice of establishing a consensus forecast, and in order to provide a more skilful and challenging reference for the ensemble probabilities to be compared to, a consensus forecast based on the mean track of the three parent deterministic models was created and used as the reference in the BSS calculation. The BSS for the overall 7-day track probability skill scores shows positive skill for all of the ensemble forecast models in all basins when compared to the reference consensus track, with the BSS being highest for the combined multi-model ensemble. This shows the added benefit to be gained in operational TC track forecasting if the full probability information provided by the ensembles can be pulled through into operational warnings, as has begun to happen with the incorporation of dynamic uncertainty information in warning products from several centres including RSMC La Réunion and RSMC Tokyo

(Titley et al., 2019). When the track probability verification is split into 24-hour periods it is harder to beat the consensus reference, particularly at short lead times, but the BSS increases with forecast range, showing the increasing value of ensemble prediction over consensus forecasting with lead time. Overall there is positive forecast skill from T+60 onwards. The BSS was also calculated for the 24-hour track probability verification using the sample climatology (not shown) and showed the reverse pattern, with the highest BSS (0.6) at shorter lead times, slowly decreasing with forecast range. This is to be expected since the sample climatology reference has no forecast capability, whereas the consensus has a very strong forecast capability at short range which reduces with increasing lead time. This illustrates how important it is to be clear about the implications and rationale of selecting a particular reference forecast.

- *When comparing basin to basin, does forecast performance vary, both overall compared to a reference consensus forecast, and relatively between the different ensembles?*

In the overall track probability forecasts BSS results (Figure 4-6), the combined multi-model ensemble has the largest BSS of all the models, and varies between 0.25 and 0.4 with each basin, showing that the multi-model ensemble is adding skill in every basin. The strongest performing individual ensemble is ECMWF ENS in all basins except the SWI basin where NCEP GEFS has the highest skill. The relative performance of MOGREPS-G and NCEP GEFS varies between basins, with MOGREPS-G performing well in the NAT, NWP and AUS basins, and NCEP GEFS performing well in the SWI and NEP basins. This result would provide useful guidance to operational TC forecasters in these regions, as it provides a breakdown of the current levels of performance of each model in their area of interest, and confirms that all areas would see additional value in computing track probability forecasts from multi-model ensemble forecast data. It is also of interest to model developers, who could investigate why their model performs better relatively for TCs in one basin over another, leading to potential ideas on how to improve forecast skill. As discussed earlier, each model has different perturbation strategies and data assimilation schemes, with differences in which observations are included (including differences in quality-control systems and in the assimilation of RSMC TC observations, known as bogussing, which is not currently carried out at ECMWF). These results show that some aspects of each model system may be better suited to one region over another, and further investigations including the implications of a TC being in a data-rich vs. data-poor area may lead to further insights and improvements in the model forecasts.

The inter-basin comparison of BSS using the 24-hour track probability forecasts centred on each lead time (Figure 4-10) reveals large differences from basin to basin in the skill compared to the consensus reference forecast, with positive skill in the NAT, NWP and NEP, but not in the NI, SWI

or AUS basins. This emphasises that the traditional deterministic consensus is hard to beat in regions where the tracks of the named storms in the study period are well forecast by the deterministic models. The lower BSS for the 24-hour track probability forecasts highlights the need to focus on improving the ensemble forecasts for the translation speed of the storm, as the results when splitting the track probability forecasts in to 24-hour periods will be impacted by along-track errors in addition to the cross-track errors, whereas the cross-track errors are the most influential factor in the overall 7-day track probability results. Although forecast users predominantly need to know whether or not they will be impacted by a hurricane rather than when, the timings are also important in forecast preparedness activities, and future verification could investigate the relative along- and cross-track errors in the ensemble vs. the deterministic consensus.

- *When comparing storm to storm, how does forecast performance vary, and what does it tell us about the benefit of multi-model ensembles?*

Although when averaged across all TCs the ECMWF ENS is clearly currently the most skilful of the individual ensembles, the case-study analysis in Figure 4-11 showed that in a particular case this will not always hold true. Sometimes a different model is more skilful in a particular storm (as in the case for MOGREPS-G for Hurricane Matthew). At the time of the live forecasts, operational forecasters will not know which individual model is destined to be the most skilful for that storm, and therefore the result showing that the multi-model ensemble skill matches that of the strongest performing individual ensemble is an important result, and clearly illustrates the benefit of combining the ensemble members into a multi-model ensemble when computing track probability forecasts.

4.5 Conclusions and future plans

This study has shown that combined multi-model ensemble TC track probability forecasts, calculated from all members of ECMWF ENS, MOGREPS-G and NCEP GEFS, have increased skill and value over the best-performing individual ensemble. This result is consistent when verifying all global named TCs together, and when the verification is carried out for each individual basin. The improved skill and value of the multi-model ensemble is found for both the full (up to 7-day) track probability forecasts, and for track probabilities split into 24-hour forecast periods. The verification results from the three individual ensembles show that the track probability forecasts from ECMWF ENS display the best reliability, skill and value. MOGREPS-G and NCEP GEFS become increasingly over-confident and under-dispersive with increasing forecast range, emphasising the importance of ongoing work at both centres to improve the perturbation strategy and increase the spread in the ensemble forecasts. However, even when there is an individual ensemble model that on

average performs better in a particular storm, basin, or overall, there is still skill, reliability and value to be gained from adding additional ensemble models into a combined multi-model ensemble, indicating that they have independent systematic strengths and errors and collectively provide more realistic estimates of TC track probabilities than individually.

Storm-based verification illustrated that the best-performing individual ensemble can change from case to case, but that the multi-model ensemble matches the best-performing individual ensemble, which would not be known in advance, in terms of overall probabilistic forecast skill. The mean, or consensus, of the three higher-resolution deterministic forecasts is hard to beat in some basins, but overall the additional probabilistic skill of the ensembles is shown, particularly at longer lead times or when computing the full 7-day track probability forecasts. This study helps to illustrate the potential value and skill to be gained if operational TC forecasting can continue to migrate away from a deterministic-focussed forecasting environment to one where the probabilistic situation-based uncertainty information provided by the ensembles can be pulled through into operational forecasts and warnings.

There are many ideas for where this work could be further applied and extended in the future. In this study no weighting is applied to the ensemble members in the multi-model ensemble, with each member from each model given an equal weight. It would be interesting to look at different options for combining together the ensemble members, and also investigate how much of the additional value is from additional members as opposed to the inclusion of members from other models. For example, would the multi-model ensemble still add value to the ECMWF ENS if it were restricted to the same number of members? Additional global ensemble forecast models, including those available in the TIGGE cyclone CXML archive from the Japan Meteorological Agency, the Canadian Meteorological Centre and Météo-France, could be added to the multi-model ensemble, to investigate the optimal combination of ensembles in track probability forecasts. At the Met Office the ensemble TC products and verification will continue to be used to evaluate important model upgrades and trials, such as the impact of the forthcoming move from ETKF to Ens-4D-En-Var perturbations. The products and verification are also run on high-resolution convective-permitting ensembles in Southeast Asia. Work is also underway to extend the ensemble TC verification capability at the Met Office to incorporate a verification of forming storms (TC genesis) and intensity trends. It is also important to look beyond the traditional TC track and intensity forecasts and move towards verifying the associated hazards, for example to assess how the uncertainty and predictability of the track translates through to uncertainty and predictability of the precipitation, and the downstream flood hazard. Ongoing collaborations between global numerical weather prediction centres, researchers and operational forecasting centres continue to

be essential to ensure that future research and ensemble model developments are of maximum benefit to operational TC forecasting.

Chapter 5 Key factors influencing the severity of fluvial flood hazard from tropical cyclones

The third objective of the PhD was to identify the key factors that influence flood-related hazards from tropical cyclones using observations and reanalyses. The paper covering this objective has been published in the Journal of Hydrometeorology (JHM) with the following reference:

Titley, H.A., Cloke, H.L., Harrigan, S., Pappenberger, F., Prudhomme, C., Robbins, J.C., Stephens, E.M. and Zsótér, E. (2021) Key factors influencing the severity of fluvial flood hazard from tropical cyclones. Journal of Hydrometeorology, 22 (7): 1801–1817. doi:10.1175/JHM-D-20-0250.1.

The contributions of the authors of this paper are as follows: H.T. designed the study with advice from H.C., E.S. and F.P.. H.T. obtained the datasets, with assistance from S.H. and E.Z. for the ERA5 and GloFAS data. H.T. carried out the analysis and led the writing of the manuscript. All authors assisted with writing the manuscript. Overall, 85% of the writing was undertaken by H.T..

The published article can be found in Appendix A3.

Abstract. Knowledge of the key drivers of the severity of river flooding from tropical cyclones (TCs) is vital for emergency preparedness and disaster risk reduction activities. This global study examines landfalling TCs in the decade from 2010 to 2019, to identify those characteristics that influence whether a storm has an increased flood hazard. The highest positive correlations are found between flood severity and the total precipitation associated with the TC. Significant negative correlations are found between flood severity and the translation speed of the TC, indicating that slower moving storms, that rain over an area for longer, tend to have higher flood severity. Larger and more intense TCs increase the likelihood of having a larger area affected by severe flooding but not its duration or magnitude, and it is found that the fluvial flood hazard can be severe in all intensity categories of TC, including those of tropical storm strength. Catchment characteristics such as antecedent soil moisture and slope also play a role in modulating flood severity, and severe flooding is more likely in cases where multiple drivers are present. The improved knowledge of the key drivers of fluvial flooding in TCs can help to inform research priorities to help with flood early warning, such as increasing the focus on translation speed in model evaluation and impact-based forecasting.

Significance statement. Knowing ahead of landfall which tropical cyclones are likely to lead to significant river flooding will help those responsible for emergency planning make appropriate

decisions to minimize loss of life and property. We compare 280 tropical cyclones and find that the cases with slow-moving, large, and intense cyclones, affecting areas with wet antecedent conditions, have the highest likelihood of experiencing widespread flooding. Slower-moving storms also have an increased risk of longer and more extreme floods. Our results show the importance of considering aspects such as the speed of forward movement along the whole flood early warning chain, from model evaluation and development, through to warning design and communication, to better inform forecast-based action prior to tropical cyclone landfall.

5.1 Introduction

Considering fluvial flood hazards in tropical cyclone (TC) forecasting and warning is important because this is a leading cause of mortality and damages (Rezapour and Baldock, 2014). In the U.S., drowning from excessive rainfall occurs in more TCs than deaths from any other cause (Rappaport, 2014). Many of these fatalities occur outside of landfall counties (Czajkowski and Kennedy, 2010) and in inland counties (Rappaport, 2000). U.S. residential losses from TC freshwater flooding are twice as high compared to TC storm surge losses, with nearly half of these being in inland areas (Czajkowski et al., 2017). A multi-hazard approach considering both wind-speed and rainfall has been shown to be more appropriate for risk-informed decision-making (Song et al., 2020), but studies investigating evacuation decision-making during hurricanes in the U.S. have shown that the key determining factor is the intensity of the storm on the Saffir Simpson scale based on wind speed (Whitehead et al., 2000), with no significant relationship between the perceived risk of flooding and evacuation (Stein et al., 2010). Therefore, it is important to increase public awareness of the dangers of inland flooding, and provide a better understanding of those factors that influence the severity of flood hazard to those involved in emergency preparedness and disaster risk reduction activities.

Heavy rainfall can present a risk to communities regardless of the storm's intensity category. The U.S. National Hurricane Center (NHC) and the U.S. National Weather Service (NWS) both indicate the importance of other information beyond the storm's intensity-based category. The NHC state that "Rainfall amounts are not directly related to the strength of tropical cyclones but rather to the speed and size of the storm, as well as the geography of the area." (https://www.nhc.noaa.gov/prepare/hazards.php#rain). The NWS state that "It is common to think the stronger the storm the greater the potential for flooding. However, this is not always the case. A weak, slow moving tropical storm can cause more damage due to flooding than a more powerful fast-moving hurricane." (https://www.weather.gov/jetstream/tc_hazards). The total rainfall from a TC in a given location is largely determined by the length of time a TC spends over that location,

which is dependent on the size of the rainfall area and translational speed of the storm (Rogers et al., 2009). A slow along-track motion, or stalling near or after landfall, can lead to higher amounts of rainfall and a greater flood hazard, as seen during Hurricanes Harvey (2017) and Florence (2018). Larger TCs can also increase the rainfall hazard as they precipitate upon one spot for a longer time than smaller TCs moving at the same speed. They can also lead to more widespread flooding, leading to increased challenges for responders. Forecasting rainfall induced by TC landfall is determined by many factors, including the TC track, intensity, size and structure, as well as interaction with topography (Qiu et al., 2019), vertical wind shear (Chen et al., 2006), and other meteorological systems in the wider atmospheric environment. This can lead to seemingly similar landfall locations having different rainfall distributions (Cheung et al., 2018; Shi et al., 2017). Hydrological factors, such as soil conditions and orography, are also thought to have been important in determining which TC cases had elevated impacts from flooding (Rappaport, 2000; Sturdevant-Rees et al., 2001).

While there is broad understanding, largely through case study or regional analysis (Saha et al., 2015; Hernández Ayala and Matyas, 2016; Touma et al., 2019; Yu et al., 2017), of those factors influencing the rainfall related to TCs, systematic global analysis to objectively confirm these drivers and their relationships to downstream flood hazard is lacking. This study aims to address this gap and help provide more specific information in support of efforts to raise awareness of fluvial flood risk from TCs, by undertaking a systematic global analysis of the key meteorological and hydrological factors that lead to an increased fluvial flood hazard from TCs.

An improved understanding of the key factors that influence the severity of flood hazard from TCs is important to guide research work aiming to understand the predictability of fluvial flooding from TCs, to inform research priorities to improve the forecasts of flooding, and to guide planning and preparedness activities in the event of an advancing TC. Information on which TCs are likely to have storm surge and winds as their main hazard and which will also have substantial flood hazard is of vital importance for disaster risk reduction. While the strongest winds and largest storm surge usually occur near the center of intense TCs, rainfall and flooding often occur far from the center, spread far inland, and last beyond when the cyclone has weakened or dissipated (Villarini et al., 2014a; Khouakhi et al., 2017). This has important consequences for evacuation, emergency management and recovery programs. If planners and responders better understand the likely severity of fluvial flooding for a given storm, and the locations likely to be at greatest risk from flooding, this can help with evacuation and emergency planning, response and recovery efforts.

This study develops three indices to represent the severity of fluvial flooding for TC cases in terms of the flood area, duration and magnitude. These three flood severity indices are compared against

TC and catchment characteristics in 280 landfalling TCs in the 10-year period from 2010 to 2019 in order to identify the key factors that influence the severity of river flooding in TCs.

The remainder of this paper is set out as follows: Section 4.2 details the data sources used in the study, and Section 4.3 describes the methods used to select the TC cases, and calculate the TC characteristics, catchment characteristics and flood severity metrics for each case. Section 4.4 then compares the flood severity metrics for each storm with the TC and catchment characteristics to analyze the key factors that influence the severity of flood hazard from TCs. Section 4.5 discusses some important aspects of the results and where future work would be beneficial, while Section 4.6 contains the main conclusions of the study.

5.2 Data sources

5.2.1 *IBTrACS*

The main observed TC track dataset used in this study is the International Best Track Archive for Climate Stewardship (IBTrACS) (Knapp et al., 2010, 2018). Where IBTrACS data are still listed as provisional (the 2018/2019 and 2019/2020 southern hemisphere seasons, and the 2019 season in the North Indian and North West Pacific basins), the track points are supplemented by the initial positions in the real-time advisories from each Regional Specialized Met Center (RSMC), which are collated at the Met Office for use in verifying TC track forecasts (Heming, 2017; Titley et al., 2020).

The IBTrACS data are used to calculate the land footprint for each TC case (section 5.3.1), over which to calculate the flood severity indices. The IBTrACS data also provide the TC intensity data (Section 5.3.2.1), and is used to calculate the translation speed (Section 5.3.2.2).

5.2.2 *Global Precipitation Measurement (GPM) IMERG*

Precipitation data are taken from the latest Integrated Multi-satellitE Retrievals for GPM (GPM IMERG Final Run V06) (Huffman et al., 2019). The “Final Run” data are used as this includes gauge data where available to calibrate the rainfall satellite-derived observations, and has been shown to capture TC precipitation patterns well, with closer agreement with gauge-based measurements than its predecessor for extreme events (Yuan et al., 2021). The GPM data, with a horizontal resolution of 0.1˚, are extracted at a 30-minute temporal resolution and then accumulated to give 24-hour precipitation accumulation data. The precipitation summary metrics are described in Section 5.3.2.4.

5.2.3 *ERA5*

ERA5 is ECMWF’s latest comprehensive atmospheric reanalysis with a global horizontal grid resolution of 0.28125° (about 31km) (Hersbach et al., 2020). ERA5 mean sea level pressure (MSLP)

data were extracted from the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) (Hersbach et al., 2018a) at 6-hourly intervals and used to calculate the size of the TC (Section 5.3.2.3). Daily ERA5 soil moisture content data were also extracted, along with soil type, for use in calculating the antecedent soil moisture saturation (Section 5.3.3.1). Precipitation accumulations were extracted hourly and summed to give 24-hour precipitation totals to compare to GPM precipitation accumulations. Several studies have found large improvements in performance for precipitation in ERA5 compared to ERA-Interim (e.g. Beck et al., 2019; Tarek, et al., 2020; Nogueira, 2020), including for TC cases (Hersbach et al., 2018b). A recent study found that ERA5 agrees well with gridded gauge data in terms of the spatial distribution of typhoon precipitation, although it can underestimate the most extreme precipitation (Jiang et al., 2021).

5.2.4 *GloFAS-ERA5 reanalysis*

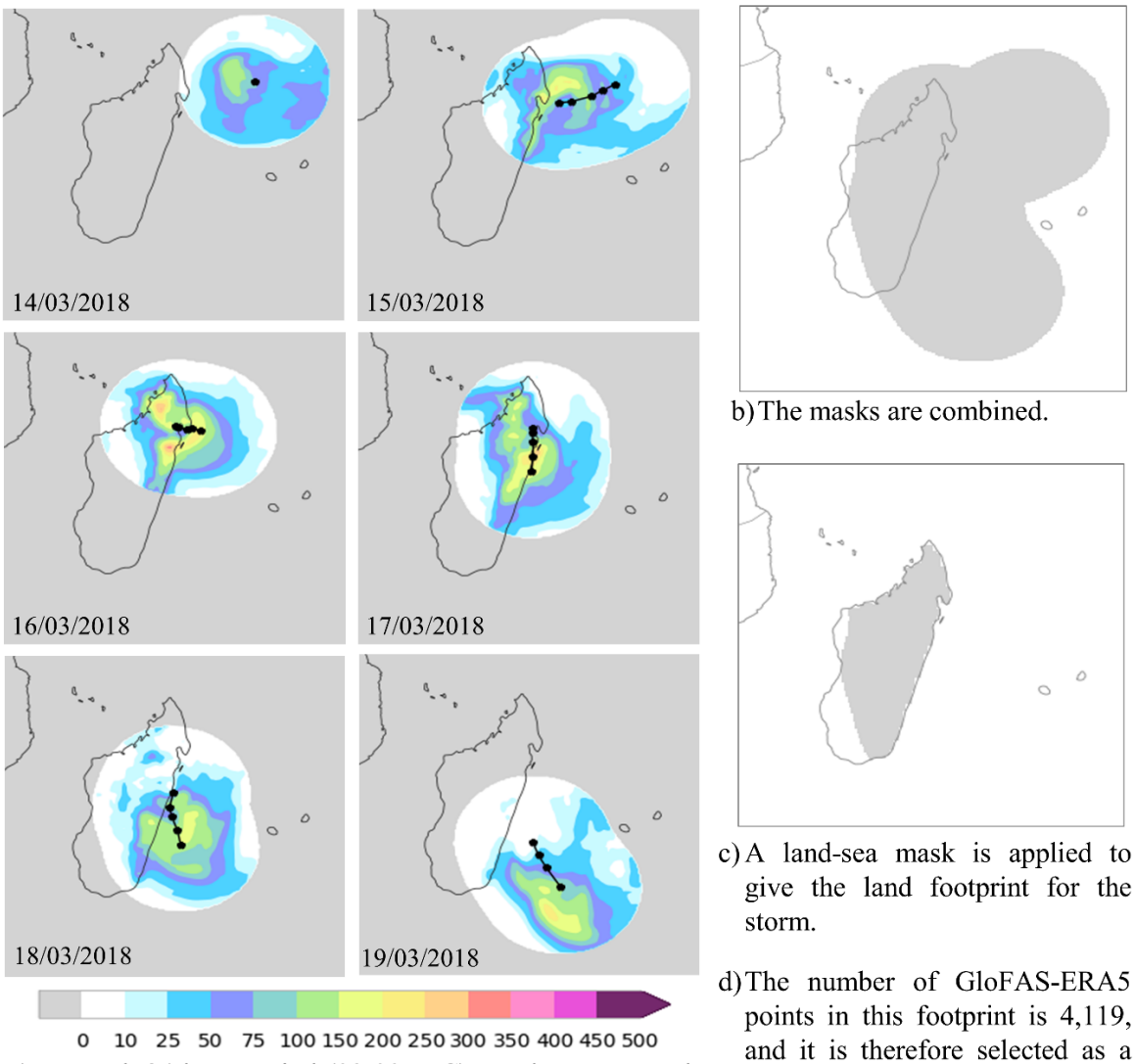
The severity of fluvial flooding from TCs is difficult to calculate in a consistent way from observations given the relative sparsity of river discharge observations, particularly in many areas impacted by TCs (Lavers et al., 2019). Therefore, this study uses the GloFAS-ERA5 global river discharge reanalysis (Harrigan et al., 2020), as a proxy for river discharge observations. The Global Flood Awareness System (GloFAS) is designed to provide a global overview of upcoming flood events to decision makers such as humanitarian organizations (e.g. Coughlan de Perez et al., 2016) as part of the Copernicus Emergency Management Service for floods. In GloFAS, ensemble meteorological forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) are processed by the revised land surface hydrology scheme (HTESSEL) to create land surface runoff fields, with additional hydrological processes such as flow routing provided by the LISFLOOD hydrological model in order to forecast river discharge (Harrigan et al., 2023; Alfieri et al., 2013; Balsamo et al., 2011; Van Der Knijff et al., 2010; Hirpa et al., 2018). Daily GloFAS-ERA5 river discharge reanalysis data, with a horizontal resolution of 0.1°, were extracted from the C3S CDS (Harrigan et al., 2019), for each day 2010-2019, and used to define the flood severity in each landfalling TC case (Section 5.3.4).

Although GloFAS-ERA5 is partially dependent on model-derived precipitation and soil moisture, the groundwater and river routing parameters were calibrated against daily river discharge from 1287 observation stations worldwide (Hirpa et al., 2018) and a recent comprehensive global evaluation found it was skillful against a mean flow benchmark in 86% of catchments (Harrigan et al., 2020). Other studies have compared GloFAS-ERA5 and observed discharge across regions (Ficchì and Stephens, 2019; Towner et al., 2021), or in case studies e.g. Emerton et al. (2020), for Cyclones Idai and Kenneth in Mozambique. The use of GloFAS-ERA5 reanalysis data allows an analysis of the flooding across the whole area affected by each TC, rather than just at isolated points on the river

network. It also allows a direct comparison of flood severity across all TC cases, using openly available data, in a way that would not be possible using discharge observations.

5.3 Methods

5.3.1 *Selecting tropical cyclone cases and calculating their land footprint*



a) For each 24-hour period (00-00UTC), masks are created for the area within 500km of that day's storm track segment. This plot shows these areas in white, overlain with 24-hour ERA5 precipitation accumulations.

b) The masks are combined.

c) A land-sea mask is applied to give the land footprint for the storm.

d) The number of GloFAS-ERA5 points in this footprint is 4,119, and it is therefore selected as a case to be included in this study.

Figure 5-1. Method for calculating the land footprint of each TC case, illustrated for Cyclone Eliakim, which impacted Madagascar in March 2018.

For the period 2010-2019, the IBTrACS TC track data are filtered to only include those TCs with a landfall at tropical storm strength or above (>=34 knots). The 6-hourly latitude and longitude track position data are extracted where the storm is at or above tropical storm strength, and is over land or within 500km of land. In order to decide the area, henceforth termed the 'footprint', in which to

calculate the flood severity for each TC, the following steps are taken (illustrated for Cyclone Eliakim, which impacted Madagascar in 2018, in Figure 5-1):

a) The TC positions at the start and the end of each day are extracted and used to create a daily mask representing the area of precipitation defined as being associated with the TC, by calculating all grid points that lie within 500km of that 24-hour storm track segment. Most studies that have assigned a set radius for defining TC-related rainfall, have used a radius of either 500km (Prat and Nelson, 2013, 2016; Jiang et al., 2011; Luitel et al., 2018), 550km (Zhou and Matyas, 2017) or 5 degrees (Guo et al., 2017).
b) The daily masks are combined to create an overall mask for the track of each TC.
c) The overall mask is combined with a land sea mask to create the final land footprint used to analyze the flood severity.
d) The number of valid 0.1˚ by 0.1˚ points (the GloFAS-ERA5 reanalysis resolution) in the land footprint is calculated, and the small number of TCs with fewer than 1,000 points in their footprint (equating to an area of around 100,000km2) are excluded to ensure the footprints are large enough to allow a sensible comparison of the area affected by severe flooding.

This process results in 280 landfalling TC cases from 2010-2019 being included in the study (broken down by basin in Figure 5-2). For each case a set of TC characteristics, catchment characteristics and flood severity measures are calculated, to allow an analysis across the cases of the key factors influencing the severity of fluvial flooding.

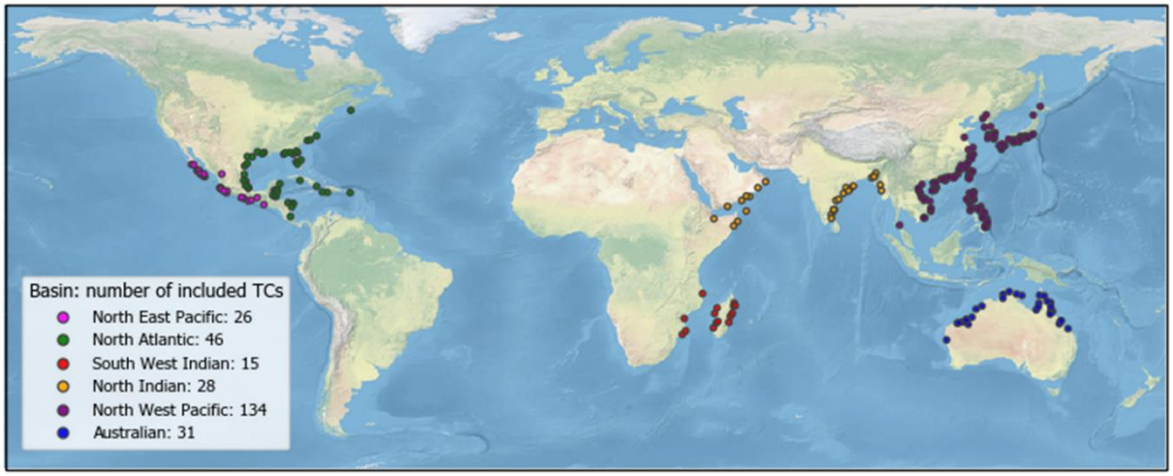


Figure 5-2. The locations of the first land point in the track of the 280 TCs included in this study (2010-2019), including the breakdown of cases by the TC basin. NB. No TCs are included in the South Pacific basin, where the only landfalls are in small island nations and therefore the land footprints were too small to allow for a fair comparison with the other cases.

5.3.2 *Tropical cyclone characteristics*

5.3.2.1 Intensity

The minimum mean sea level pressure (MSLP) and maximum sustained wind speed are extracted from IBTrACS at each track point and the pre-landfall value is used to identify the storm intensity for consideration in each case. To ensure wind speed data are comparable between basins, 1-minute maximum sustained winds are used where available.

5.3.2.2 Translation speed

Translation speed (speed in a forward motion) is calculated for each 6-hour track segment by dividing the distance between the track points by 6 hours. Three speed measures are then taken forward to consider in the analysis: i) translation speed during the track segment where it makes landfall; ii) average speed while the storm track is over land (the mean of the translation speeds from all 6-hourly track segments over land); iii) minimum translation speed while the storm is over land (the minimum of the translation speeds from all 6-hourly track segments over land), to take account of storms that stagnate after landfall.

5.3.2.3 Size and shape

TCs span a huge range of sizes, with gale force winds extending anywhere from around 20km from the center to over 1000km. A number of methods are commonly used to calculate storm size and shape, including the radial extent of winds reaching certain threshold values in knots (R43, R50, R64), the radius of the maximum winds (RMW), and the radius of the outermost closed isobar (ROCI) (Merrill, 1984; Weber et al., 2014). The ROCI delimits the outermost extent of a TC's wind circulation and is the most relevant of these metrics for this study. Each RSMC will have its own, often subjective, procedures to calculate the ROCI recorded in IBTrACS, so to allow global comparisons this study re-calculates the ROCI using ERA5 MSLP fields along the TC tracks. For each track point the MSLP field is centered on the storm location, and then decomposed into tangential and radial co-ordinates, essentially unwrapping the field around the storm center. A search is carried out for the outermost closed MSLP isobar on each radial angle. The distance to this isobar is then averaged for all radial angles to give the overall ROCI, and averaged for each quadrant (left-front, right-front, right-rear and left-rear) to give shape information. Storm asymmetry is calculated by the ratio of the quadrant with the highest ROCI to that with the lowest ROCI. The average size and asymmetry in the day prior to landfall in each case is used in the subsequent analysis.

5.3.2.4 Precipitation

The daily masks containing all points within 500km of the storm track in each 24-hour period (Figure 5-1a) are applied to the 24-hour precipitation accumulation data from GPM and ERA5, to obtain the precipitation considered to be associated with the TC. Three subsequent days are also included using the final mask to ensure that precipitation from any slow-moving remnants of the storm that

may still contribute to flooding are included in the total storm precipitation. For the day prior to landfall, the proportion of the masked area with precipitation accumulations greater than 100mm is calculated and included as a characteristic, to see how this information, which could be better estimated ahead of landfall, compares to flood severity.

To calculate the total storm precipitation, the masked daily precipitation fields are combined to give the overall storm-total GPM and ERA5 precipitation in the land footprint (illustrated for Cyclone Eliakim in Figure 5-3). To investigate a range of precipitation characteristics in terms of their influence on the flood severity metrics, the following values are calculated from both GPM and ERA5:

- The highest precipitation accumulation at any point in the land footprint area
- The average precipitation accumulation across the land footprint area
- The average precipitation accumulation for points in the land footprint area where the total is greater than 25mm (giving an estimate of the average precipitation accumulated in those areas where the storm led to rainfall)
- The proportion of the land footprint area with precipitation accumulations greater than 25/100/200mm

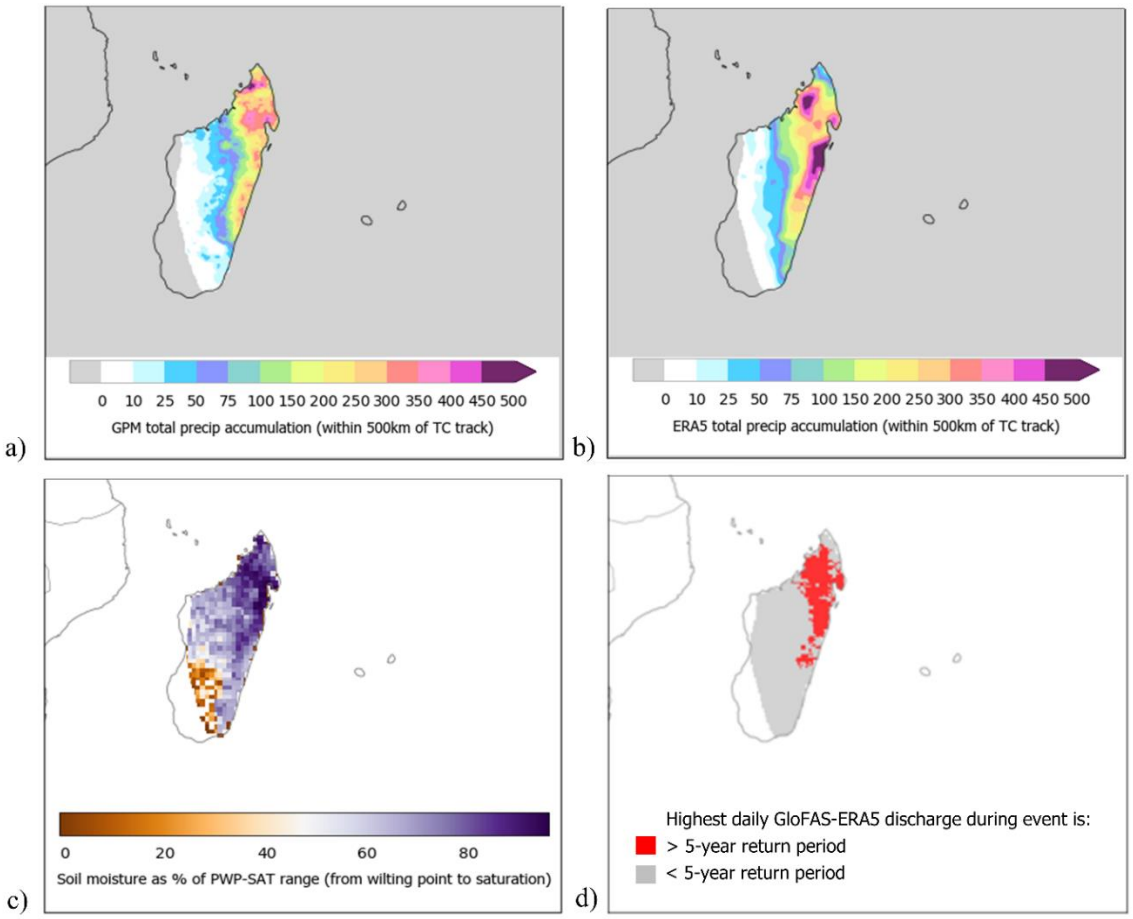


Figure 5-3. Cyclone Eliakim (2018): a) Total event precipitation calculated from GPM; b) Total event precipitation calculated from ERA5; c) Antecedent soil moisture saturation in the top layer of soil; d) GloFAS-ERA5 points that exceeded the 5-year return period during the event (22% of land footprint).

5.3.3 *Catchment characteristics*

5.3.3.1 Antecedent soil moisture saturation

ERA5-derived soil moisture saturation fields were calculated as a percentage of the progress along a scale from the permanent wilting point soil water content to the saturation soil water content (as specified by soil type in table 8.9 of ECMWF (2020a). The data were calculated on two levels: layer 1 (0-7 cm), and layers 1-3 combined (0-100 cm). For each TC case, the antecedent soil moisture saturation data were taken from the date prior to the storm coming within 500km of land, to ensure the pre-storm saturation levels were being recorded. The antecedent soil moisture saturation for the Cyclone Eliakim case is shown in Figure 5-3c. For each layer, the average value in the storm land footprint was calculated, along with the proportion of the footprint with saturation greater than 90%.

5.3.3.2 Orography and gradient

Two of the GloFAS input variables, detailing the orography and the gradient of each grid point, were extracted for each footprint area. The average orography and gradient within the storm footprint area were calculated for consideration in the data analysis.

5.3.4 *Calculating flood severity*

The period used to calculate the flood severity for each TC case was defined by setting the first date as the landfall date and the last date as a week after the final point in the cyclone track, to allow time for the affected watersheds to respond to the rainfall. For each day within this case period the GloFAS-ERA5 discharge data in the land footprint area were compared to return period flood thresholds to calculate the return period exceeded at each point. The return period thresholds were taken directly from GloFAS, where they are determined from the GloFAS-ERA5 river discharge reanalysis, by fitting a Gumbel extreme value distribution on the annual maxima time series over the 1979-2018 period (Alfieri et al., 2019; Harrigan et al., 2023; Zsótér et al., 2020).

Flood severity can be defined in several ways, each measuring a contrasting property of the flood hazard which may be of particular relevance to different stakeholders. In this study the flood severity is calculated in the following three ways for each storm:

a) *Flood severity (Area)*: This is defined as the percentage of GloFAS grid points within the storm's footprint where the discharge exceeds the 5-year return period value for that point at any date during the case period. This has been termed 'area' to simplify the decomposition of flood severity into three component parts, but it is more akin to 'floodiness' (Stephens et al., 2015) defined over a storm's footprint than it is to measures of inundation extent. For Cyclone Eliakim, the flood severity (area) was 22% (Figure 5-3d).

b) *Flood severity (Duration)*: This is defined by selecting those points where the 5-year return period was exceeded on at least one day, and then calculating the average number of days for which the 5-year return period was exceeded at those points. This value is set to zero for cases where no points exceed the threshold at all during the case period.

c) *Flood severity (Magnitude)*: For this magnitude-based definition the data were filtered to only include points that have an upstream area of greater than 1,000km2, to focus on the magnitude of the flooding on larger rivers in the footprint. The magnitude of the flood is defined by first calculating the maximum return period that is exceeded at each of these river points during this TC event, and then ranking the points from highest to lowest and calculating the average return period exceeded over the top 100 points. The return periods are capped at 200 years as a sensible upper limit, given that the return period thresholds are generated on a 40-year sample (Harrigan et al., 2023). In this way the magnitude-based definition of flood severity compares the most severely affected river points in each TC case, in terms of the extremeness of the discharge levels experienced.

5.4 Results

5.4.1 *Comparison of flood severity measures*

The flood severity of all the TC cases, split geographically by the TC basin, is summarized in Figure 5-4. For all basins other than the South Indian Ocean, most of the cases have relatively low flood severity (Area) (median of 1-5%), but each basin has several cases with higher flood severity. However, in the South Indian Ocean basin, more of the cases have a higher flood severity (Area), with the median of 13% (Figure 5-4a) found by a Kruskal-Wall H-test to be significantly different from the other basins ($p<0.01$). For flood severity (Duration) the median duration for those points exceeding the 5-year return period threshold is between one and three days in all regions (Figure 5-4b). The median duration was higher for the Australian basin and the lower for the North West Pacific (both significant at $p<0.01$). For flood severity (Magnitude) the South Indian Ocean is again shown to have a higher median value compared to other regions (significant at $p<0.01$). All basins have some cases where very high return periods are exceeded (Figure 5-4c).

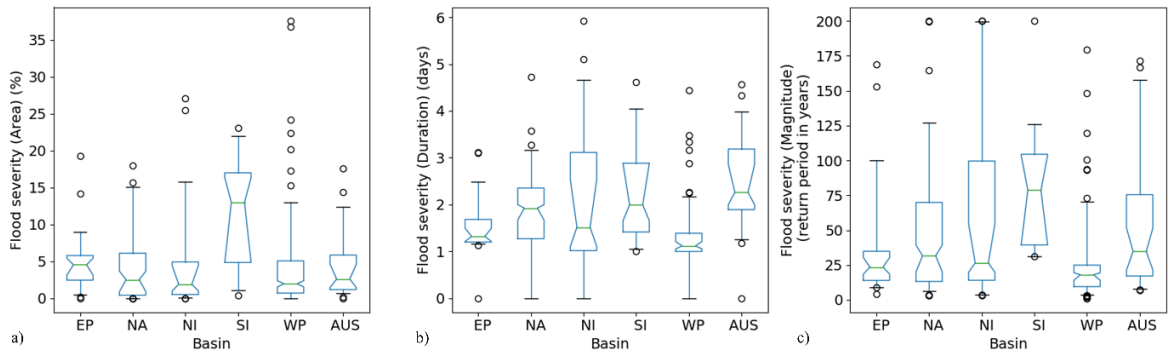


Figure 5-4. Box-and-whisker plot showing a) Flood severity (Area), b) Flood severity (Duration), and Flood severity (Magnitude) grouped by TC basin (AUS=Australian, EP=Eastern Pacific, NA=North Atlantic, NI=North Indian, SI=South Indian, WP=North West Pacific). In all the box-and-whisker plots in this study, the box extends from the Q1 to Q3 quartile values of the data and the whiskers extend from the 5th to the 95th percentiles. Outliers beyond this point are plotted as circles. The median is shown by a green line, with the 95% confidence bounds in that median shown by the notched area of the box, calculated by bootstrapping.

There are significant positive Spearman correlations between the three flood severity measures (Figure 5-5). Although the correlations are strong (> 0.5 in each case), there are some cases where storms have a very high flood severity in one measure and not in another. Good examples of this are Typhoon Hagibis (2019) and Typhoon Lan (2017), which can be seen in the bottom right of the plot in Figure 5-5a, with highest flood severity (Area), but where most of that flooding was very short lived, with flood severity (Duration) of only around 1 day. This emphasizes the importance of considering multiple indices, to take into consideration all characteristics of the flood severity.

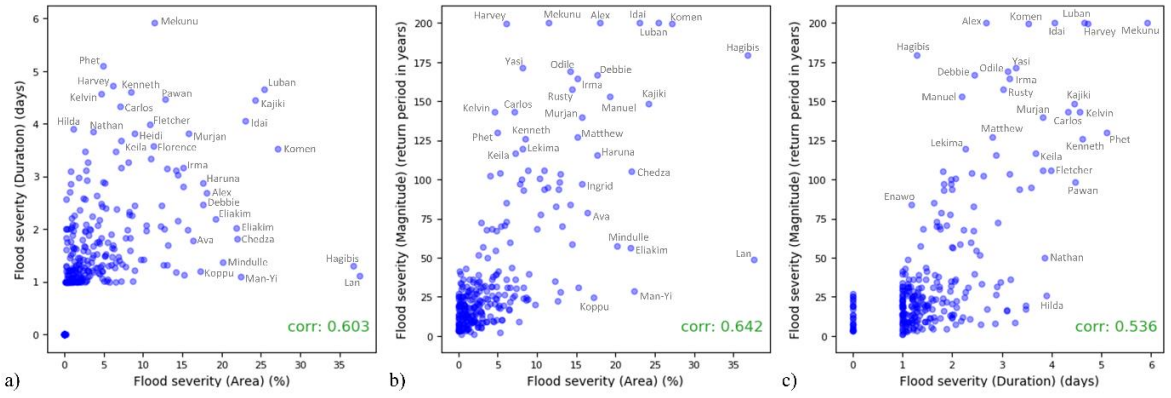


Figure 5-5. Scatter plots of the different flood severity indices: a) Flood severity (Area) vs Flood severity (Duration); b) Flood severity (Area) vs Flood severity (Magnitude); and c) Flood severity (Duration) vs Flood severity (Magnitude). The Spearman's Rank Correlation Coefficient is displayed on each plot. Points are annotated with the storm name if they have high values on either axis or are outliers to the general pattern.

The top ten most severe cases according to each of the flood severity indices are shown in Table 5-1. All basins have storms represented in the lists. Many storms appear in the top ten for two of the indices, and two storms, Cyclone Luban (2018) and Cyclone Idai (2019) appear in all three lists.

Table 5-2 contains a summary of the impacts from the TC cases that appear in at least two of the top tens. While it is harder to ascertain whether there were cases with severe impacts that were not highlighted, the fact that all of the storms in Table 5-2 had significant impacts from flooding indicates that the methods used in this paper have correctly highlighted high-impact cases of fluvial flooding from TCs.

Rank	Flood severity (area)	Flood severity (duration)	Flood severity (magnitude)
1	Lan (WP; 2017)	**Mekunu** (NI; 2018)	Alex (NA; 2010)
2	**Hagibis** (WP; 2019)	Phet (NI; 2010)	***Luban*** (NI; 2018)
3	**Komen** (NI; 2015)	**Harvey** (NA; 2017)	**Mekunu** (NI; 2018)
4	***Luban*** (NI; 2018)	***Luban*** (NI; 2018)	***Idai*** (SI; 2019)
5	**Kajiki** (WP; 2019)	Kenneth (SI; 2019)	**Harvey** (NA; 2017)
6	***Idai*** (SI; 2019)	Kelvin (AUS; 2018)	**Komen** (NI; 2015)
7	Man-Yi (WP; 2013)	Pawan (NI; 2019)	**Hagibis** (WP; 2019)
8	Chedza (SI; 2015)	**Kajiki** (WP; 2019)	Yasi (AUS; 2011)
9	Eliakim (SI; 2018)	Carlos (AUS; 2011)	Odile (EP; 2014)
10	Mindulle (WP; 2016)	***Idai*** (SI; 2019)	Debbie (AUS; 2017)

Table 5-1. The 10 highest-ranked TCs for each flood severity measure (in boldface type if that TC case is in the top 10 for two of the three indices, and in boldface italic if it is in the top 10 of all three indices). The regions are defined in the caption of Fig. 4.

Case	Summary of impacts
Idai (SI; 2019)	Severe flooding in Mozambique, Malawi, and Zimbabwe, leading to a major humanitarian crisis; damages of at least $2.2 billion (2019 USD), 1.85 million people in need of humanitarian assistance, and over 600 fatalities in Mozambique alone (Emerton et al. 2020); deadliest TC recorded in the South West Indian Ocean basin
Luban (NI; 2018)	Flooding in Somalia, Oman, and Yemen, with 14 deaths in Yemen, with damages in the country estimated at U.S. $1 billion (Associated Press 2018)
Hagibis (WP; 2019)	98 confirmed deaths in Japan and more than 270 000 households left without power across the country; severe damage to transport infrastructures
Komen (NI; 2015)	Komen led to 45 deaths in Bangladesh, 83 in India, and 39 deaths in Myanmar
Kajiki (WP; 2019)	At least 6 fatalities in Vietnam, with agricultural losses estimated at U.S. $12.9 million
Mekunu (NI; 2018)	31 fatalities (Socotra island, Yemen, and Oman); severe flooding led to power outages, and landslides
Harvey (NA; 2017)	Unprecedented flooding in Texas inundating hundreds of thousands of homes, displacing more than 30 000 people, and prompting more than 17 000 rescues; over 100 fatalities

Table 5-2. A summary of the impacts from those TCs that appear in the top 10 for at least two of the flood severity measures (in boldface type if that TC case is in the top 10 for two of the three indices, and in boldface italic if it is in the top 10 of all three indices). All are TCs that led to significant impacts.

5.4.2 *Relationships between flood severity and case characteristics*

5.4.2.1 Correlations

The Spearman's Rank Correlation Coefficients between the flood severity metrics and all the calculated storm and catchment characteristics are shown in Figure 5-6. A comparison between GPM and ERA5 rainfall (not shown) revealed very strong correlations, so only the ERA5 precipitation variables are included. All the flood severity variables are significantly correlated with each other, as was seen in Figure 5-5. All precipitation variables (both for the precipitation field prior to landfall and for the accumulated precipitation variables within the footprint) are significantly positively correlated with the flood severity (Area) index, with the highest correlations for the average accumulation in the footprint where it rained (0.58) and for the percentage of the footprint over

100/200mm accumulations (0.57/0.54). For flood severity (Duration), there is no significant correlation with the precipitation prior to landfall, but there are significant correlations with most total storm accumulation measures. For flood severity (Magnitude), there are positive correlations with all precipitation variables apart from the percentage of the footprint over the lowest value of 25mm.



Figure 5-6. Spearman's Rank Correlation Coefficients of the flood severity variables with the TC characteristics and catchment characteristics fields for each TC case. Those with a Spearman's Rank Correlation Coefficient with a significance value <0.05 are marked in bold with an asterisk.

The storm *intensity* variables, MSLP and wind speed at landfall, are significantly negatively (-0.18) and positively (0.13) correlated with flood severity (Area) respectively. This shows that generally the more intense the storm (lower MSLP minima and stronger winds), the greater the flood severity. However, neither intensity variable is significantly correlated for flood severity (Duration) or flood severity (Magnitude), so it is not always the case that a strong storm will lead to severe flooding in terms of duration or magnitude, or that a weaker storm cannot cause significant flood events. A comparison of the distribution of flood severity values for TC cases split into those that are tropical storms at landfall with those that have equivalent hurricane or major hurricane strength (Figure 5-7) confirms that all storm categories have some cases with both low and high flood severities. There are no significant differences between the severity of flooding in tropical storms and hurricanes, but for major hurricanes there is a notable increase in the upper quartile range and the median is significantly higher for the area and magnitude definitions of flood severity (Kruskal-Wall H-test with $p<0.01$).

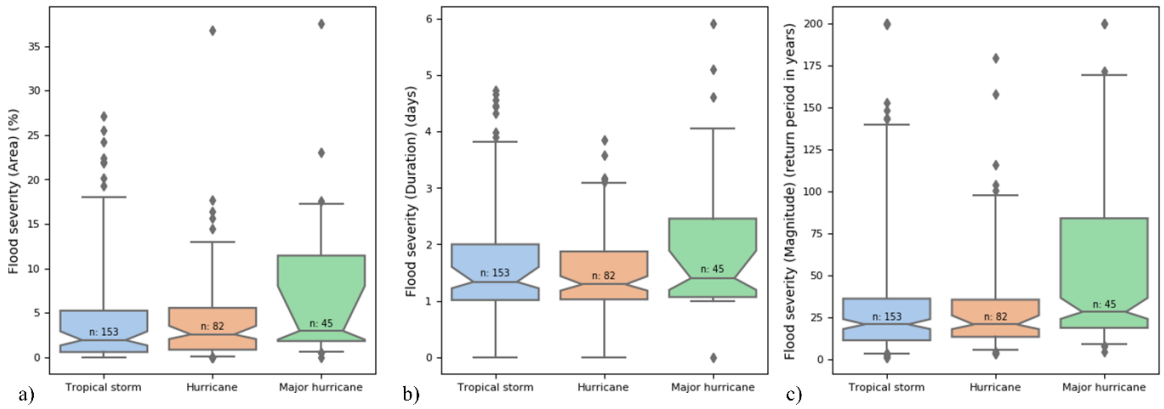


Figure 5-7. Flood severity (Area (a), Duration (b), and Magnitude (c)), for three categories of TC calculated based on the maximum wind speed at landfall taken from IBTrACS: i) Tropical storm (34–63 knots), ii) Equivalent hurricane strength (64-95 knots), iii) Equivalent major hurricane strength (96+ knots). In actuality different storm categories and thresholds are used by different RSMCs when defining intensity, but in order to compare across basins, the categories based on the Saffir-Simson scale are used here. The 95% confidence bounds in the median is shown by the notched area of the box, calculated by bootstrapping. The number of cases in each category is overlain.

The three *translation speed* variables are all significantly negatively corelated with all the flood severity indices i.e. slower-moving storms tending to have greater flood severity when considered in terms of area, duration and magnitude. The strongest correlations are with the duration of flooding (-0.31, -0.32, and -0.33 respectively for the translation speed at landfall, the average translation speed over land, and the minimum translation speed over land).

The *storm size* as measured by the ROCI is significantly positively correlated with flood severity (Area) (0.12), indicating that larger storms tend to have a greater percentage of the land footprint exceeding the 5-year return period. There is no significant correlation with flood severity (Duration) or flood severity (Magnitude), and no significant correlation of *storm asymmetry* with any of flood severity measures.

The relationship between flood severity and *soil saturation* is complex and varies with each of the flood severity indices. For flood severity (Area), all the soil saturation metrics have positive correlations, showing that cases with more saturated antecedent soil conditions tend to have greater flood severity. However, a different pattern is shown for the other two flood severity indices (Duration and Magnitude), where there are significant negative correlations with the average soil moisture saturation in the top 100cm layer (-0.27 and -0.15 respectively).

There is a significant positive relationship of average *orography* and flood severity (Area) (0.24), and there are significant negative correlations of the average *gradient* with flood severity (Duration) (-0.44) and flood severity (Magnitude) (-0.31), so lower gradients would tend to have greater flood severity in terms of the flood duration and magnitude.

5.4.2.2 Conditional probabilities

Conventional wisdom is that translation speed is a significant risk factor leading to increased flood severity. Significant relationships were shown between the flood severity indices and the translation speed of the TC in Figure 5-6, but the correlations are perhaps lower than anticipated, at -0.22, -0.33, and -0.26 for the minimum translation speed over land when compared to flood severity as measured by area, duration, and magnitude respectively. However, if we look in more detail at scatter plots for these relationships, and apply a conditional probability technique to split the plot into four quadrants, further useful information can be found (Figure 5-8). The x-axis is split by the median value of the minimum translation speed over land, to designate slow-moving and fast-moving storms. A sensitivity analysis (not shown) was used to select which percentile of the y-axes (the flood severity scores) to use to designate non-severe and severe flooding cases. The 90th percentile was selected as the threshold that gave the best combination of having sufficient cases included in the upper section while also focusing in on the more severe cases (the top 10%). The conditional probability of having severe flooding if the storm is slow moving is then calculated. For flood severity (Area), the conditional probability of having severe flooding if the storm is slow moving is 14%, more than double the conditional probability if the storm is fast moving (6%) (Figure 8a). So although there are a few cases (e.g. Hagibis, Man-Yi, and Lan) in the upper right quadrant of Figure 5-8a where flood severity is high in fast-moving storms, and lots of cases where slow-moving storms do not lead to severe flooding, overall there is a greater risk of severe flooding in slow-moving storms, with a significant cluster of storms with high flood severities and low translation speeds (e.g. Komen, Idai, Luban, Kajiki, Eliakim). The results when we apply the same conditional probability technique to the other two flood severity metrics are even clearer, with three times the probability of having a severe flooding event for slow-moving storms for flood severity (duration), and four times the probability of a severe flooding event for slow-moving storms for flood severity (magnitude) (Figure 5-8b and 5-8c).

Fast movement of TCs appears to often mitigate severe flooding, and slow movement increases the likelihood of severe flooding, but these patterns are not seen in all cases. The role of other factors in moderating this relationship can be examined by comparing conditional probability plots of flood severity against several different characteristics, selected based on the correlation results shown in Figure 5-6 (Figure 5-9). Flood severity (Area) is used to examine in more detail the interplay between characteristics, as the extent of the severe flooding across the whole area affected by the TC is perhaps the most relevant when considering preparedness and recovery activities such as evacuation planning. The strong relationship between flood severity and the percentage of the footprint with high precipitation totals can be seen in Figure 5-9a, with only 2% of the cases lower

than the median precipitation having severe flooding. When the conditional probability technique is applied to the average soil saturation in the top layer of the soil, there is found to be more than double the risk (14% compared to 6%) of having severe flooding where there is high soil saturation (Figure 5-9b). The bottom three plots (6-9c-e) show three of the TC characteristics that may play a role in increasing the overall precipitation accumulation: intensity, translation speed, and size. For the more intense storms, there is a small increased probability of severe flooding (12%) when compared with the less intense storms (7%) (Figure 5-9c). The doubling of the likelihood of severe flooding for slow translations speeds, as shown in Figure 5-8a, are repeated in Figure 5-9d, while Figure 5-9e shows the conditional probability results for TC size, with nearly double the risk of severe flooding for large storms (13%) compared to small storms (7%).

Typhoons Man-Yi (2013), Hagibis (2019) and Lan (2017) appear as relative outliers in the upper right quadrant of Figure 5-9d (fast-moving storms where flood severities were unexpectedly high or very high). On Figure 5-9e it can be seen that these are all very large storms, indicating that the size of these storms may have been a significant factor contributing to their high flood severity, acting to override the mitigative nature of their fast translation speed. They are also categorized in the more intense category for intensity (Figure 5-9c), and the high soil saturation category (Figure 5-9b). Thus, in these cases the high translation speed of these typhoons did not prevent a severe flood event due to the presence of several other risk factors.

A similar conditional probability analysis for the other two flood severity measures (duration and magnitude) (not shown) found that the main driving factor from the TC characteristics was the translation speed of the storm. The other main risk factors were from the catchment characteristics, with drier soils and lower gradients being associated with a large increase in the risk of having a longer and higher magnitude flood event. These results show that the influence of the catchment characteristics on the duration and magnitude of flooding is greater than for the flooding measured in terms of the area.

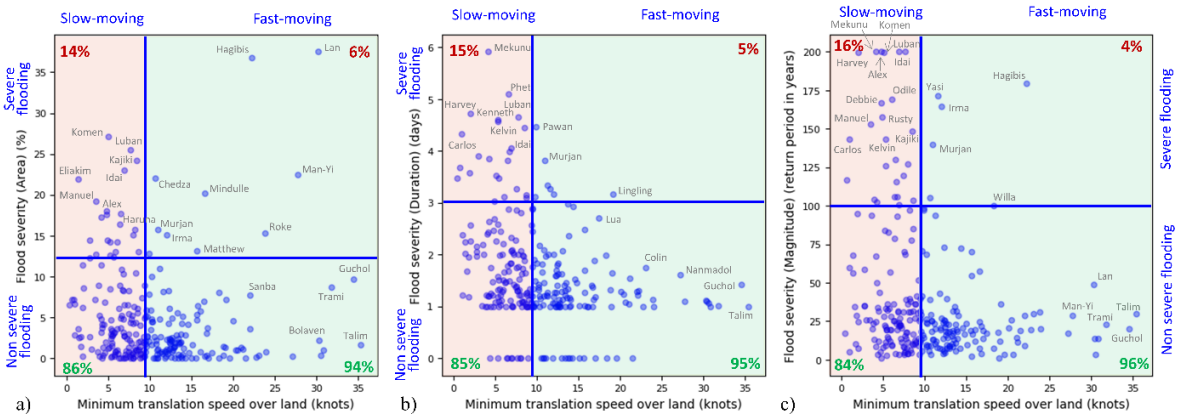


Figure 5-8. Illustration of the conditional probability technique for examining the relationship between minimum translation speed over land and flood severity. The blue lines divide the plot into quadrants based on the median translation speed (9.19 knots) and the 90th percentile of the flood indices (12.86% for flood severity (Area), 3.10 days for flood severity (Duration) and 102.59 for flood severity (Magnitude)). There are therefore equal numbers of cases in the orange and green halves of the plots, while the upper section of the plot displays the 10% of the TC cases with the most severe flooding. The numbers in the top right and top left quadrants can be compared to see if there is an elevated risk of severe flooding for slow moving storms. Points are annotated with the storm name if they have high values on either axis or are outliers to the general pattern. The Spearman's Rank Correlation Coefficients for each plot are as follows: a) -0.22; b) -0.33; c) -0.26.

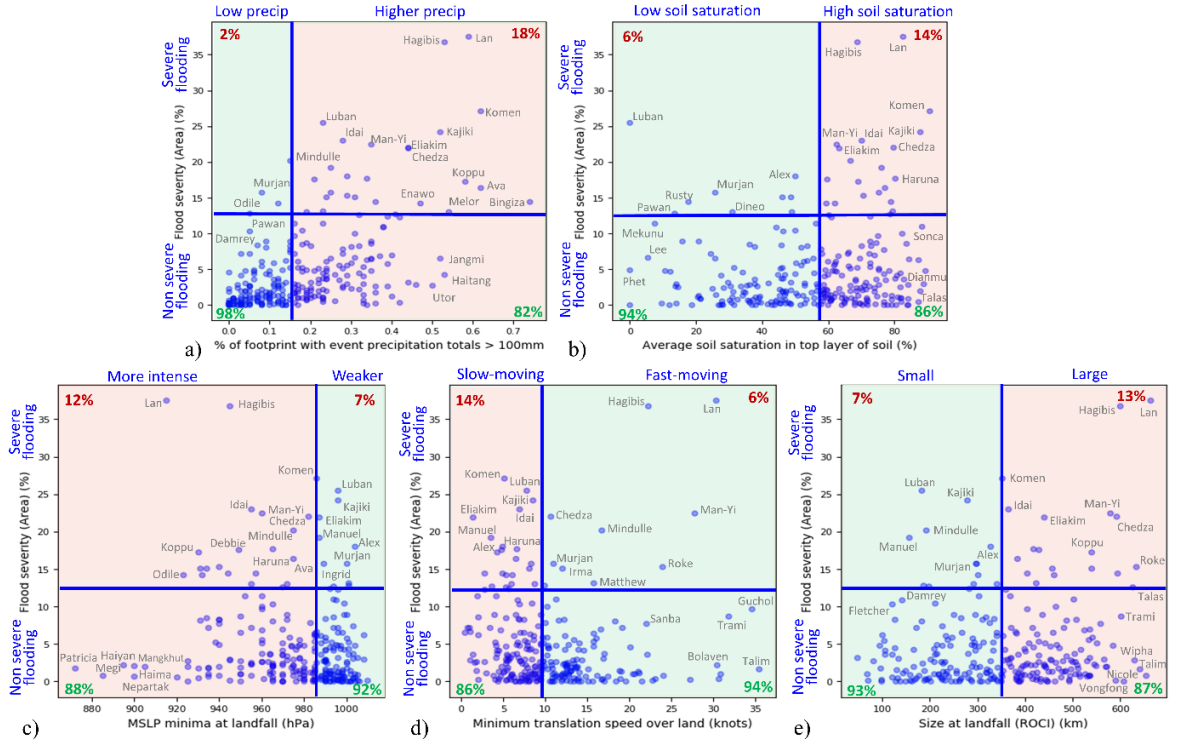


Figure 5-9. Scatter plots for the flood severity (Area) indices against a) the percentage of the footprint with event precipitation totals greater than 100mm; b) average soil saturation in the top layer of soil; c) mean sea level pressure minima at landfall; d) minimum translation speed over land, and e) size at landfall (ROCI). Each plot has the conditional probability technique applied using the median of the variable and the 90th percentile of the flood severity to create the quadrants. The Spearman's Rank Correlation Coefficients for each plot are as follows: a) 0.57; b) 0.15; c) -0.18; d) -0.22; e) 0.12.

5.4.2.3 Interplay of characteristics

To investigate further how the combination of several factors helps to determine the overall flood severity, two characteristics, TC translation speed and TC size, have been used to split the cases into four categories based on whether they are larger/smaller than the median size, and faster/slower than the median translation speed (Figure 5-10). For flood severity (Area), cases where the TC is both slow and large have the highest median and upper quartile range, followed by slow and small, then fast and large, and finally fast and small cases. Meanwhile for flood severity (Duration) and flood severity (Magnitude) there is a clearer split between the fast types and the slow types, showing that for flood duration and magnitude the translation speed of the storm is a more important factor than the size of the storm.

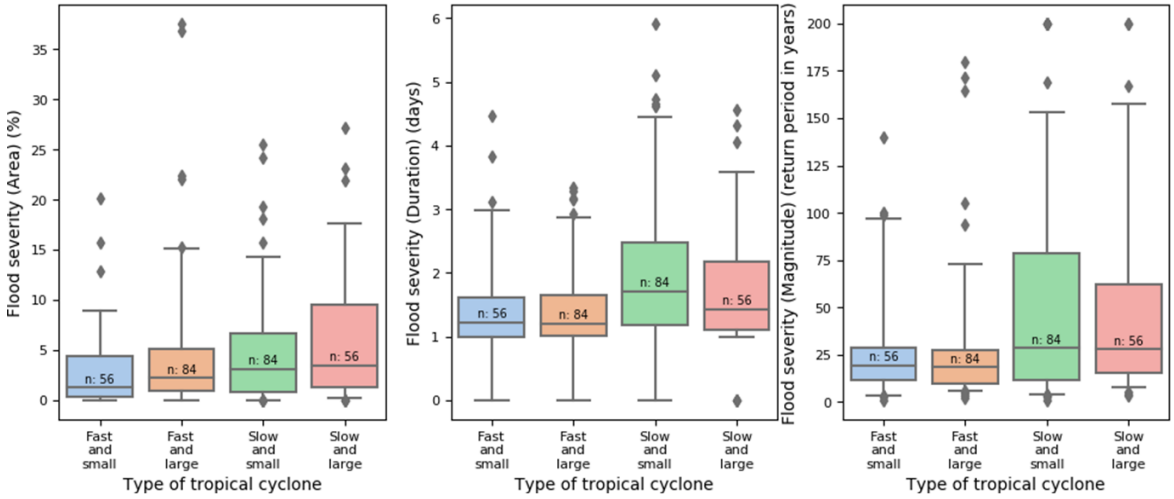


Figure 5-10. Flood severity (Area (left), Duration (center), and Magnitude (right)), for four categories of TC: i) fast and small, ii) fast and large, iii) slow and small, iv) slow and large. The number of cases in each category is overlain.

This process is taken further by grouping each TC case depending on how many of the four risk factors found in Figure 5-9 (slow-moving storm, large storm, intense storm and wet antecedent conditions) are present (Figure 5-11). Total precipitation is not included here as we are focusing on the factors that influence that precipitation, that are easier to estimate a-priori when assessing the risk associated with a storm in forecast mode. Figure 5-11a shows that the likelihood of a higher flood severity increases the more risk factors are present. Figure 5-11b splits the categories by which of the risk factors are present. The storms that are classed as having all four risk factors (slow, large, strong and wet) have the highest median flood severity and the highest upper quartile range. Within those cases that have three risk factors, cyclones that are slow and large with wet antecedent conditions stand out as having a high upper quartile range. For those cases with two risk factors, the highest median value is for cases with slow TCs with wet antecedent conditions, while where only one risk factor is present, there are higher flood severities for slow storms.

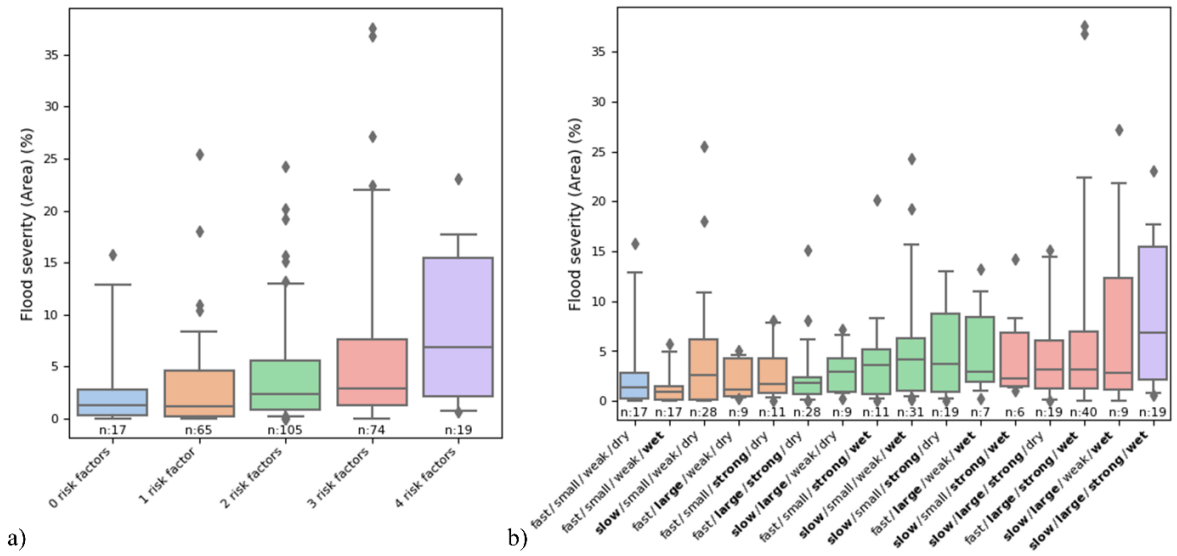


Figure 5-11. Flood severity (Area), for TCs grouped by risk factors (slow-moving storm, large storm, intense storm, wet antecedent conditions): a) grouped according to how many of the risk factors they possess; b) further split by which risk factors (highlighted in bold) are present in each group. The number of cases in each category is displayed below the box plot.

5.5 Discussion of the key outcomes

5.5.1 *Applying flood severity metrics to tropical cyclones footprints*

This study has implemented three flood severity indices designed to measure the severity of flooding in TCs, based on i) the area affected; ii) the duration of the flooding; and iii) the magnitude of the flooding in the most severely affected areas. Although the different flood severity measures are all significantly positively correlated with each other, there are cases where a storm has a very high flood severity in one measure and not in another e.g. where a large area is affected but flood duration is low. Some of the drivers of an enhanced flood severity found in this study are consistent whether considering the area affected, duration, or magnitude e.g. high total precipitation and low translation speed, whereas some drivers are only significant for one or two of the indices e.g. large storms and the area-based index, and gentler slopes with the duration- and magnitude-based indices. This emphasizes the importance of considering multiple indices to take into consideration all characteristics of the flood hazard, especially given that stakeholders may have differing priorities. An organization with responsibility for emergency planning across a large geographical area would be particularly concerned about the extent of the area being affected by flooding, whereas an individual in an area prone to flooding may be most concerned about the magnitude of flooding likely to be experienced.

5.5.2 *Importance of precipitation and translation speed*

The strongest relationships are found with the metrics summarizing the event total precipitation, showing the importance of increasing the availability and profile of precipitation forecasts in TC forecast communication and verifying the ability of models that drive flood forecasting systems to forecast the precipitation associated with TCs.

Of the other TC characteristics, the speed of forward movement of the TC is found to be a key factor influencing the severity of the fluvial flood hazard, in terms of its area, duration and magnitude. This confirms that slower moving storms, which rain over a given area for longer, tend to have higher flood severity, and that greater focus should be given to translation speed in forecast warnings and communication. Climate change is thought to have caused a general weakening of summertime tropical circulation, decreasing the translation speed of TCs (Kossin, 2018). This, combined with an expected increase in the magnitude of rainfall rates in TCs (Knutson et al., 2010), is likely to lead to an increased fluvial flood hazard from these events in the future.

5.5.3 *Presence of flood hazard at all levels of storm intensity*

TC intensity is positively correlated with the flood severity as defined by area, but not significantly correlated with the duration or magnitude of the flooding. A further analysis splitting cases into those of equivalent tropical storm, hurricane, and major hurricane strength, showed that all three have cases spanning the full range of flood severity values. This shows the importance of forecasting agencies continuing to emphasize to both the public and to organizations involved with disaster risk reduction preparedness and response activities that the fluvial flood hazard can be severe in all categories of TC.

5.5.4 *Complex relationship between flood severity and antecedent soil moisture*

Conventional wisdom, established by looking at small to medium sized flood events, is that higher levels of soil moisture saturation would lead to an increased risk of flooding (Berghuijs et al., 2019). However, when dealing with the extreme precipitation totals that can be found in TC events, the relationship appears more complex. For flood severity (Area), there are indeed more cases of severe flooding in those cases with high soil saturation. However, a different pattern is shown for the other two flood severity indices (duration and magnitude), where there is a higher conditional probability of severe flooding for lower (drier) antecedent soil moisture. Areas that are drier climatologically (e.g. the Arabian Peninsula) show higher return periods being exceeded, which is likely due to a combination of flash flooding, and the difficulty of accurately estimating return period thresholds in arid areas. Using soil moisture anomaly data rather than soil moisture saturation in future studies may help to further unpick this complex relationship.

5.5.5 *Interplay of characteristics in modulating flood severity*

The interplay between different characteristics acts to modulate the overall flood severity. For example, although fast-moving storms can mitigate fluvial flooding, there can still be a significant flood event if there are other risk factors present. An example of this is Typhoon Hagibis, which led to widespread and severe flooding in Japan in 2019. Here, the relative mitigation of Hagibis being fast-moving was outweighed by the storm's size, intensity, and wider atmospheric environment, leading to very heavy precipitation and severe flooding, albeit over a relatively short period (Takemi and Unuma, 2020). Contrastingly, although slow-moving storms significantly increase the conditional probability of a severe flood event, many slow-moving storms do not lead to significant flooding e.g. Tropical Storm Beryl (2012), which moved very slowly over northern Florida, but was relatively small and weak, and passed over an area where the antecedent conditions were dry. As this study included over 250 cases it was not possible to explore in detail the varied complexities that contributed to the detailed flood response in each individual case, such as the wider atmospheric environment, in a way that would be possible if only examining one or two cases. Nevertheless it provides a systematic analysis of potential meteorological and hydrological drivers, and a key finding is that the more driving factors a case possesses, the higher the likelihood of severe flooding, with in particular TCs that were slow, large, and strong, with wet antecedent conditions, having a higher flood severity in terms of the area that was affected.

5.5.6 *Implications for research into the predictability of fluvial flooding from tropical cyclones*

The severity of flooding in TC cases is shown to be dependent on a range of both meteorological and hydrological factors, showing the importance of a multi-disciplinary Earth system approach to global flood prediction (Harrigan et al., 2020). The flood forecasts from hydrometeorological forecasting systems will be highly dependent on how reliably the input meteorological forecasts are able to predict not only the track and the intensity of TCs, but also their precipitation extent and intensity. This in turn will be dependent on the size and in particular the translation speed of the storm. There has generally been less focus on verifying these characteristics of TCs in ensemble forecasts, with most studies focusing on the predictability of track, intensity, and genesis. In order to understand and improve the forecasts of fluvial flooding from TCs, studies evaluating these aspects of forecast performance are recommended. It would also be beneficial to examine how uncertainty in each aspect of the forecast, and in each part of the forecasting chain, impacts on the overall predictability of fluvial flooding for TCs.

5.5.7 *Implications for disaster risk reduction, preparedness and response*

The finding that the forward speed of the storm is of more importance than the intensity in modulating the severity of fluvial flooding is an important result. The expected speed of movement

of the storm tends to have far less coverage in both traditional and social media than the intensity of the storm. This means that the public, local emergency responders, and those involved in planning disaster risk reduction activities may become too focused on storm intensity categories when making evacuation decisions, undertaking storm preparedness activities, or planning deployment of resources. Significant impacts from fluvial flooding can occur regardless of whether it is a strong storm, where there will likely also be significant wind and storm surge hazards, or a weak storm, where the fluvial flooding may be the main hazard. Both circumstances have important implications for forecast-based early action. An example of a TC that caused significant flooding while being relatively weak is Cyclone Komen, which impacted Myanmar, Bangladesh and India in 2015. Komen only reached equivalent strength of a tropical storm, but was very slow-moving, with a broad circulation leading to prolonged heavy rainfall over areas of Myanmar already saturated due to weeks of monsoonal flooding, leading to severe impacts. Raising the profile of cases such as these where the wind hazards may be relatively low, but there is potential for a slow-moving storm impacting vulnerable areas, meaning the fluvial flood hazard and impacts are likely to be more severe, will help to ensure that preparations are made to ameliorate impact. For intense TCs, where evacuations and deployment of aid may be taking place in advance of landfall due to the anticipated wind impacts, it is also vital to know whether that storm also has an elevated likelihood of fluvial flooding due to other risk factors such as a slow translation speed or saturated antecedent soil conditions. If it does it would be important to select evacuation routes, shelters and aid deployment sites that would not be impacted in the event of extensive fluvial flooding, which may extend far from the landfall location. For example, Cyclone Idai, which led to catastrophic flooding in Mozambique in March 2019, exhibited most of the factors shown in this study to be key drivers of fluvial flood hazard: it was a strong storm, but significantly it was also a slow-moving storm at and after landfall, a relatively large storm, and also had wet antecedent soil moisture conditions. Highlighting this co-occurrence of flood drivers to decision-makers in future cases where a storm is approaching landfall will help to ensure early action can be taken to plan for and mitigate the impacts of severe flooding.

5.6 Conclusions

Landfalling TCs over the last decade have been classified in terms of the severity of fluvial flooding that was associated with each storm using GloFAS-ERA5 reanalyses. Flood severity has been calculated using three different indices, each focusing on a different characteristic of the associated fluvial flooding (area, duration and magnitude). The key factors that impact the severity of fluvial flooding were investigated by comparing these flood severities with a range of TC characteristics

(storm speed, size, shape, intensity, and precipitation), and catchment characteristics (antecedent soil moisture, orography, and gradient). The key findings are:

- There are cases that have very high flood severity in one measure and not in another, highlighting the importance of considering the area, duration and magnitude characteristics of TC flood events.
- Flood severity is strongly positively correlated with metrics summarizing the event total precipitation, showing the importance of increasing the availability and profile of precipitation forecasts in TC forecast communication and verifying the ability of models that drive flood forecasting systems to forecast the precipitation associated with TCs.
- Slow-moving storms were found to have a much higher conditional probability of a severe flood event compared to fast-moving storms, confirming that slower moving storms, where storms are raining over an area for longer, tend to have higher flood severity.
- All intensities of TC have cases spanning the full range of flood severity values, including several cases of tropical storm strength that were associated with severe flooding. This confirms the need for forecasting agencies to continue emphasizing that the fluvial flood hazard can be severe in all categories of TC.
- Large storms have a higher conditional probability of severe flooding than small storms in terms of the area affected. The duration and magnitude of the flooding are found to be more dependent on the catchment characteristics within the land footprint of the storm, with gentler slopes having a much greater conditional probability of severe flooding. The relationship between flood severity and antecedent soil moisture is found to be complex and not consistent across the different flood severity indices.
- Several TC and catchment characteristics often combine to influence the overall flood severity and negate or enhance other factors e.g. fast-moving storms still having a severe flood hazard if they are large, intense and passing over saturated ground. The more driving factors that a case possesses the higher the flood risk is found to be, with the highest likelihood of having a large area affected by flooding found for TCs that were slow, large, intense, and affecting areas with wet antecedent conditions.

The reanalysis data used in this study provides a proxy for the observed severity of flooding, and although this allows a direct comparison of flood severity across all TC cases, studies of fluvial flooding in TCs would benefit hugely from an increase in the availability and consistency of hydrological observations (Lavers et al., 2019).

This study has highlighted the characteristics of TCs that have the most control on flood hazard. For those involved in communicating early warnings and taking forecast-based action before a storm

our results show the importance of considering aspects such as storm speed when assessing the risks. For developers of hydrometeorological ensemble forecasts of river flows our work underlines that the input meteorological ensemble forecasts for TCs need to be able to reliably predict not just their track and intensity, but also their precipitation, size and translation speed. Future research needs to focus on verifying and improving these characteristics for agencies to be able to provide more accurate forecasts of flood hazard.

Chapter 6 Using ensembles to analyse predictability links in the tropical cyclone flood forecast chain

The fourth objective of the PhD was to analyse predictability links in the tropical cyclone flood forecast chain using an ensemble-based method. The paper covering this objective has been submitted to the Journal of Hydrometeorology (JHM) with the following reference:

Titley, H.A., Cloke, H.L., Stephens, E.M., Pappenberger, F. and Zsótér, E. (2023) Using ensembles to analyse predictability links in the tropical cyclone flood forecast chain. Journal of Hydrometeorology (re-submitted June 2023 following review).

The contributions of the authors of this paper are as follows: H.T. designed the study with advice from H.C., E.S. and E.Z.. H.T. obtained the datasets, with assistance from E.Z. for the GloFAS data. H.T. carried out the analysis and led the writing of the manuscript. All authors assisted with writing the manuscript. Overall, 90% of the writing was undertaken by H.T..

The submitted article can be found in Appendix A3.

Abstract. Fluvial flooding is a major cause of death and damages from tropical cyclones (TCs), so it is important to understand the predictability of river flooding in TC cases, and the potential of global ensemble flood forecast systems to inform warning and preparedness activities. This paper demonstrates a methodology using ensemble forecasts to follow predictability and uncertainty through the forecast chain in the Global Flood Awareness System (GloFAS), to explore the connections between the skill of the TC track, intensity, precipitation and river discharge forecasts. Using the case of Hurricane Iota, which brought severe flooding to Central America in November 2020, I assess the performance of each ensemble member at each stage of the forecast, along with the overall spread and change between forecast runs, and analyse the connections between each forecast component. Strong relationships are found between track, precipitation and river discharge skill. Changes in TC intensity skill only result in significant improvements in discharge skill in river catchments close to the landfall location that are impacted by the heavy rains around the eye wall. The rainfall from the wider storm circulation is crucial to flood impacts in most of the affected river basins, with a stronger relationship with the post-landfall track error rather than the precise landfall location. We recommend the wider application of this technique in TC cases, to investigate how this cascade of predictability varies with different forecast and geographical contexts, to help inform flood early warning in TCs.

Significance statement. This study demonstrates a methodology to analyse the cascade of predictability and uncertainty through the various stages of the tropical cyclone (TC) flood forecasting chain, illustrating how it can provide useful information to modellers interested in optimising flood forecast skill, and to those who prepare and communicate flood forecasts with stakeholders and end-users in TC cases. The results highlight the importance of improving verification of ensemble TC precipitation forecasts, and of focusing on more than just the category of the storm and landfall location when forecasting and communicating flood impacts in TC cases.

6.1 Introduction

Despite the major role of rain-induced flooding in deaths and damages related to tropical cyclones (TCs) (Rappaport, 2014; Czajkowski and Kennedy, 2010; Czajkowski et al., 2013) and predictions that TC rainfall rates and associated flooding impacts will increase due to climate change (Knutson et al., 2010, 2013, 2015, 2020; Wang et al., 2015; Wright et al., 2015), the main determining factor in evacuation decision-making remains the intensity of the TC (Stein et al., 2010; Whitehead et al., 2000; Senkbeil et al., 2019). The potential for heavy rainfall is present regardless of the storm's category (Titley et al., 2021) so it is vital to increase public awareness of the dangers of river flooding in TCs. The importance of increasing the focus on the probabilistic forecasting of downstream hazards has been recognised by key international research programmes (Golding et al., 2022; Titley et al., 2019). Improving the understanding of the factors that influence the predictability of flood-related hazards during TCs can help optimise forecast guidance and therefore improve preparedness in advance of an approaching storm. However, investigating this predictability is hampered by a limited sample size of forecast TCs affecting a single region, and the difficulties of aggregating across regions due to geographical differences (e.g. river length and orientation, topography). Ensemble forecasts have the potential to help overcome this issue by providing a larger sample of potential outcomes for what are rare events in any given location.

This work therefore aims to develop a novel methodology to use ensemble forecasts to explore how each part of the TC flood forecast chain (track, intensity, precipitation, and hydrology) influences the overall predictability of river flooding in TC cases. The ensemble forecasts are examined across several forecast runs to create hundreds of sample forecasts for a particular high-impact case. Taking inspiration from previous ensemble-based sensitivity analysis (Hakim and Torn, 2008; Lynch and Schumacher, 2014; Zhang and Meng, 2018; Shen et al., 2020), in this study we develop methods to use the ensemble forecasts to investigate the interconnectivity of the various forecast components of the TC flood forecast chain and give valuable insights as to the controls on TC flood predictability. The following questions will be addressed: i) How do the errors and the

spread in the TC track and intensity forecasts relate to the errors and spread in the precipitation forecasts and the river discharge forecasts?; and ii) Do the members with the best/worst track forecasts necessarily lead to good/poor precipitation and river discharge forecasts?

The forecast system selected for use in this paper is the Global Flood Awareness System (GloFAS, www.globalfloods.eu), an operational system of the European Commission's Copernicus Emergency Management Service, which is used for producing flood forecast bulletins for humanitarian users for TCs (e.g. Emerton et al., 2020; Speight et al., 2023). GloFAS produces probabilistic river discharge forecasts driven by ensemble weather forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF), providing early indication of upcoming flood events. The 51 ensemble members within the GloFAS forecasting system can be traced through the complete forecast chain from TC track and intensity, through to their rainfall and river discharge forecasts, and over several forecast runs, providing a larger sample size of forecasts that can be used to assess the relationships between each forecast component. Using the GloFAS forecasting system in this way can also allow an initial assessment of the lead times at which it can provide useful guidance for potential flooding in TC cases, and what forecast and geographical contexts can cause variations in this. The methods are demonstrated using the case of Hurricane Iota which affected over 5 million people in Central America in November 2020 (UN-OCHA, 2020b, 2020a).

6.2 Case background

Hurricane Iota was a strong category 4 hurricane on the Saffir-Simpson scale that caused severe and widespread river flooding in Nicaragua, Honduras and Guatemala, resulting in at least 67 deaths and total damages estimated at $1.4 billion (Stewart, 2021). Iota originated as a tropical wave that moved into the Eastern Caribbean on 10th November 2020 before strengthening and being named Tropical Storm Iota on 13th November. Iota became hurricane-strength on 15th November, and rapidly intensified to a Category 4 hurricane, with peak maximum wind speeds of 135 kt, before making landfall with an intensity of 125 kt on 17 November at 03:40 UTC, south of Puerto Cabezas in northeast Nicaragua. Several sites across Nicaragua and Honduras reported event rainfall totals over 200mm, with a maximum total of 510mm in Guatemala (Stewart, 2021). The extreme rainfall led to estimated floods exceeding the 1 in 20-year thresholds across major rivers in these countries (Figure 6-1), as well as widespread flash flooding and mudslides. Flooding was exacerbated by pre-existing flood conditions caused by Hurricane Eta which had made landfall in a similar location two weeks earlier. In total 5.2 million people were affected with widespread damage to properties and infrastructure (IFRC, 2020; UN-OCHA, 2020b, 2020a).

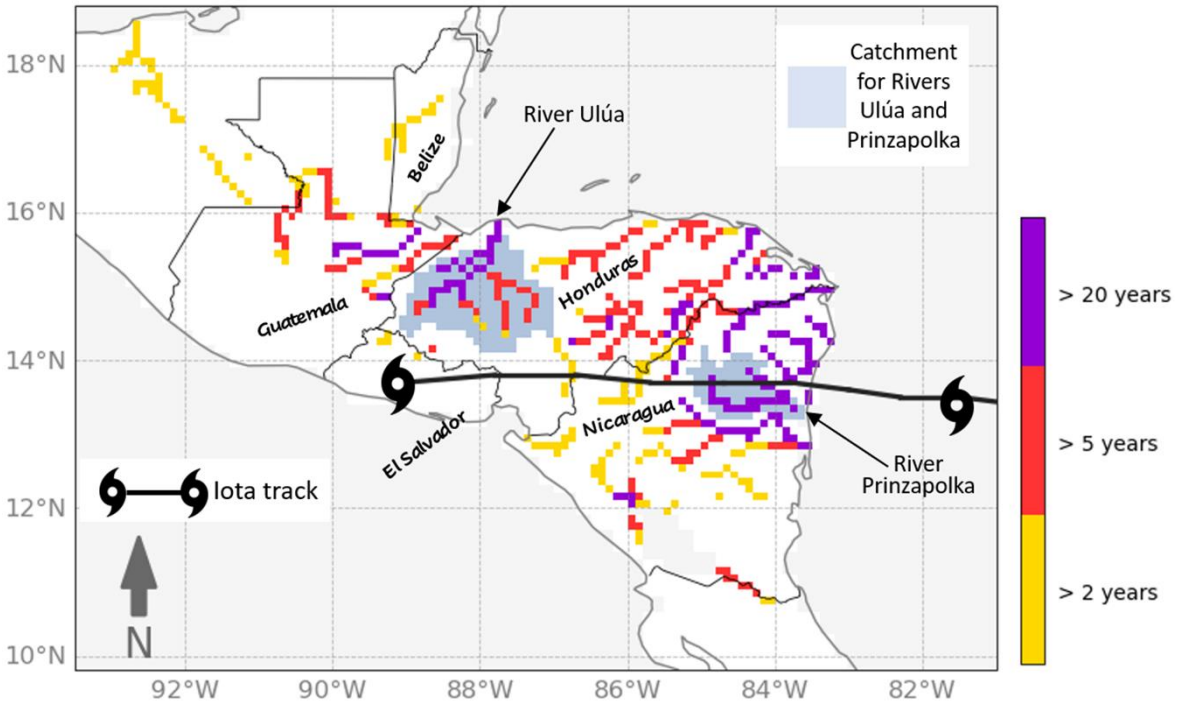


Figure 6-1. The return period estimated to have been exceeded from 16th-20th November 2020 in the GloFAS-ERA5 reanalysis. Two of the most severely affected rivers are labelled (the Ulúa and the Prinzapolka, for which the forecasts are examined in Section 6.4), with their catchment areas shown in light blue.

6.3 Data and methods

The GloFAS forecast system is driven by ECMWF ensemble (ENS) forecasts. The ECMWF ENS consists of an ensemble of 51 forecasts with a current spatial resolution of 18km; one unperturbed control member and 50 perturbed members whose initial states and model physics have been perturbed to explore the currently understood range of uncertainty in the observations and the model. The initial condition uncertainty is sampled using perturbations from an ensemble of data assimilations (EDA) and perturbations constructed from the leading singular vectors, while the model uncertainties are represented with the Stochastically Perturbed Parameterization Tendencies scheme (SPPT) that simulates the effect on forecast uncertainty of random model errors due to the parametrized physical processes (ECMWF, 2020b). The ensemble members provide a range of possible future weather states, with the methods designed to ensure that each ENS member is truly independent of all the others (Owens and Hewson, 2018).

From the set-up of the GloFAS forecast system it is possible to assess the performance of each component of the flood forecast chain (track, intensity, precipitation and discharge) within each ensemble member from each successive forecast in the lead up to the flood event. Comparing the performance of the ensemble forecasts between components and from run to run allows for an assessment of the links between the predictability of each of the different forecast components

(Figure 6-2), with a particular focus on how the discharge forecast skill relates to the errors in the TC track, intensity, and precipitation forecasts. Table 6-1 summarizes the forecast and verification data that was sourced for each of the components in the forecast chain. It also describes the verification metric that was selected and calculated for each ensemble member from each forecast run. Further detail on each component and the evaluation of the predictability linkages is given in the following sections.

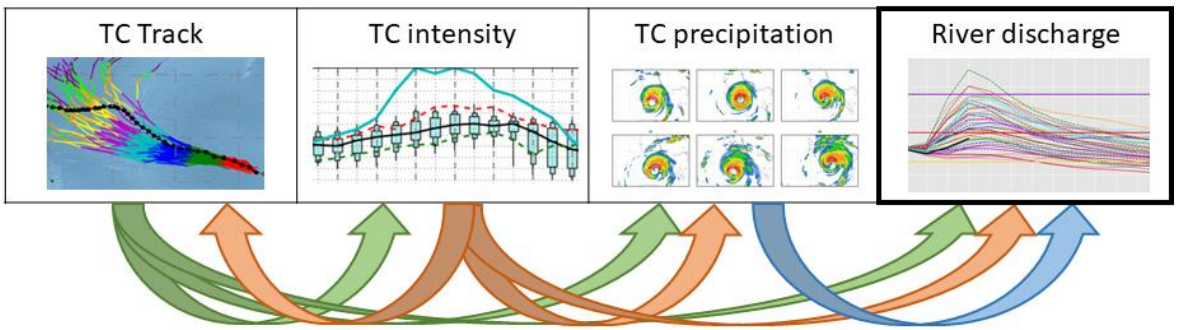


Figure 6-2. Conceptual diagram illustrating four of the main components of the TC river flood forecast process, along with the potential links in predictability between them. The boxes contain the four aspects of the forecast that are verified in each ensemble member in each forecast run in the lead up to the flood event, while the arrows represent the potential of each component to influence the skill of the others (green arrows = from TC track; orange arrows = from TC intensity; blue arrows = from TC precipitation). It is these links and dependencies that are evaluated in this paper (e.g. each arrow has a matching scatter plot in Figure 6-8).

	TC Track	TC intensity	TC precipitation	River discharge
Forecast data	ECMWF ENS tracks from TIGGE CXML archive (NCEP et al., 2008)	ECMWF ENS tracks from TIGGE CXML archive (NCEP et al., 2008)	ECMWF ENS precipitation forecast data (0.2°x0.2° resolution) (ECMWF, 2022)	Global Flood Awareness System (GloFAS) (v2.1, 0.1°x0.1° resolution), (Zsótér et al., 2019)
Verification data	International Best Track Archive for Climate Stewardship (IBTrACS) (Knapp et al., 2010, 2018)	International Best Track Archive for Climate Stewardship (IBTrACS) (Knapp et al., 2010, 2018)	GPM IMERG (final version, 0.1°x0.1° resolution) precipitation data (Huffman et al., 2019)	GloFAS-ERA5 operational global river discharge reanalysis (0.1°x0.1° resolution) (Harrigan et al., 2019, 2020)
Verification metric	• Track error at landfall • Track error in the first 24 hours over land	Error in mean sea level pressure minima at landfall	Mean absolute error of storm total precipitation	Modified Kling-Gupta efficiency (KGE) (Kling et al., 2012) computed for key river points and across all severely impacted river points

Table 6-1. A summary of the forecast data, verification data and metrics computed for each component of the TC flood forecast chain.

6.3.1 *Tropical cyclone track (position and intensity)*

1) FORECAST DATA

The forecasted TC track position and intensity information from each of the 51 ECMWF ENS ensemble members were extracted from the TIGGE CXML archive (NCEP et al., 2008), for each 00 UTC forecast run from 8th to 17th November 2020.

2) VERIFICATION DATA AND METHOD

The verification data for TC track position and intensity for Hurricane Iota were obtained from the International Best Track Archive for Climate Stewardship (IBTrACS) (Knapp et al., 2010, 2018). Forecast TC tracks are deemed to be matched with Hurricane Iota if any point on the track is within 500km of a matching observed position from IBTrACS. All points from that track are then included in the analysis providing their forecast position is within 1,000km of the matching observed position. The TC track error in each of the ensemble members with a matching TC was then calculated as follows: i) the landfall track error, defined as the distance between the observed and forecast position, in km, at the time of the last pre-landfall observed track point (00 UTC 17th November 2020); ii) the average track error across land; in the Iota case this is defined as the average of the distances (in km) from each of the 6-hourly forecast and observed track positions in the first 24 hours that the storm was over land, to ensure that storm tracks are available for most

members. The intensity error calculated is the error in mean sea level pressure (mslp) minima at the last pre-landfall point, prior to any weakening on landfall.

6.3.2 *Tropical cyclone precipitation*

1) FORECAST DATA

ECMWF ENS precipitation forecast data for each 00UTC forecast run from 8th to 17th November 2020 was extracted from the ECMWF archive (ECMWF, 2022). The storm total precipitation was calculated as being the 96-hour accumulation from 00 UTC 16th to 00 UTC 20th November within 500km of the Hurricane Iota track. This represents the area of precipitation defined as being associated with the cyclone, as used in multiple studies that assign a set radius when defining TC-related rainfall (Prat and Nelson, 2013, 2016; Jiang et al., 2011; Luitel et al., 2018; Titley et al., 2021). A land/sea mask was applied to focus on precipitation over land, as relevant for river flooding.

2) VERIFICATION DATA AND METHOD

The observed precipitation for the matching storm total 96-hour period was calculated from the Integrated Multi-satellitE Retrievals for GPM (IMERG) (Huffman et al., 2019), selecting the IMERG Final Run data. The IMERG Final Run product, recommended for research uses, is available around 3 months after valid time and incorporates additional microwave passes and a monthly gauge correction. The same masks (land/sea and within 500 km of the observed track) were applied as in the forecast fields. Prior to verification, the IMERG data is re-gridded to the model grid using linear interpolation. The Mean Absolute Error (MAE) was then calculated for each ensemble member by comparing the storm total precipitation in the forecasted and observed fields.

6.3.3 *River discharge*

1) FORECAST DATA

The set-up of the operational GloFAS forecast system at the time of Hurricane Iota (v2.1) (Alfieri et al., 2013; Harrigan et al., 2023), utilised ECMWF ENS meteorological forecasts coupled with the land surface module HTESSEL (Balsamo et al., 2011) and the Lisflood hydrological routing model (Van Der Knijff et al., 2010), to produce 51 possible forecast evolutions of the streamflow through the global river network. GloFAS version 2.1 forecasts (Zsótér et al., 2019) were downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) for each 00 UTC forecast run from 8th to 17th November 2020.

2) VERIFICATION DATA AND METHOD

Ongoing difficulties in the availability of reliable, spatially distributed, river discharge observations from river gauges (Lavers et al., 2019) mean that reanalyses provide the only practical source of

river discharge verification data for the purposes of this study. Therefore, the verification dataset utilised in this study is the GloFAS-ERA5 operational global river discharge reanalysis which accompanies GloFAS v2.1 (Harrigan et al., 2019, 2020). An objective analysis of the GloFAS forecasts is carried out by calculating the Modified Kling-Gupta efficiency (KGE') (Gupta et al., 2009; Kling et al., 2012) to compare the forecast and verification discharge data. The KGE' is an expression of the distance away from ideal model performance as described by three components (correlation, variability bias and mean bias). KGE' = 1 indicates perfect agreement between simulations and observations. The KGE' score for a mean flow benchmark is KGE'≈-0.41 (Knoben et al., 2019). In our study the KGE' is calculated for the 1-to-15-day forecasts from all GloFAS ensemble members for key river points and for all severely impacted river points across the region (defined as those points with an upstream area greater than 1,000km^2 that experienced greater than a 5-year return period during the event). The return period thresholds used in the study were as adopted within the GloFAS web interface, calculated by fitting a Gumbel extreme value distribution to the GloFAS-ERA5 river discharge reanalysis over the 1979-2018 period (Alfieri et al., 2019; Harrigan et al., 2023; Zsótér et al., 2020).

6.3.4 *Proposed methodology to evaluate the predictability linkages along the forecast chain*

Once all of the forecast components have been verified for each ensemble member at each forecast run, the key part of the proposed methodology is to use this information to analyse the links between forecast track, intensity, precipitation and discharge skill across hundreds of sample forecasts for a particular high-impact case; in this case Hurricane Iota. Several steps are carried out as follows:

i. Box and whisker plots are used to represent how the track, intensity and precipitation errors and the discharge KGE' scores (across all ensemble members) compare in terms of their mean and spread for each forecast run in the lead up to the event.
ii. A “divergence” score (Richardson et al., 2020) is also calculated to allow a comparison of the run-to-run differences between each component of the forecast. In this score the difference between two ensembles is calculated using the divergence function associated with the Continuous Ranked Probability Score (CRPS) (Gneiting and Raftery, 2007). In our case the two ensembles we are comparing are the ensemble member scores from each component on two successive runs, with the divergence score indicating how much the forecast performance is changing from run to run.
iii. Comparing the divergence scores from each part of the forecast (the track, intensity, precipitation and discharge components) allows us to assess how closely matched the run-to-run jumps in forecast performance are between each component.

iv. The correlation between the individual ensemble member scores for each component from each forecast lead time is also calculated and displayed alongside scatter plots to show how closely each aspect of forecast performance is tied to each other.

v. 'Good' and 'bad' members are identified from their performance in each part of the forecast chain and their performance in other components examined, in order to delve deeper into the controls on predictability and how these vary from place to place and from run to run.

6.4 Results

6.4.1 *Evolution of forecasts for Iota*

A summary of the TC position and intensity, precipitation and discharge forecasts for Hurricane Iota is shown in Figure 6-3. At seven days ahead of landfall around two thirds of members are forecasting a TC to develop, but with a very large spread in forecast position. By five days ahead of landfall over 90% of members are forecasting a TC, and by three days ahead the forecast storm location is much more confident. The intensity of Hurricane Iota is underpredicted at all lead times, as the forecast models struggled to resolve the rapid intensification process. The observed mean sea level pressure (mslp) minima at the pre-landfall point was 918mb, but even at a one-day lead time the minimum mslp forecasted by any of the ensemble members was 951mb (Figure 6-3 top row). Despite this, the heavy rainfall along the north coast of Honduras is predicted with high probabilities (> 60%) around 5 days ahead of landfall (Figure 6-3 middle row), but heavy rainfall close to the landfall location on the east coast of Nicaragua was not predicted with high probabilities until later forecast runs closer to the time of landfall. This suggests that the heavy rainfall close to the landfall location was more reliant on having a more accurate forecast for the landfall location and intensity, which was only possible at shorter lead times. The northward bias of some earlier runs (see top row Figure 6-3) still produced forecasts of significant rainfall for Northern Honduras.

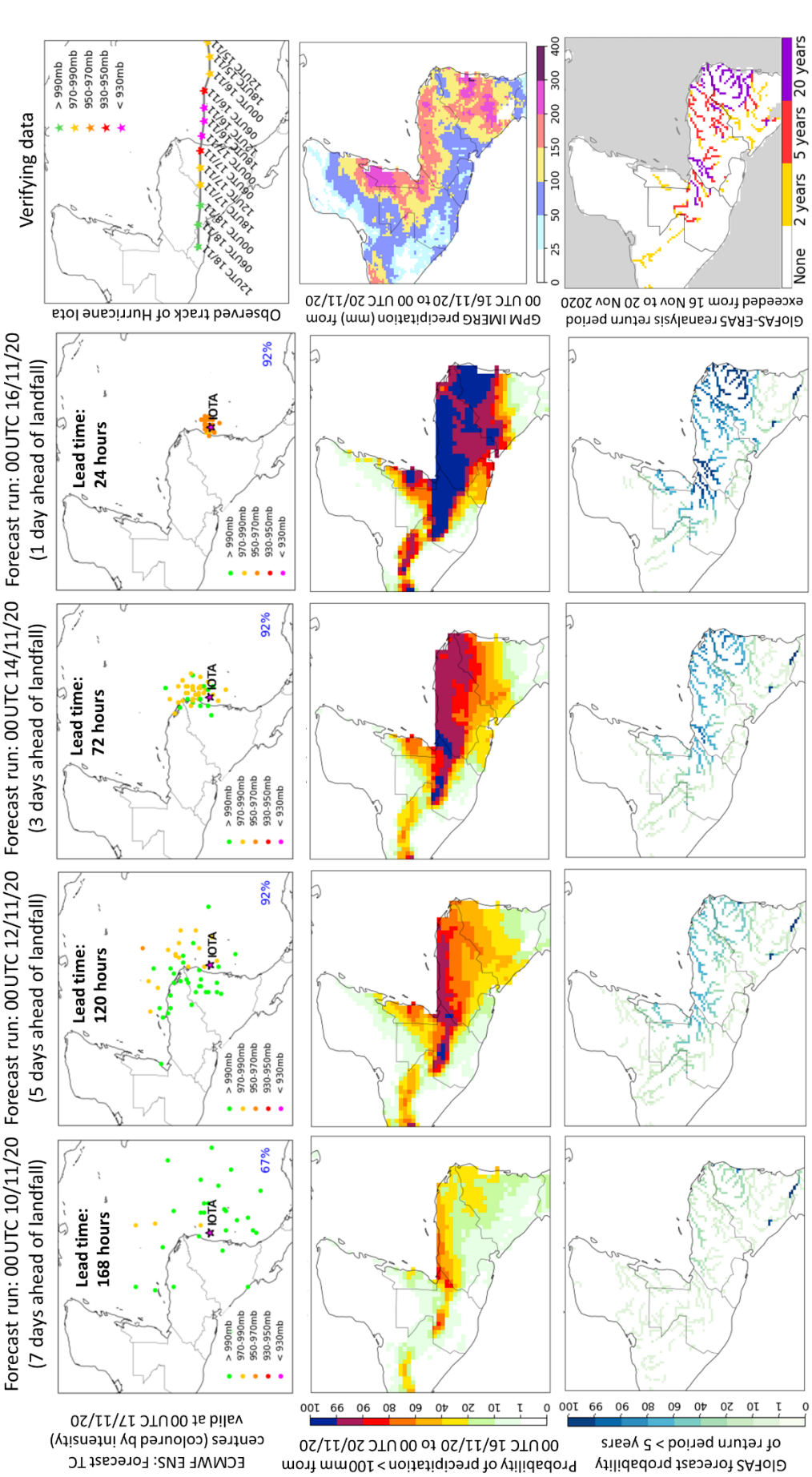


Figure 6-3. Hurricane Iota cyclone position (top row), rainfall (middle row) and flood (bottom) forecasts from the forecast runs of ECMWF ENS and GloFAS in the lead up to the event. The first four columns show the forecast runs from 7, 5, 3 and 1 days prior to landfall, while the right column shows the corresponding verification data for track, rainfall, and flood return period exceedance. In the top row the star shows the observed centre position of TC Iota at 00UTC 17/11/2020 (just prior to landfall) and the dots show the forecast positions valid at this time from each ensemble member, coloured by the mean sea level pressure minima. The percentages indicate the percentage of members that had a matched cyclone centre forecast at this lead time. The second row shows the precipitation probability forecasts from EC ENS for 4-day event precipitation totals > 100mm. The bottom row shows the GloFAS forecast probability for river discharge exceeding the 5-year return period on each river point.

A signal for significant flooding appears 5 days ahead of landfall (Figure 6-3 bottom row), with probabilities of 5-year flood return periods being exceeded being greater than 40% across a large part of the region. This coincides with the increase in confidence in the TC position and rainfall accumulations. However, the forecast confidence varies from river to river, and even one day prior to landfall not all the areas simulated to have exceeded the 20-year return period (the highest threshold displayed in GloFAS) have high probabilities. For example the Río Grande de Matagalpa in Central Nicaragua (the most southerly river that experienced return periods greater than 20 years as shown in the bottom right plot in Figure 6-3) only has a probability of 40-60% of exceeding a 5-year return period, which is much lower than the other impacted rivers. Although the rainfall forecast one day prior to landfall is correctly identifying two main areas of high precipitation, on the north coast of Honduras and near the landfall location in Nicaragua, the area near the forecast track is displaced too far to the north, and the forecast has missed the extension of the rainfall area to the southwest of the TC track. These results indicate that the factors controlling the forecast skill for the flood forecasts for Iota vary with lead time and scale. Once the forecast confidence increased for the TC genesis and track of Iota (around 5 days ahead of landfall) there is a clear signal for a significant flood event in the overall region that was affected. At the catchment scale, the detailed structure and precipitation forecasts for the storm influence how well the flooding is forecast on each individual river, even at shorter lead times.

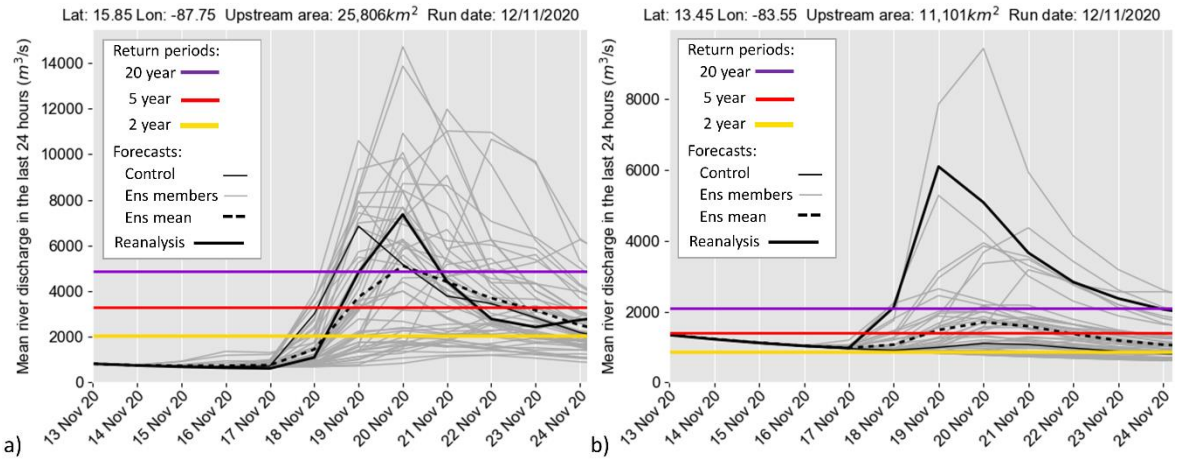


Figure 6-4. GloFAS forecast hydrographs from 00UTC run on 12th November 2020 for a) River Ulúa and b) River Prinzapolka (ensemble members: grey; control member (unperturbed): thin black; ensemble mean discharge forecast: black dashed). The yellow, red and purple horizontal lines show the 2-, 5- and 20-year return period discharge levels for that river. The reanalysis discharge is overlain in a thick black line.

Two of the rivers which experienced the most severe flooding from Hurricane Iota were the River Ulúa in northern Honduras and the River Prinzapolka in eastern Nicaragua. To examine the river discharge forecasts for these rivers, the GloFAS outlet river point was selected for each river:

15.85N, 87.75W for the Ulúa and 13.45N, 83.55W for Prinzapolka, with upstream areas of 25,806km^2 and 11,101km^2 respectively (see annotated map in Figure 6-1 for the locations and catchment areas). GloFAS forecast hydrographs for these points from 5 days prior to landfall are shown in Figure 6-4 and illustrate that the majority of members were predicting severe flooding on the Ulúa, whereas only a small number of members predicted the extreme discharge levels experienced on the Prinzapolka. This shows the sensitivity of rivers closer to the landfall location to the precise forecast of the landfall, as these were areas affected by the heavy rainfall close to the centre of the storm. The upstream area of the Prinzapolka site is also smaller, resulting in a faster response to heavy rainfall compared to the Ulúa site, and increased sensitivity to accurate timing of the landfall.

6.4.2 *Flow of uncertainty and predictability through the forecast chain*

The previous section presented the evolution of forecasts for Hurricane Iota, identifying potential linkages between the different forecast components. To investigate the flow of uncertainty and predictability more thoroughly, a detailed analysis comparing the ensemble performance at each part of the forecast chain is proposed and illustrated. Forecast performance across the ensemble members is presented side-by-side for each component of the forecast: track error (Figure 6-5a), intensity error (Figure 6-5b), precipitation error (Figure 6-5c), and KGE' score for the discharge forecasts (6-5d) across all severely impacted river points as well as the two rivers that were examined in Figure 6-4 (the Ulúa and the Prinzapolka). Flooding on the River Ulúa was well predicted further in advance than the flooding on the River Prinzapolka, which was closer to the landfall location, confirming the findings in Section 6.4.1. Using the Towner et al. (2019) thresholds to evaluate the KGE' scores ("Good" (KGE' $\geqslant$ 0.75), "Intermediate" (0.75 > KGE' $\geqslant$ 0.5), "Poor" (0.5 > KGE' > 0) and "Very poor" (KGE' $\leqslant$ 0), the majority of ensemble members have "intermediate" performance 5 days ahead of landfall for the River Ulúa, but only 2 days ahead of landfall for the River Prinzapolka. An examination of IMERG satellite data (Figure 6-6) reveals that the Ulúa catchment rainfall was mainly associated with onshore flow of the outer rainbands with orographic enhancement, which fell over a longer period and is less sensitive to the exact landfall location. For the River Prinzapolka, the centre of the storm passed directly over the catchment leading to very heavy rainfall in a shorter period during and after landfall. Successful flood forecasts therefore relied on the accurate prediction of landfall location and of the heavy rainfall close to the TC centre, which was less predictable at longer lead times. The average KGE' score across all impacted points (all river points where the 5-year return period was exceeded in the reanalysis) falls between these two extremes.

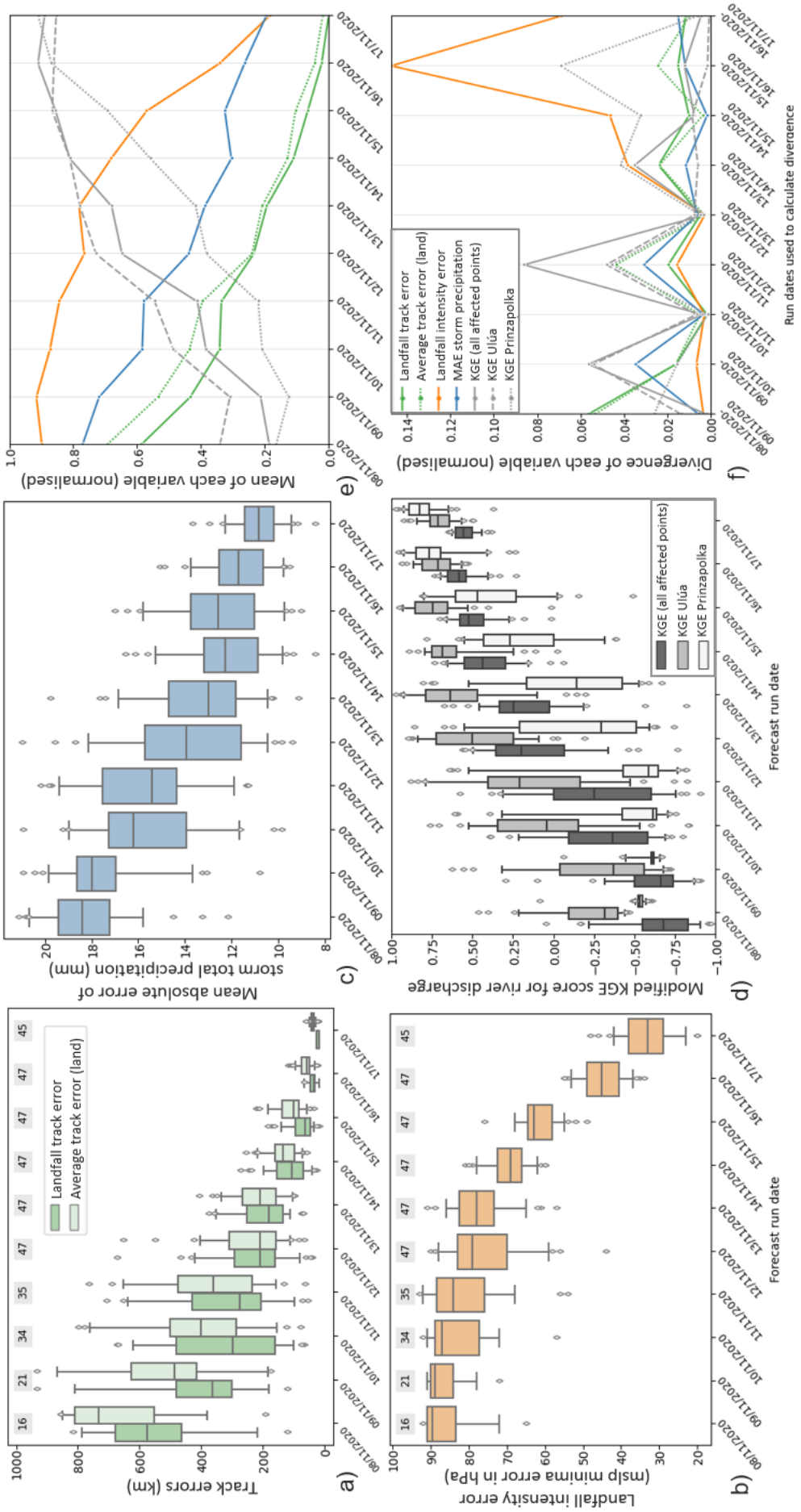


Figure 6-5. Box plots comparing forecast performance across the ensemble members for each 00UTC forecast run in the lead up to the Hurricane Iota event: a) forecast track error (landfall error and average track error over land); b) forecast landfall intensity (mslp minima) error; c) forecast precipitation error (mean absolute error of the storm total precipitation (within 500km of TC track); d) Discharge forecast performance as measured by the average KGE' score at all severely impacted river points (where 5-year return period is exceeded in the GloFAS reanalysis), and at the River Ulúa / River Prinzapolka outlet river points; e) a comparison of the means of these metrics, normalized by the y-axis range in each subplot and colour-coded from the plots and key as displayed in (f); f) divergence score of these metrics, normalized by the y-axis range and colour coded as shown in the key. In (a) and (b), errors can only be computed for those ensemble members with a tracked TC (numbers overlain in grey).

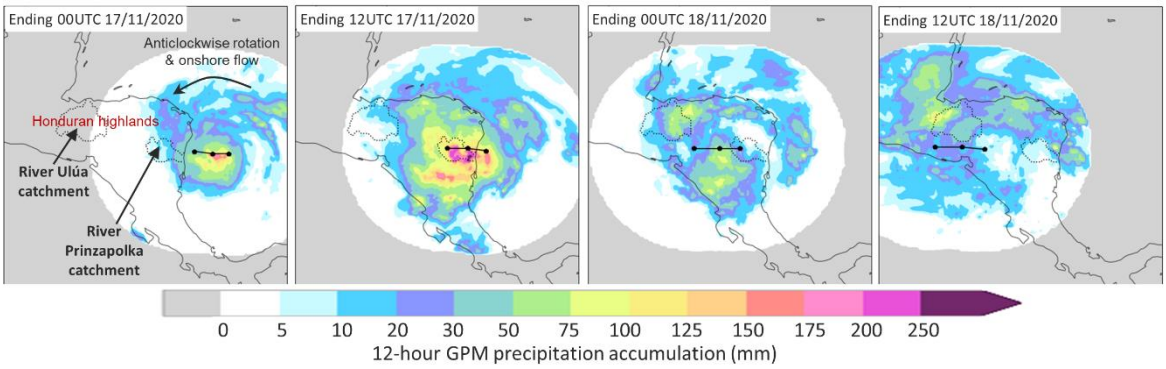


Figure 6-6. Precipitation accumulation across Central America from Hurricane Iota in four 12-hour periods from 12 UTC 16th November 2020 to 12 UTC 18th November 2020. The river catchments for the River Ulúa and River Prinzapolka are shown in dashed lines, with labels on the left plot.

The right hand two plots on Figure 6-5 then compare the means of these scores (Figure 6-5e), and the divergence score (i.e. change in performance from run to run (Figure 6-5f)). The idea is to objectively study how closely the skill of each part of the forecast is linked to the others. The biggest change in the KGE' of all river points (from 11th to 12th) coincides with changes in the precipitation, landfall position, and intensity errors, and in particular to the biggest change seen in the average track error over land, showing the importance of verifying and improving forecasts of TC track after landfall. Comparing the distribution of the track errors from 11th to 12th in Figure 5a shows there was both an overall reduction in track error and a significant reduction in the spread. But although there are matching improvements in the precipitation and discharge scores, the spread is not significantly reduced and remains relatively high until very short lead times, showing there is greater variation in precipitation and discharge forecasts even after the track forecasts become more confident. For the River Prinzapolka, which was close to the landfall location, the biggest jump in KGE' scores coincided with the biggest change in landfall intensity error (from 15-16th, where the forecasts begin to pick up on the ongoing intensification of the storm). This large improvement in the intensity forecast in these short-range forecasts has only a modest impact on the track, precipitation and overall discharge forecast scores, but has a large impact for the KGE' score for the River Prinzapolka. As the River Prinzapolka catchment is much closer to the landfall point, the rainfall it experiences is predominately from the area close to the centre of the storm e.g. the eye wall, and the intensity of the precipitation in this part of the storm is greater for intense storms (Rios Gaona et al., 2018).

6.4.3 *Member-specific analysis*

The consecutive forecast runs on 11th and 12th November were found in Figure 6-5f to have high divergence scores i.e. significant run-to-run changes in ensemble performance. The increase in forecast confidence from 11th to 12th November in Iota's position forecast the day after landfall is notable with a much more concentrated pattern of ensemble member TC positions (Figure 6-7). The additional member-specific analysis here illustrated in Figure 6-7 is to colour the forecast positions by their KGE' scores on two rivers points that have been focused on: the Ulúa (left) and the Prinzapolka (right) outlet points. This allows us to provide an impact-relevant hydrological assessment of which ensemble members provided the most accurate flood forecast for each river (those coloured in blue).

Both before and after the increase in forecast confidence, the members with a good forecast for the River Prinzapolka all have the TC forecast centre relatively close to the observed location, although there are also several ensemble members with forecast centres close to the observed location that don't have good KGE scores. A broader range of ensemble members based on TC position perform well for the Ulúa catchment. The TC centres in Figure 6-7 are sized according to the MAE in storm total precipitation to allow an inspection of the relationship between three aspects of the forecast: the TC track (position), precipitation, and river discharge skill. The ensemble members with better overall precipitation forecasts (lower MAE), shown as larger dots, are mainly those that are forecasting the TC centre to be close to its observed position. Several of the ensemble members with TC positions too far to the north are shaded blue in the left plots as they score well for river discharge on the Ulúa, but they have a small size indicating they do not score well overall for precipitation. Their better score for the Ulúa discharge is partly due to their more northerly tracks bringing forecasts of heavy rain from the central part of the storm over the Ulúa catchment. The centre of the storm tracked further south in reality and so the overall precipitation errors in these members are higher as they fail to forecast the precipitation further south.

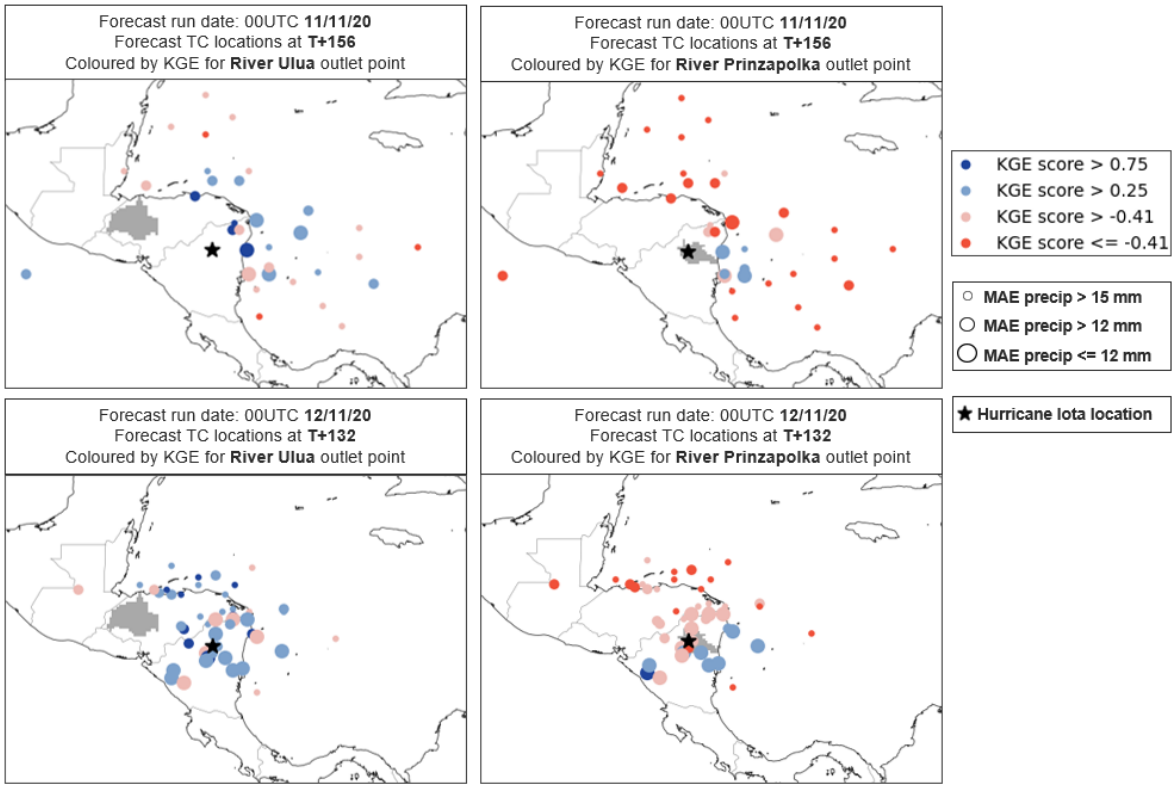


Figure 6-7. Forecast TC locations in the ECMWF ENS forecasts from 00UTC 11th November 2020 (top row) and 00UTC 12th November 2020 (bottom row). The centres are coloured according to the modified KGE score for the River Ulúa (left column) and River Prinzapolka (right column) outlet river points as defined in Section 6.4.1. The centres are sized by the mean absolute error in the storm total precipitation field (larger circles = better forecasts). The verifying position of Hurricane Iota at this time is marked with a black star, and the catchment areas for the two river outlet points are shown in grey for reference.

To take this member-specific analysis a step further, the discharge scores from each ensemble member from a particular run can be ranked to determine the 'best' performing members. This allows us to examine the other aspects of forecast performance for those members, to better understand why they led to good flood forecasts. This is illustrated for the 00UTC 12th November run in Figure 6-8. Interestingly in this case, the landfall position of all three of the identified 'best' members is similar, and is close to the observed landfall location. However, their forecast tracks after landfall are different, with the best performing member for the River Ulúa being closest to the observed track (Figure 6-8a), while the best member for the River Prinzapolka turns southwest shortly after landfall, and therefore underestimates precipitation totals further north (Figure 6-8b). Despite their similar landfall locations, the forecast precipitation patterns are very different, with large variation in the precipitation totals even close to the landfall location in all three. When comparing to the verifying IMERG data, the member that had the highest KGE' score when averaged over affected river points across the region (Figure 6-8c) had the best match in terms of the overall extent of the rain event, but was forecasting significantly greater totals in several areas compared

to IMERG. However, rain totals of up to 510mm were recorded locally (Stewart, 2021), indicating that IMERG may be underestimating the most extreme rainfall totals. All three of the 'best' members from this forecast run failed to significantly intensify the storm. Member 9 (Figure 6-8c) was the strongest at landfall (986mb compared to 1001 and 1006mb respectively for the other two 'best' members), which may have contributed to the larger forecast rainfall totals nearer the landfall site, but it was still much weaker than observed (918mb). This analysis shows that the ensemble forecasts of TC rainfall totals depend on much more than having an accurate landfall location, and can vary significantly even between members with similar track forecasts, emphasizing the need for an increased focus on evaluating and understanding ensemble TC rainfall forecasts.

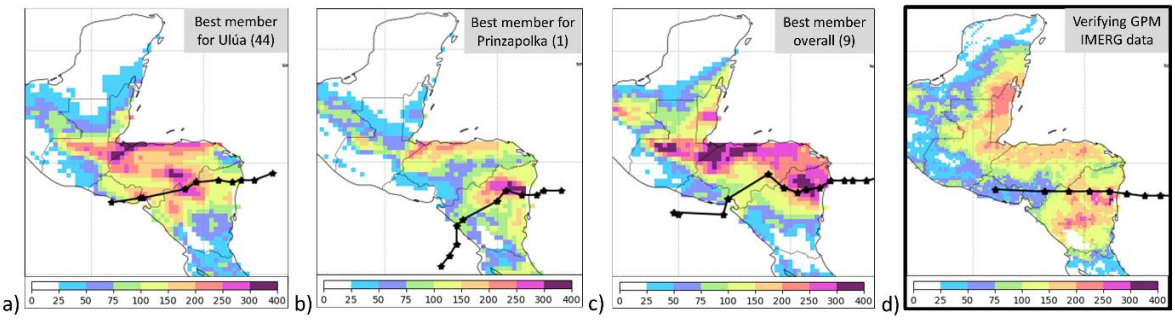


Figure 6-8. Hurricane Iota forecast track and forecast event precipitation forecast totals (in mm) from the 00UTC 12/11/2020 forecast run from the EC ENS ensemble members with the highest KGE' scores for a) River Ulúa outlet point, b) River Prinzapolka outlet point, and c) the average of all severely impacted river points; d) displays the verifying GPM IMERG precipitation (in mm) along with the observed track of Hurricane Iota.

The correlation of the member scores between different aspects of forecast performance helps to quantify the degree to which the forecast performance aspects are tied to each other (Figure 6-9). The average track error over land was selected as the track measure to plot as initial investigation (not shown) found a stronger correlation between this measure and both the precipitation error and hydrology skill (KGE') compared to against the landfall error. This further emphasises the earlier finding from Figure 6-5 of the importance of track after landfall to the wider impacts in the region. Each scatter plot in Figure 6-9 includes the ensemble members across all lead times allowing us to examine the link between each forecast component using a large sample of hundreds of forecasts. All the relationships have a significant correlation when all lead times are included (shown in blue), with the strongest correlations being between the track error and the river discharge skill, and the precipitation errors and the river discharge skill. Later forecasts have lower errors / greater skill, and reduced spread in the scores, but the shape of the relationship varies considerably between plots. For example, the points in the track vs discharge and rainfall vs discharge plots (Figure 6-9c and 6-9f) are mainly clustered along the diagonal, whereas in the intensity vs discharge plot (Figure

6-9e) the points are spread across the upper right half of the plot. This indicates that good forecasts of river discharge are possible even with high intensity errors, but not with high rainfall or track errors.

Looking at the correlations at each forecast run time (in tables on Figure 6-9, with average of these in green) shows how predictability links between each component for particular forecast runs. The correlation between track and intensity error is not significant at any lead time (Figure 6-9a). There are however significant correlations between track and precipitation, and track and river discharge at all forecast times other than the latest two, by which point the track errors were very small (Figure 6-9b and 6-9c). Interestingly, the correlation between the track errors and river discharge skill is stronger than track errors against the precipitation errors, which may be because discharge skill depends on the total catchment rainfall, where the smaller grid point to grid point differences that influence the precipitation errors are balanced out to some degree. The slightly lower correlation with precipitation may also be due to the independence of the precipitation verification data, as opposed to the model-based reanalysis data that is used to verify the discharge forecast, and may also be impacted by potential underestimates of the rainfall totals by GPM IMERG. The intensity error at landfall is not significantly correlated with river discharge skill at any forecast lead time (Figure 6-9e). Although there is a positive correlation overall between the intensity error and rainfall error, there is actually a counter-intuitive negative correlation at three of the forecast runs, and when you take an average of the correlations at each forecast run (Figure 6-9d). So, in a particular forecast cycle, the members with better forecasts for TC intensity do not tend to be those with better precipitation forecasts or river discharge forecasts. There are however significant correlations as expected between the precipitation error and the river discharge score (Figure 6-9f). It is therefore crucial to verify and understand the skill of ensemble TC rainfall forecasts, and focus future research on understanding and improving rainfall prediction in addition to track and intensity in order to advance flood forecasting in TC cases.

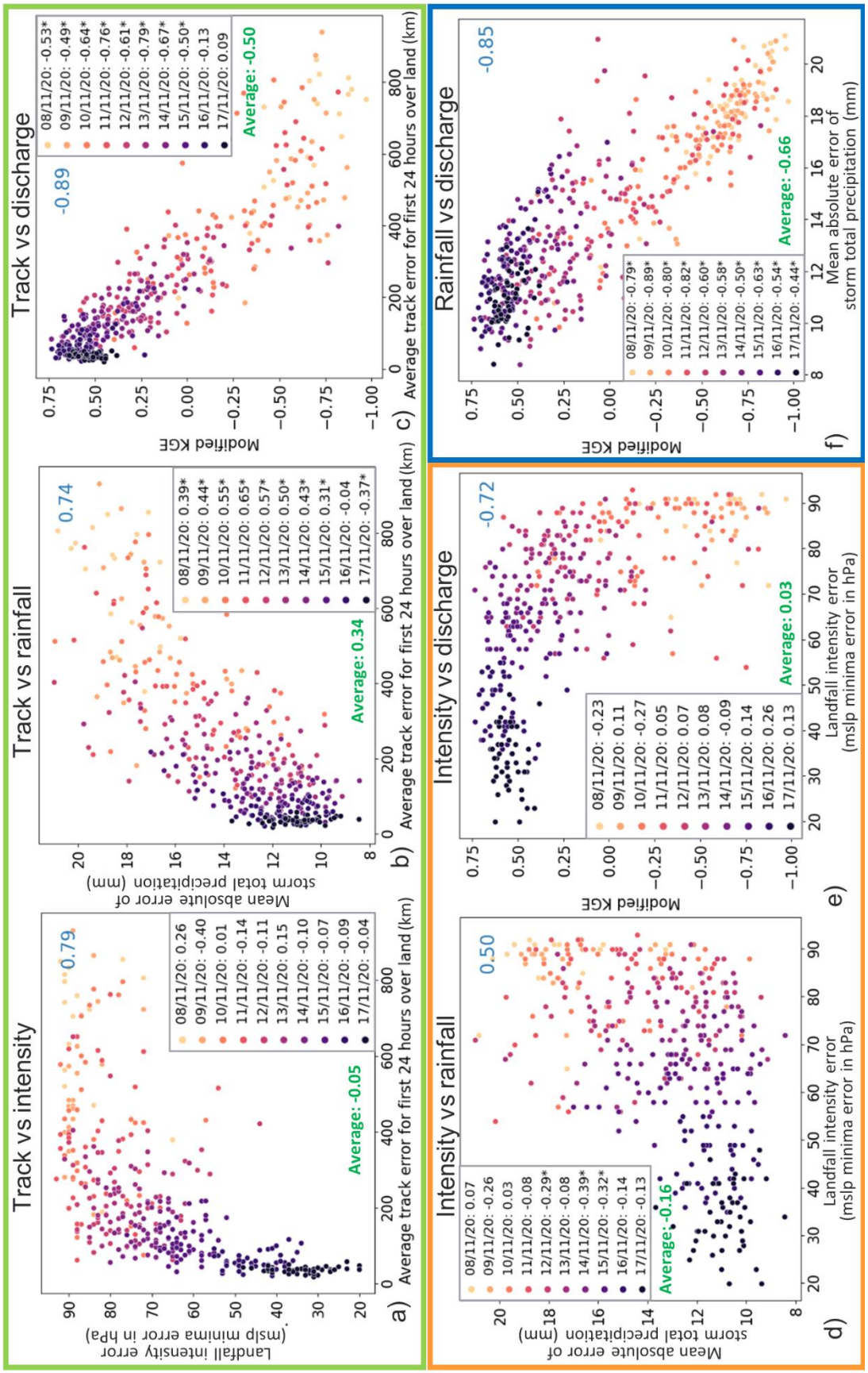
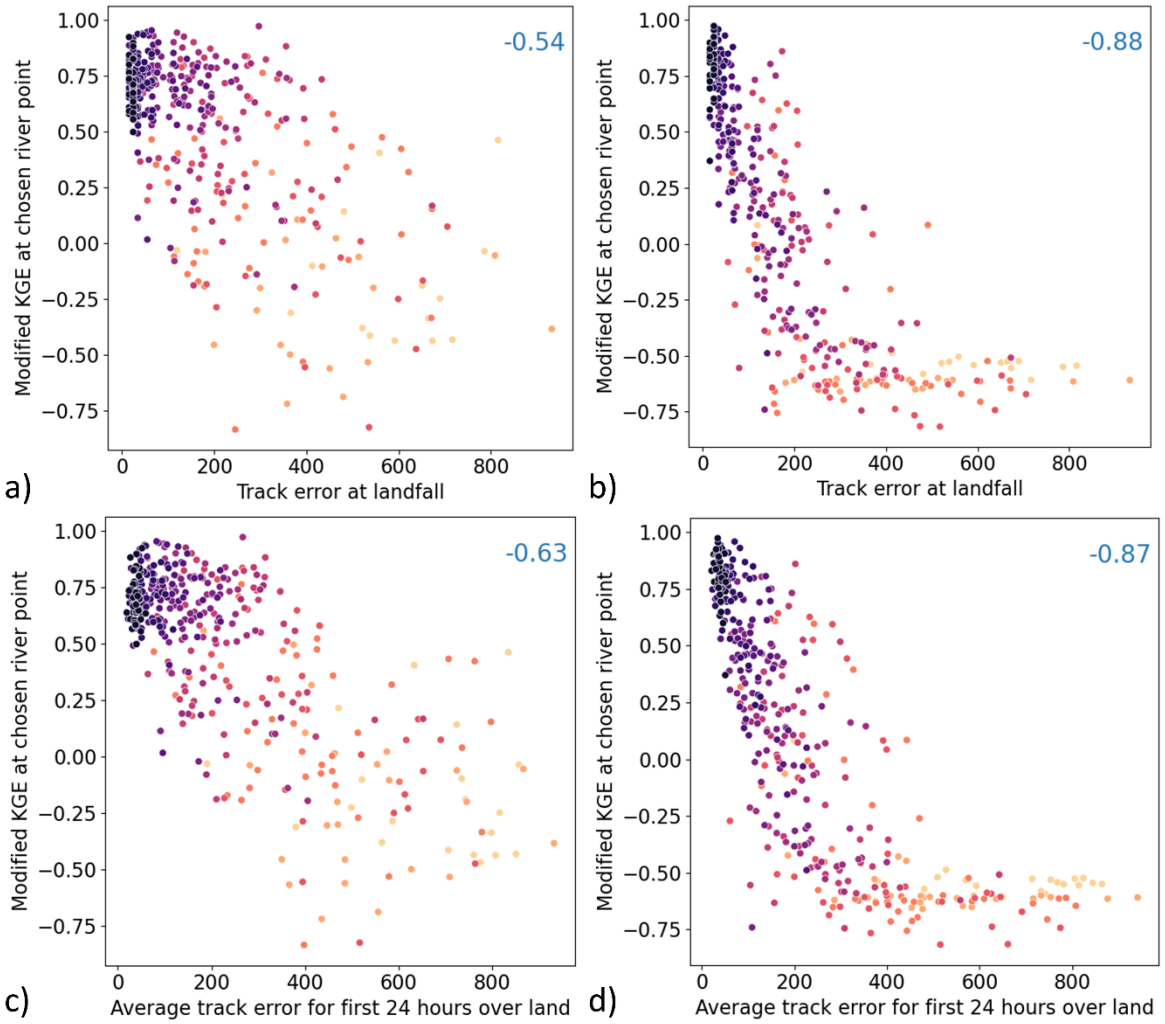


Figure 6-9. Scatter plots of each error type / score where each dot is one member: a) Average track error over land vs landfall intensity error; b) Average track error over land vs MAE of storm total precipitation; c) Average track error over land vs average KGE' score at all severely impacted river points (where 5-year return period is exceeded in the GloFAS reanalysis); d) landfall intensity error vs MAE of storm total precipitation; e) landfall intensity error vs average KGE'; f) MAE of storm total precipitation vs average KGE'. Overall Spearman correlations are shown on each plot in blue. Members are coloured by their forecast run and the Spearman correlation for each forecast run is shown in the legend, with a star indicating when this is significant (<0.05). The average of these forecast run correlations is shown in green. The green, orange and blue outlines show those relationships that were represented by this colour arrow in Figure 7-2.



e)

Forecast run (00 UTC)	a) Landfall track error vs KGE River Ulúa	b) Landfall track error vs KGE River Prinzapolka	c) Land track error vs KGE River Ulúa	d) Land track error vs KGE River Prinzapolka
08/11/20	-0.17	-0.29	-0.23	-0.36
09/11/20	-0.32	-0.25	-0.32	-0.27
10/11/20	0.11	-0.37	-0.23	-0.51
11/11/20	-0.51	-0.72	-0.65	-0.62
12/11/20	-0.03	-0.67	-0.06	-0.57
13/11/20	-0.22	-0.73	-0.33	-0.62
14/11/20	-0.08	-0.78	-0.33	-0.41
15/11/20	-0.00	-0.54	-0.03	-0.48
16/11/20	0.05	-0.27	0.17	-0.10
17/11/20	0.00	0.00	-0.03	-0.28
Average of run correlations	-0.12	-0.46	-0.20	-0.42

Figure 6-10. Scatter plots of the various errors / scores where each dot is one member: a) Track error at landfall vs KGE' at the River Ulúa outlet point; b) Track error at landfall vs KGE' at the River Prinzapolka outlet point; c) Average track error over land vs KGE' at the River Ulúa outlet point; d) Average track error over land vs KGE' at the River Prinzapolka outlet point. Overall Spearman correlations are shown on each plot in blue. The Spearman correlations for each run time (and the average of these) are shown in the table in (e). Members are coloured by forecast run as shown in the table in (e).

Scatter plots of the two track metrics (landfall error and track error over land) against the KGE' river discharge score for the River Prinzapolka reveal a strong 'l-shaped' relationship, with a very quick drop off in forecast skill as track error increases, no forecast skill for higher track errors, and little difference between the plots for the two track error metrics (with just a slightly stronger correlation with landfall track error compared to the average track error over land both overall and in the run-to-run correlations) (Figure 6-10, b, d, and e). This shows how crucial an accurate landfall and track forecast was to a skilful flood forecast for this river. In the equivalent plots for the River Ulúa (Figure 6-10, a, c, and e), this l-shaped relationship is not observed and there is a more gradual drop in KGE as track error increases, and so an accurate landfall forecast was less crucial for this river. The correlation between track error and river discharge skill is lower compared to the Prinzapolka, and there is a stronger relationship with average land track error rather than landfall error (overall and in the run-to-run correlations), showing the greater influence of the detail of the forecast track after landfall on how much rain fell in the Ulúa catchment in northern Honduras.

6.5 Discussion

The ensemble-based methods developed in this study to analyse the cascade of predictability and uncertainty through the various stages of the TC flood forecasting chain provide useful information to both modellers interested in optimising flood forecast skill and to those who prepare and communicate flood forecasts with stakeholders and end-users in TC events.

The results showed that shifts in track skill do correspond to shifts in precipitation and discharge skill. The stronger relationship between discharge skill and the mean track error over land compared to the landfall location error (as indicated by the divergence score analysis and by the correlation scores), shows the importance of verifying and improving forecasts of TC behaviour after landfall, and crucially of not only focusing on the landfall location when communicating the forecasts in advance of a TC. The track after landfall is crucial, as the rainfall totals and flood extent and severity will vary depending on the path the cyclone takes after landfall and the translation speed at which it travels.

The large changes in TC intensity skill, which generally do not happen until shorter lead times, do not result in significant improvements in precipitation and discharge skill, apart from in catchments close to the landfall location that are impacted by the heavy rains around the eye wall. More widely it is the wider rainfall from the storm that is crucial to flood impacts, which is controlled more by the track (location and translation speed) and other factors such as the size of the storm and the orography than it is by the intensity forecasts (Titley et al., 2021). This also has important

implications for forecast communication, as it reinforces the need to focus on more than just the intensity category of the storm when forecasting flood impacts.

The average correlation between rainfall and river discharge skill is stronger than just track error against river discharge at most forecast runs and the analysis of a few strong performing members from one of the forecast runs reveals that even similar track forecasts can have very different precipitation patterns. This, combined with the finding that the spread in precipitation forecast errors across the ensemble remains relatively large even after the track spread reduces, highlights the importance of understanding and optimising precipitation forecasts in the ensemble models that are used to drive ensemble flood forecasts. Traditionally ensemble TC forecast verification at numerical weather prediction (NWP) centres has focused on track and intensity (Magnusson et al., 2019, 2021; Titley et al., 2020; Zhou et al., 2022b), and the relatively crude precipitation verification done here does not provide all the answers as to which aspects of the precipitation forecasts would benefit from improvement. More detailed verification of ensemble precipitation forecasts in TC cases is recommended, particularly using spatial or object-oriented methods to highlight which specific aspects of rainfall forecasts would benefit from improvement in order to feed into improvements in flood forecasts.

This study shows how closely the flood forecast skill is tied to both TC track and precipitation forecast skill and how uncertainty in these forecast aspects feeds through to uncertainty in the flood forecasts. However traditionally forecasts of precipitation and downstream flood hazards in TCs have used deterministic forecast models, leading to overconfident and unreliable flood predictions. The use of an ensemble flood forecasting system, such as the GloFAS system used in this paper, where ensemble forecasts are carried through the forecast chain to create probabilistic flood forecasts, are vital in representing the uncertainty in flood predictions, and preparing for a range of possible flood scenarios when there is an approaching TC. Some of the learning of this study can be directly applied in a flood forecasting capacity by providing additional context and advice to decision makers. For example, Figure 6-11 shows an example forecast TC centre plot that could be produced in the run up to an event to show the specific forecast tracks that were likely to lead to significant flooding in a particular river, by colouring the forecast centres by the discharge return period that is forecast to be exceeded. Being able to follow specific members through the forecast chain in this way could help in scenario planning and forecast communication as a supplementary tool alongside the probabilistic ensemble forecasts.

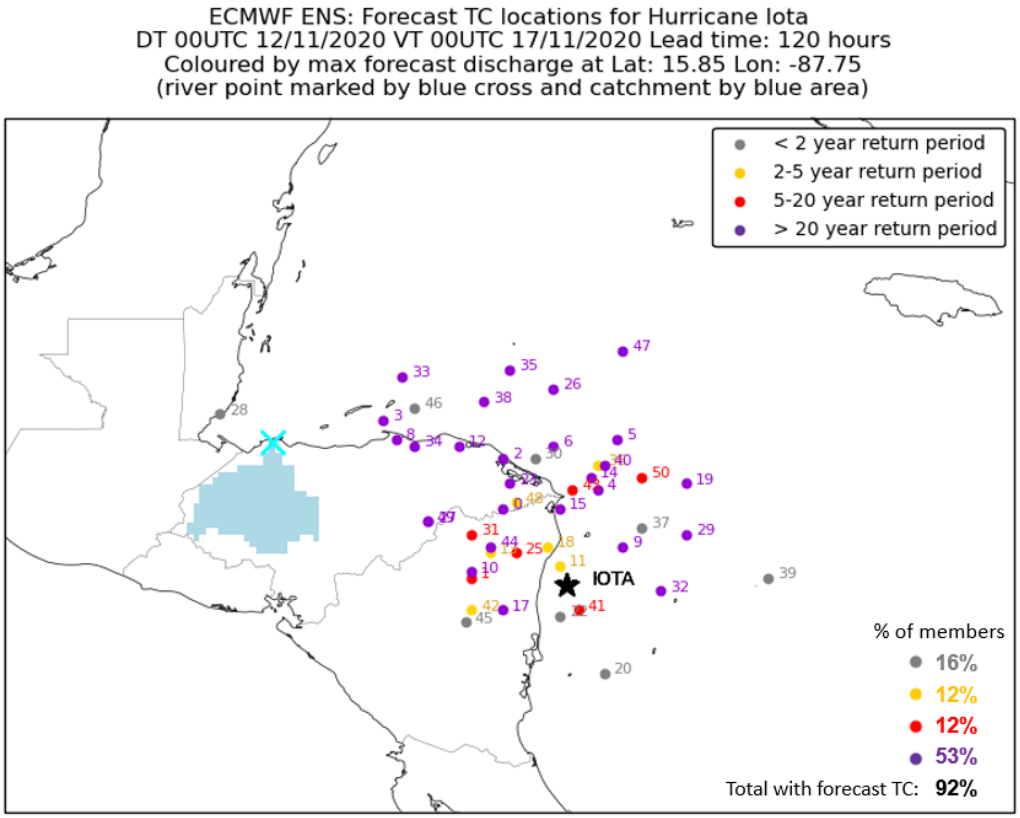


Figure 6-11. Prototype plot that could be produced in the run up to a TC event to highlight the forecast TC locations of those ensemble members that are forecasting a significant flood event at a particular river point. Here the outlet point on the River Ulúa has been used (cyan cross). The Day 5 ECMWF ENS forecast positions for Hurricane Iota from 00UTC 12th November 2020 forecast run are displayed (valid at 00UTC 17th November 2020 close to the time of landfall). The forecast positions are numbered by the ensemble member number (to allow the individual ensemble members to be followed through the forecast chain), and coloured specifically by the return period that is forecast to be exceeded on the River Ulúa. The thresholds and colours are the same as those used on the GloFAS web interface https://www.globalfloods.eu/.

The main strength of the techniques described in this study is that they allow for an analysis of the links between forecast track, intensity, precipitation and discharge skill across hundreds of sample forecasts for a particular high-impact case. We are using the ensemble as a larger sample of potential outcomes for what are rare events in any given location, and using this sample to investigate the interconnectivity of the various forecast components and give valuable insights as to the controls on TC flood predictability. All the data used in this study are freely available, meaning the analysis is repeatable and scalable, and can be applied to other cases in any TC basin around the world.

There are limitations to the methods as described here that could be improved in subsequent case studies. The use of reanalyses as the river discharge verification data set has the advantage of being available at every global river point, and further work is ongoing to address their known limitations (Harrigan et al., 2020; Zsótér et al., 2020), but if reliable discharge observations were more

widespread and openly available this would benefit studies of this kind. The rainfall verification metric used here is a very simple point-based metric and other metrics that remove the point-by-point sensitivity such as the Fractions Skill Score (Roberts and Lean, 2008) could be used. It would also be useful to calculate the forecast catchment-total rainfall in addition to the overall storm-total rainfall when analysing the forecasts for specific rivers. A further limitation of this study is the reliance on GPM IMERG data to verify the TC rainfall. GPM is an incredibly useful resource due to its global coverage and wide accessibility, and the Final Run data used here incorporates additional microwave passes and a monthly gauge correction. However, some of the ground-based observations in Iota indicate potential errors in capturing the true rain rates in some regions, which may impact the robustness of the analysis of the errors of the precipitation forecasts. More detailed investigation of biases in observations of TC precipitation data is recommended, comparing IMERG data with other sources of precipitation 'truth' data such as radar and gauge-based datasets, in order to develop better understanding of observation uncertainty and improve the confidence in TC rainfall verification results.

This study has focused on a particular case, Hurricane Iota, and the resultant flooding that occurred in Honduras and Nicaragua. In the event of a cyclone with a similar forecast track impacting the region, the findings of this study could help in giving greater confidence to longer-range flood forecasts for those catchments that are forecasted to be impacted by the wider storm circulation and orographically-enhanced rainfall, whilst understanding that the locations and severity of flooding closer to the potential landfall site will be more difficult to predict until closer to the landfall time. However, it is important to understand that even in the same region, no two cyclones are the same in terms of their track, speed, size and wider environmental conditions, and so care should be taken in applying the understanding gained in this study to future storms. The wider applicability of these results in other geographical contexts is also uncertain because flood severity also depends on the local landscape and catchment characteristics. The purpose of this study was to provide a suggested approach and set of methods to carry out an analysis of the flow of predictability through the flood forecast chain in TC cases. The wider application of these methods across a range of cases would help to inform a) the robustness and wider applicability of the findings of this case; b) how the flow of predictability through the flood forecast chain varies with different geographical and forecast contexts.

6.6 Conclusion

The work adds a new hydrological perspective to the traditional methods of verifying global ensemble TC forecast performance. It proposes the use of ensemble-based methods to analyse the

predictability links between the track, intensity, precipitation and discharge forecast components of the ensemble flood forecast chain in TC cases. These methods are illustrated using Hurricane Iota, which led to widespread and severe impacts from fluvial flooding in Central America in November 2020. We examined the performance of the ensemble members at ten forecast runs in the lead up to the flood event, and found strong correlations between the members' TC track error, precipitation error and river discharge KGE' scores. An analysis of the run-to-run changes in skill for each forecast component found that the largest run-to-run shift in forecast track errors corresponded to the largest change in the discharge KGE' scores.

Once the forecasts began to show increased confidence for the TC genesis and track of Iota (around 5 days ahead of landfall) there is a clear signal for a significant flood event in the overall region that was affected. But at the catchment scale, the study highlights how the sensitivity of discharge forecasts to the landfall track and intensity forecasts varies considerably depending on the catchment's location relative to the TC, with a greater sensitivity to precise landfall and intensity forecasts for river catchments close to the landfall location. The stronger relationship found between the flood forecast skill across the impacted region and the mean track error over land, when compared to only the landfall location error, shows the importance of predicting TC behaviour after landfall and not only focusing on landfall location when communicating TC forecasts. Changes in TC intensity only result in significant improvements in discharge skill for river catchments impacted by the heavy rains around the eye wall. The rainfall from the storm's wider circulation is crucial to flood impacts in most of the affected river catchments in the region, and is less sensitive to the TC intensity, showing the importance of focusing on more than just the category of the storm when forecasting flood impacts.

The spread in precipitation errors and discharge scores is shown to reduce at a lower rate than that of the track forecasts, with relatively large spread found even after the forecast track confidence began to increase. A more detailed look at the precipitation forecasts from some individual ensemble members reveals that the forecast precipitation patterns and total accumulations can vary considerably even when the forecast landfall locations are similar, emphasizing how vital it is to increase the focus on evaluating and understanding ensemble TC rainfall forecasts in order to improve TC flood forecasting.

The application of these methods to additional TC cases is recommended and will allow a broader investigation of the lead times at which ensemble flood forecasting systems such as GloFAS are able to provide useful early warning of river flooding in TC cases, and what forecast and geographical contexts lead to variations in the predictability and the strength of the links between each component of the forecast chain.

Chapter 7 Discussion

The overall aim of this PhD is to design and conduct research to facilitate increased pull-through of ensemble forecast data in tropical cyclone forecasting and early flood warning. There were four main objectives:

Objective 1: To identify the current and potential use of ensemble forecasts by operational tropical cyclone forecast centres (Chapter 3).
Objective 2: To evaluate the benefits of multi-model ensemble tropical cyclone track probability forecasts (Chapter 4).
Objective 3: To identify the key factors that influence flood-related hazards from tropical cyclones using observations and reanalyses (Chapter 5).
Objective 4: To analyse predictability links in the tropical cyclone flood forecast chain using an ensemble-based method (Chapter 6).

Each of these objectives led to a research study which was written up in a published paper, and included in the thesis as Chapters, 3, 4, 5 and 6. This section now discusses each of these in turn, focussing on i) scientific knowledge contribution of each paper, ii) limitations of the study; and iii) additional outcomes and impact of the research. There is then a discussion of additional research that I have been involved with through the course of this study, which is related to the thesis aim. The final section provides ideas for next steps in taking the research forwards.

7.1 Use of ensemble forecasts in operational tropical cyclone forecasting

This study, published in Tropical Cyclone Research and Review in 2019, is contained in full in Chapter 3 and in appendix A1 in published form. It addressed the first objective: “To identify the current and potential use of ensemble forecasts by operational TC forecast centres”. The key question it was answering was: *How are ensemble forecasts currently used in operational tropical cyclone forecasting and warnings, and what potential is there for this to be improved?*

7.1.1 *Scientific knowledge contribution*

The research contributed an important baseline record of the current and potential use of ensemble forecast data by operational TC forecast centres such as Regional Specialised Met Centres (RSMCs) and Tropical Cyclone Warning Centres (TCWCs) (as of 2018). The mechanism for obtaining this knowledge was a questionnaire to operational TC forecast centres aiming to a) document the availability and use of ensemble forecasts in TC forecast centres; b) ascertain how uncertainty is calculated in operational warnings from their centre; c) obtain examples where ensemble forecasts have been successfully integrated into operations, and/or hurdles that prevent their full utilisation; and d) collate feedback on where future research and development should focus to enable wider use of ensemble forecasts in operational TC forecasting.

The results show that ensemble forecasts are used by nearly all respondents, particularly in TC track and genesis forecasting. There was, however, a notable difference between the high proportion of operational TC forecasters who used and valued ensemble forecast information, and the slower pull-through of ensemble-based probability and uncertainty information into operational forecast warnings and products for use by end users such as emergency management communities and the public. This is important because it shows that valuable situation-dependent information on the forecast confidence is currently being withheld from many users of TC warnings, with uncertainty information instead provided by historical error statistics that do not vary based on the forecast situation. Examples of where ensemble forecasts had been successfully pulled through into the operational TC forecasting process included the use of ensembles in the RSMC La Réunion operational track uncertainty cone (Dupont et al., 2011), and the work of the North Western Pacific Tropical Cyclone Ensemble Forecast Project (NWP-TCEFP, Yamaguchi et al., 2014). Four main hurdles preventing ensemble forecasts from being fully utilised were: i) lack of availability of ensemble data; ii) concerns about ensemble performance; iii) lack of familiarity with ensemble forecasts; and iv) challenge of communicating ensemble forecasts to customers.

The areas of research and development that would help TC forecasters to make increased use of ensemble forecast information in the future were those that would address these hurdles, including improved access to ensemble forecast data, verification and visualizations, the development of hazard and impact-based products, an improvement in the skill of the ensembles (particularly for intensity and structure), and improved guidance on how to use ensembles and optimally combine forecasts from all available models. A change in operational working practices towards using probabilistic information, and providing and communicating dynamic uncertainty information in operational forecasts and warnings, is also required. The research showed that there is huge potential to increase the use of ensemble forecasts in operational TC forecasts and warnings if the identified hurdles can be overcome by implementing the suggested recommendations.

7.1.2 *Limitations*

There are two main limitations to this research. Firstly, a questionnaire approach can only capture the current situation at the particular time point at which the survey is conducted i.e. 2018 in this case. The use of ensemble forecasts at operational TC forecast centres is constantly evolving due to a combination of factors including changes in the skill of ensemble forecasts as models are upgraded, changes in data availability and accessibility, advances in training, and updating of operational working practices. For this reason, some of the factual information gained will become outdated within a few years of the questionnaire being carried out. However, it still provides a

baseline that could be looked back at in future years, or repeated to see the changes in the intervening period.

Secondly, there exists a possibility of bias, with those forecasters who are more interested in ensemble forecasting, or more experienced in using them in their forecasting practices, being more likely to be interested and motivated to fill in the questionnaire. This study tried to minimise this issue by contacting specific contacts provided by WMO for each operational TC forecast centre, explaining all responses were valued whatever the levels of ensemble use, and providing information for them to forward on to all of their forecasters to reiterate this point. Several reminders were sent, particularly to regions where fewer responses had been received, to try to reduce any regional bias.

7.1.3 *Additional outcomes and impact of the research*

Perhaps some of the most important findings of this research were the identified hurdles that were preventing a more complete pull-through of ensemble forecasts in operational TC forecasting, particularly as these were consistent across several forecast centres. This led to several recommendations for future work to try to address and overcome these hurdles to enable increased pull-through of ensemble forecast data.

These recommendations were put forward in a sub-topic report (6.3: "Current and Potential Use of Ensemble Forecasts in Operational TC Forecasting") at the 9th meeting of the International Workshop on Tropical Cyclones (IWTC-9) in Hawaii in December 2018 (WMO, 2018), ahead of being published in the IWTC-9 special issue of the journal TCRR. IWTC is a unique 4-yearly gathering that brings together TC researchers and operational specialists from all regions, to discuss the latest research and forecast advances and share best practices to improve TC forecasts globally (Rogers et al., 2023). The IWTC recommendations set the agenda for research and operations on the key issues on which to focus for the next four years, so aligning this research with the IWTC process enabled the issues raised to gain a much greater profile and response.

As a direct result of the issues identified in this research and reported to IWTC-9, the WMO established a new GDPFS (Global Data-processing and Forecasting System) pilot project called TC-PFP (Tropical Cyclone Probabilistic Forecast Products), aiming to coordinate across forecast centres to identify best practice guidance for probabilistic TC forecasts. The main webpage of the TC-PFP project (WMO, 2020) directly references this research in its introductory paragraph:

> "This project was launched in response to recommendations from the 9th WMO International Workshop on Tropical Cyclones in 2018 on the use of ensemble forecasting at operations. The workshop, held every four years and attended by forecasters, researchers, and academics, identified slower pull-through into operational forecast warnings and products of probabilistic

> guidance and uncertainty information that ensembles can provide (e.g., static cones of uncertainty of track forecasts) (Titley et al., 2019) and recommended that an interdisciplinary approach be taken to address this issue. By implementing this project, we will coordinate across Regional Specialized Meteorological Centres (RSMCs) and other forecast centers to identify best practice guidance for probabilistic tropical cyclone forecasts and foster Research-to-Operations transfer to provide more accurate and reliable tropical cyclone forecasts to users."

I became a project member of TC-PFP in 2019 and a key contributor to the project. TC-PFP has started to address several of the recommendations of Chapter 3, including improving the consistency of access to ensemble forecast data and facilitating collaboration between operational forecast centres, NWP modelling centres, and researchers, to develop and share best practice on the use of ensemble forecasts in operational TC forecasting. Outputs and outcomes from the TC-PFP project are discussed further in Section 7.5. So a key outcome of this work was a change in WMO focus and resource into improving the pull-through of ensemble forecasts into operational TC forecasting.

Other outcomes relate to subsequent work to address the other research recommendations. The research paper recommended research activities to demonstrate the usefulness of dynamical ensemble forecasts for hazard-based products such as for strong wind, precipitation, flooding and storm surge, and the use of ensembles in impact-based forecasting for TCs. The work of Chapter 6 starts to address this for river flooding, while at the Met Office I have developed additional ensemble-based precipitation products for TCs for use by Weather and Climate Science for Service Partnership (WCSSP) partners in India and across SE Asia (see Figure 7-1). Exploratory work using feature-based methods to verify ensemble precipitation forecasts for TC cases are discussed further in Section 7.5.2, while the use of ensembles in impact-based forecasting for TCs is progressing in the Philippines as part of our WCSSP SE Asia collaborative work with PAGASA (Yu et al., 2022b).

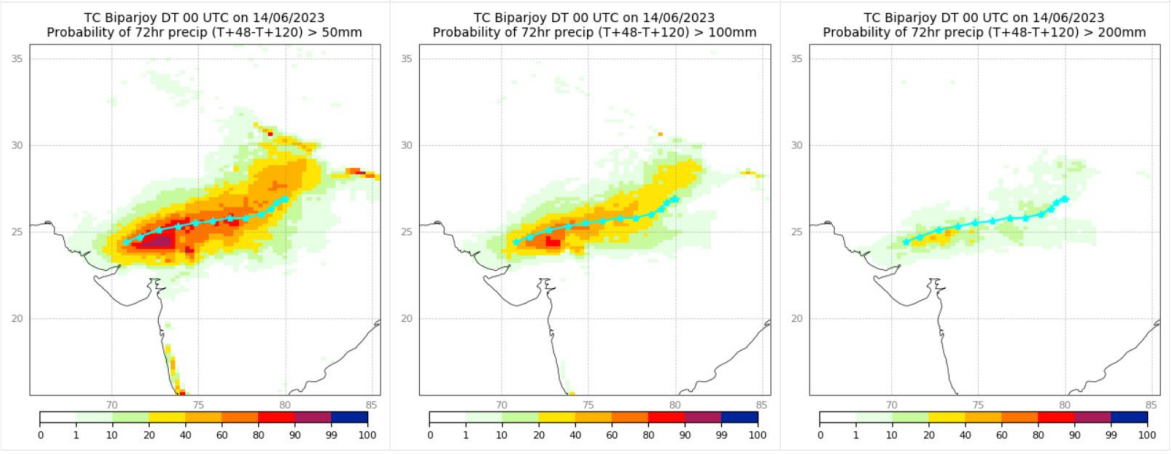


Figure 7-1. TC precipitation probability products from the Met Office MOGREPS-G ensemble are created for use by WCSSP India and South East Asia partners, with the ensemble mean track overlain for context. These are made available on a JASMIN webpage for all named storms in the North Indian and North West Pacific basins.

A further recommendation was to promote research activities to evaluate and improve probabilistic TC intensity and structure forecasts. There has been increased focus on ensemble TC intensity verification in recent years (Magnusson et al., 2021; Jia-jie et al., 2021; Torn and DeMaria, 2021; Zhou et al., 2022b; Wang et al., 2023). Since 2020 ECMWF have also made TC structure information available in the form of wind radii (the maximum distance from the centre of the cyclone to locations where the surface wind speed reaches 34, 50 and 64 knots) (Bidlot et al., 2020).

The research recommended wider sharing of code for ensemble product generation and verification, and I have subsequently shared our ensemble TC product code with NCMRWF which has enabled them to produce ensemble-based TC forecasts from their NEPS-G ensemble (Figure 7-2).

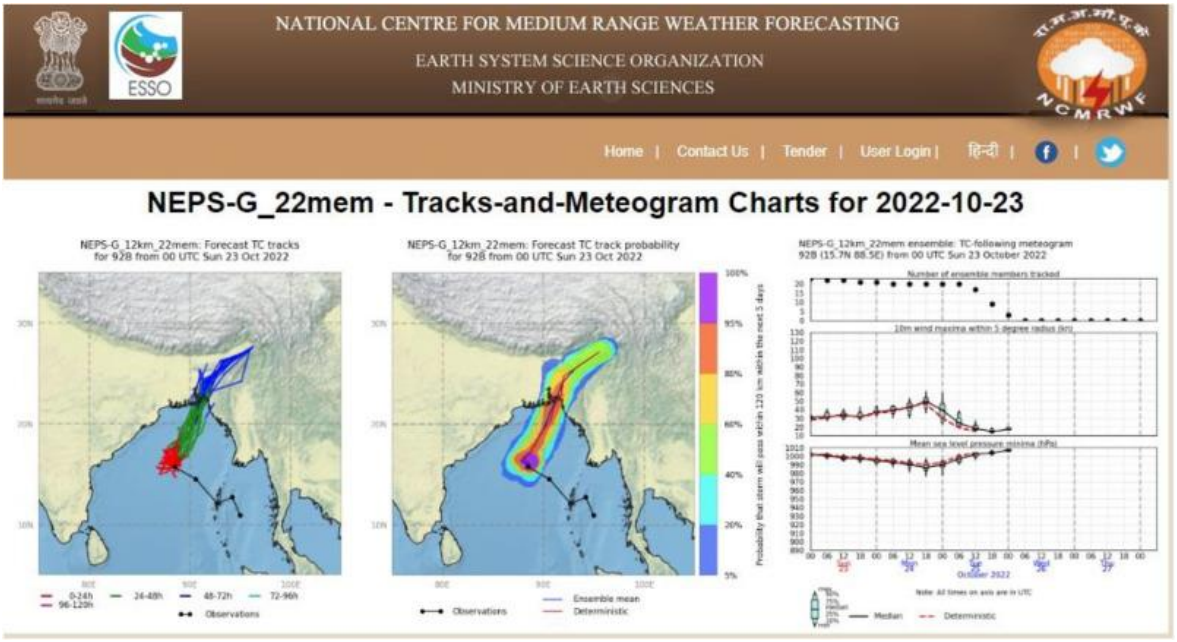


Figure 7-2. Sample of a NEPS-G ensemble TC forecast from NCMRWF (India), created using the Met Office TC ensemble post-processing code that I developed and shared (Fig. 2 from Conroy et al. (2023)).

Finally, the research recommended that the operational forecasting community facilitate a change in operational working practice towards using ensemble-based uncertainty and probabilistic information in forecasts and warnings. A great example where this has been taken forward since this research was published is in TCWC Melbourne, run by the Australian Bureau of Meteorology (ABoM). In early 2021 I held initial discussions with ABoM to advise them on combining multi-model ensemble TC data. In the 2021/22 TC season, ABoM changed their method for expressing track uncertainty from a climatologically based cone of uncertainty to a dynamically generated 'Forecast Confidence Cone', created using multi-model ('super') ensemble data (Conroy et al., 2023). This is really exciting ongoing work towards best practice in ensemble-based TC warnings, and the change

in nomenclature towards expressing the uncertainty with more positive language around 'forecast confidence' is also an important step towards explaining the benefits of dynamic ensembles to users.

7.2 Evaluation of tropical cyclone track probability forecasts

This study, published in the Quarterly Journal of the Royal Meteorological Society in 2020, is contained in full in Chapter 4 and in appendix A2 in published form. It addressed the second objective: "To evaluate the benefits of multi-model ensemble tropical cyclone track probability forecasts". The key question it was answering was: *What benefit is there in using ensemble forecast models in TC track forecasting, and does combining global ensembles into a multi-model (super) ensemble increase skill and value?*

7.2.1 *Scientific knowledge contribution*

The research provided a thorough analysis of track probability forecast skill from three global ensemble forecast models (MOGREPS-G, ECMWF ENS and NCEP GEFS), and their combined multi-model ensemble, across all global TC basins. The benefit of using ensemble forecast models over deterministic forecasts is assessed by using the mean track ('consensus') of the three deterministic model tracks as the reference forecast when calculating the forecast skill. The benefit of using ensemble track probability forecasts over this deterministic consensus track is shown, particularly at longer lead times or when computing the full 7-day track probability forecasts. This study therefore helps to illustrate the potential improved forecast guidance that could be gained if operational TC forecasting continues to migrate away from a deterministic-focussed forecasting environment towards pulling through probabilistic situation-based uncertainty information into operational forecasts and warnings.

The combined multi-model ensemble TC track probability forecasts were shown to have increased skill and value over the best-performing individual ensemble in every TC basin. Of the individual ensemble models, ECMWF ENS display the best reliability, skill and value. MOGREPS-G and NCEP GEFS become increasingly over-confident and under-dispersive with increasing forecast range, emphasising the importance of ongoing work at both centres to increase the spread in the ensemble forecasts. The added skill, reliability and value to be gained from combining the ensembles in a multi-model ensemble indicates that they have independent systematic strengths and errors and therefore collectively they can provide more realistic estimates of TC track probabilities.

7.2.2 *Limitations*

The first limitation of this work relates to the data sample size given a relatively short 2-year period was used in the verification. When undertaking studies such as this it is always a compromise between having a larger sample of data and ensuring that you are analysing the most up-to-date versions of each of the ensemble forecast modelling systems. Model and resolution upgrades occur most years, so having a larger sample would mean that you would be evaluating much older versions of the model. However, having only a 2-year period does mean that the results are sensitive to the particular storms that occurred in that period, particularly in basins where there are fewer storms each year such as the North Indian basin.

Secondly, this study used a relatively simple deterministic consensus track as the reference to calculate the probabilistic skill from the ensemble-based track probability forecasts. If this study were to be repeated, dressing the deterministic consensus forecasts using historical forecast errors and/or deterministic forecast spread statistics would provide a more sophisticated probabilistic reference forecast against which to test the dynamic ensemble-based track probabilities. Furthermore, there may be additional skill gained from combining the best of statistical and dynamical ensemble approaches. Under the U.S. Hurricane Forecast Improvement Project (HFIP) there is ongoing work to develop hybrid ensemble-statistical forecasts for both TC genesis and wind probability forecasts, aiming to incorporate the skill of statistical models with the dynamical uncertainty provided by multi-model ensemble forecasts (Brammer et al., 2021a, 2021b). Studies aiming to optimally combine deterministic, ensemble and statistical forecasts for use in operational TC forecasting are very valuable and should be encouraged.

Finally, in this study no weighting is applied to the ensemble members in the multi-model ensemble, with each member from each model given an equal weight. It would be interesting to look at different options for combining together the ensemble members, and also investigate how much of the additional value is from additional members as opposed to the inclusion of members from other models. Recent work at the Met Office on the Decider weather patterns tool (Neal et al., 2016) has shown very little difference between forecast skill when using different ways of creating a multi-model ensemble (i.e. equal member weighting vs equal model weighting vs model skill weighting) (Neal et al., 2022; Neal, 2023). But repeating this experiment using the multi-model ensemble TC forecasts would be an interesting further study.

7.2.3 *Additional outcomes and impact of the research*

One key outcome is that the benefit of a) using ensembles in TC track prediction over deterministic consensus forecasts and b) of using multi-model or super ensembles over single ensembles, has

been proven and disseminated, not just via the published paper, but also by presenting the results at various international meetings and conferences. For example, this study and two of its key figures were quoted extensively in IWTC-10 sub-topic report 4.4 ('Track forecast: operational capability and new techniques'), subsequently published as Conroy et al. (2023), helping to bring the key results to the attention of the wider TC forecasting community. This should help to encourage the use of ensembles in operational TC forecasting to try to exploit that potential benefit.

The track probability verification carried out in this study continues to be updated twice a year at the Met Office, at the end of each hemisphere's TC season, on the latest 12-month sample of storms. The key findings of the paper, regarding the benefit of ensembles over a deterministic consensus, and the benefit of a multi-model ensemble over a single ensemble, are replicated in this ongoing verification, showing that these findings are robust even when the sample of storms changes. These results, and my communication of them via Met Office TC meetings such as our weekly "Topical Tropical Teatime" gatherings have encouraged the use of the multi-model ensemble products by Met Office Global Guidance Unit (GGU) forecasters.

Ensemble TC verification activities at the Met Office are continuing to develop and expand to keep up the momentum in this area of work. For example, more recent multi-model ensemble TC product development has focussed on products that include storms forecast to form during the forecast (TC genesis) in addition to named storms. Initial subjective verification (see for example in Figure 7-3) reveals that here too the multi-model ensemble is adding benefit. Confirming this objectively is more complex as the results are highly sensitive to the specific thresholds employed by the tracker, but this continues to be an aim. Ensemble TC rainfall verification is also being developed (see Section 7.5.2), and it is hoped this will be expanded to multi-model ensemble forecasts in the future.

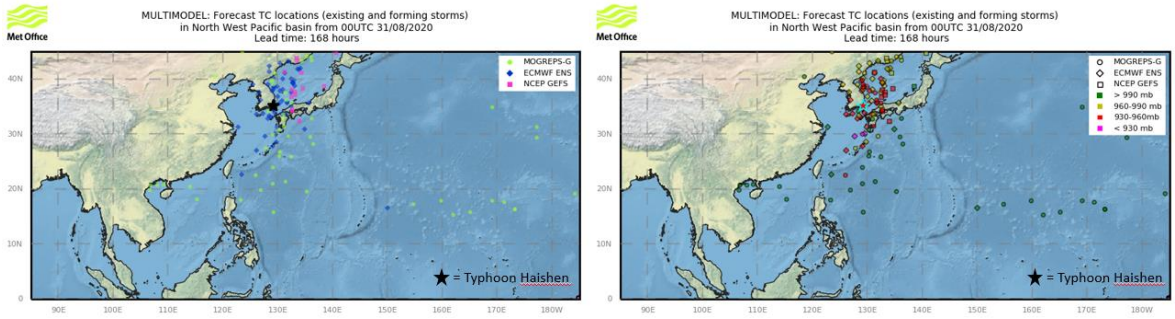


Figure 7-3. Example of a multi-model ensemble TC centre product I have developed for use by GGU and subjectively verified. Here is the case of Typhoon Haishen (2020), with TC centres from all members of the 3-model multi-model ensemble coloured by model (left) and by mslp minima (right). The model run is from 00UTC 31st August 2020. This run is prior to the storm forming, but even so the multi-model ensemble is giving a strong signal for a strong typhoon to impact South Korea at a 7-day lead time. The verifying observed position is overlain with a star on each plot.

This work has had impacts in the Met Office outside of TC forecasting, as the positive results found for combining the individual ensembles into a multi-model ensemble has encouraged the use of multi-model ensembles in other products, such as Decider (the Met Office product name given to an ensemble-based probabilistic weather pattern forecasting tool) (Neal et al., 2022).

7.3 Key factors influencing severity of tropical cyclone flood hazard

This study, published in the Journal of Hydrometeorology in 2021, is contained in full in Chapter 5 and in appendix A3 in published form. It addressed the third objective: "To identify the key factors that influence flood-related hazards from tropical cyclones using observations and reanalyses". The key question it was answering was: *Can a global analysis of TC cases and their associated fluvial flood hazard reveal the key meteorological and hydrological factors that lead to an increased TC flood severity?*

7.3.1 *Scientific knowledge contribution*

The research examines 280 landfalling TCs across the world in the decade from 2010 to 2019, to identify those characteristics that influence whether a storm has an increased flood hazard, using the GloFAS-ERA5 global river discharge reanalysis to represent the severity of fluvial flooding for TC cases in terms of the flood area, duration and magnitude. The study finds that TC cases with slow-moving, large, and intense cyclones, affecting areas with wet antecedent conditions, have the highest likelihood of causing widespread flooding, with severe flooding being more likely in cases where multiple drivers are present. Slower-moving storms in particular also have an increased risk of longer and more extreme floods. This is important because it implies that greater focus should be given to translation speed along the whole flood early warning chain, from model evaluation and development, through to warning design and communication, to better inform forecast-based action prior to TC landfall.

Larger and more intense TCs increase the likelihood of having a larger area affected by severe flooding but not its duration or magnitude, and it is found that the fluvial flood hazard can be severe in all intensity categories of TC, including those of tropical storm strength. This shows the importance of forecasting agencies continuing to emphasize to both the public and to organizations involved with disaster risk reduction preparedness and response activities that the fluvial flood hazard can be severe in all categories of TC.

Catchment characteristics such as antecedent soil moisture and slope also play a role in modulating flood severity, but this is more complex, with the conventional wisdom of higher levels of soil

saturation leading to increased flood risk not necessarily holding when dealing with the extreme precipitation totals in TC events.

The development of the three ways of classifying flood severity in TC cases via flood area, duration and magnitude, was also an important scientific contribution. Although the different flood severity measures are all significantly positively correlated with each other, there are cases where a storm has a very high flood severity in one measure and not in another e.g. where a large area is affected but flood duration is low. Some of the drivers of an enhanced flood severity are consistent across the flood indices, whereas some drivers are only significant for one or two of the indices. This emphasizes the importance of considering multiple flood indices to take into consideration all characteristics of the flood hazard, especially given that stakeholders may have differing priorities.

7.3.2 *Limitations*

The main limitations of this research stem from the data sources. Firstly, the study uses the GloFAS-ERA5 reanalysis as a proxy for river discharge observations when calculating the three flood severity indices for each TC in the study. The severity of fluvial flooding from TCs is difficult to calculate in a consistent way directly from observations given the relative sparsity of river discharge observations, particularly in many areas impacted by TCs (Lavers et al., 2019). Although GloFAS-ERA5 is partially dependent on model-derived precipitation and soil moisture, the groundwater and river routing parameters are calibrated against observation stations (Hirpa et al., 2018). The use of GloFAS-ERA5 reanalysis data allows an analysis of the flooding across the whole area affected by each TC, rather than just at isolated points on the river network, and allowed a direct comparison of flood severity across hundreds of TC cases in this study, using openly available data, in a way that would not be possible using discharge observations.

Secondly, precipitation data are taken from the GPM IMERG Final Run. Carrying out a systematic analysis collating data over hundreds of TC cases across the world using rain gauge or other local observation sources would be neither practical nor comparable. Although at times the IMERG data may fail to capture the precise rainfall intensities, its free availability across the whole of the tropics makes it an ideal choice of data for a study of this kind, and previous studies have shown it can capture the detailed spatial structure of rainfall associated with TCs (Rios Gaona et al., 2018).

A final limitation to discuss is that as this study included several hundred cases it was not possible to explore in detail the complexities that contributed to the detailed flood response in each individual case, such as the wider atmospheric environmental conditions or catchment management and land use. Case studies have shown the importance of land use changes such as urbanisation in amplifying total rainfall, altering the location of the maximum rainfall, and impacting

the hydrological response (Zhang et al., 2018). This type of detailed analysis is only possible in a more detailed case study approach. Nevertheless, the large analysis of several hundreds of TC cases carried out in Chapter 6 provides a systematic analysis of potential meteorological and hydrological drivers in a way that had previously not been carried out.

7.3.3 *Additional outcomes and impact of the research*

In the last few years, there have been positive signs for a shift in international research focus towards TC hazards such as flooding, largely driven by a number of high-profile TCs where flooding has led to major river flooding impacts such as Hurricanes Florence (2018) and Iota (2020) in the North Atlantic basin, and Cyclones Idai (2019) and Freddy (2023) in the South West Indian ocean basin. The U.S. National Hurricane Center (NHC) now places increased emphasis on rainfall and flooding information in their forecast products, web pages, and communication (Lamers et al., 2022). At the recent IWTC-10 meeting in December 2022, a specific topic was given over to forecasting TC rainfall and flooding hazards and impacts. The final recommendation of that report recognised the importance of work to better understand how interdependencies between meteorology and hydrology can affect on-the-ground impacts (Lamers et al., 2022), showing the importance of studies like the one in Chapter 5 that attempt to add knowledge in this area.

At the Met Office my involvement in this research has effected change, as my communication of my PhD topic has led to an increase in the use of GloFAS as a forecasting tool in TC cases by our GGU forecasters. Following this study there is also an increased awareness of the key factors controlling flooding, with increased focus on this both in forecast, model performance, and future research priorities. An example of this is in the Resilience and preparedness to tropical cyclones across southern Africa (REPRESA) project. REPRESA is part of the Climate Adaptation and Resilience initiative (CLARE), which aims to build resilience to climate change and natural hazards for people across Africa and the Asia-Pacific. REPRESA aims to improve TC early warning systems in Southern Africa, conducting comprehensive research on changing cyclone risks in a changing climate, and formulating adaptation options that enhance resilience. Novel climate and hydrodynamic flood modelling, coupled with risk and vulnerability reduction approaches, will work towards understanding compound TC risks and the likelihood of unprecedented TC flood impacts (University of Bristol, 2023). REPRESA runs from 2023 to 2026 and I am proving expertise to this project in understanding fluvial flood drivers in TC cases, and will be helping to advise on how to analyse the high-resolution climate runs specifically in order to better understand changing fluvial flood risk in the future. For example, it will be important to investigate any trends in translation speed and TC size alongside changes in TC frequency, track locations, TC intensity and inland propagation.

The research findings are also being used in the humanitarian sector to help with interpreting flood forecast guidance. For example, this quote from the Red Cross and Red Crescent Climate Centre illustrates the impact of the research in aiding understanding of flood forecast signals in a TC case:

> "When GloFAS was not indicating flooding from Freddy's landfall in Madagascar in February 2023, Helen's work gave us more confidence in what GloFAS was saying; while Freddy was a very intense storm, it was going to move very fast across the country and therefore would not generate rainfall accumulations which would cause significant flooding. This meant that we could focus on highlighting the wind impacts in Madagascar." (E. Stephens, personal communication, 2023)

7.4 Analysing predictability links in the tropical cyclone flood forecast chain

This study, submitted to the Journal of Hydrometeorology in 2023, is contained in full in Chapter 6 and in appendix A4 in submitted form. It addressed the fourth objective: “To analyse predictability links in the tropical cyclone flood forecast chain using an ensemble-based method”. The key question it was answering was: *How are forecasts for river flooding in TC cases affected by TC track, intensity, and precipitation forecasts, and can ensembles help to investigate this?*

7.4.1 *Scientific knowledge contribution*

This work proposes and develops ensemble-based methods to analyse the predictability links between the track, intensity, precipitation and discharge forecast components of the ensemble flood forecast chain in TC cases, using the forecasts from the Global Flood Awareness System (GloFAS) and the driving EC ENS ensemble forecasts. The performance of each ensemble member at each stage of the forecast is assessed, enabling an assessment of the connections between each forecast component, both via correlations and the overall change in ensemble behaviour from run to run. The methodology is demonstrated using Hurricane Iota, where strong relationships are found between TC track, precipitation and river discharge skill. The forecasts began to show increased confidence for the TC genesis and track of Iota around 5 days ahead of landfall, leading to a clear signal for a significant flood event in the overall region that was affected. However, at the catchment scale, the research highlights that the sensitivity of discharge forecasts to the landfall track and intensity forecasts varies considerably depending on the catchment’s location relative to the TC, with a greater sensitivity to precise landfall and intensity forecasts for river catchments close to the landfall location that are impacted by the heavy rains around the eye wall. The rainfall from the wider storm circulation is crucial to flood impacts in most of the affected river basins, with a stronger relationship with the post-landfall track error rather than the precise landfall location. This shows the importance of predicting TC behaviour after landfall and not only focusing on landfall location and the category of the storm when forecasting and communicating TC impacts. The recent

case of Cyclone Freddy (2023) provides a further illustration of this: the ECMWF ensemble forecasts were indicating the storm would travel further inland than it actually did, indicating a flood risk to Zimbabwe, but this didn't occur because the storm didn't propagate that far inland.

Looking across the forecast runs in the run up to the Hurricane Iota flood event, the spread in precipitation errors and discharge scores reduces at a lower rate than that of the track forecasts, with relatively large spread found even after the forecast track confidence began to increase. The forecast rainfall patterns and total accumulations are shown to vary considerably even when the forecast landfall locations are similar, emphasizing how important it is to increase the focus on evaluating and understanding the skill and reliability of ensemble TC rainfall forecasts in order to improve TC flood forecasting.

The scientific contribution of this study is therefore threefold: i) a new method using global freely available datasets to analyse the flow of predictability through the TC flood forecast chain, meaning it could be repeated over a range of cases; ii) knowledge about the principle drivers of the flood predictability in this case, how this varied from catchment to catchment, and the performance of the GloFAS forecasts in predicting the flooding; and iii) evidence-based suggestions on where future research and forecast communication work could focus, such as a detailed evaluation of ensemble TC rainfall forecasts, and on focussing on hazard-based and impact-based communication given that landfall location and storm category were not crucial to the flood impacts in the wider region.

7.4.2 *Limitations*

As with the previous study the main limitations relate to the various datasets selected for use, and in particular the reliance on reanalyses rather than observations for the river discharge verification, and satellite-based GPM IMERG data for the rainfall verification. The use of reanalyses as the river discharge verification data set has the advantage of being available at every global river point, and in the Hurricane Iota case this was the only available source of data for the main impacted rivers. If reliable discharge observations from river gauges were more widespread and openly available at the impacted rivers in a future application of this method to a different TC case it would be interesting to carry out the analysis using both the river gauge and reanalysis discharge data, to investigate how sensitive the results are to the choice of river discharge verification data. Similarly, the study relies on GPM IMERG data to verify the TC rainfall, and ground-based observations in Iota indicate potential errors in capturing the true rain rates in some regions, which may impact the robustness of the analysis of the errors of the precipitation forecasts. If a future case study had access to alternative sources of precipitation 'truth' data such as radar and gauge-based datasets, an investigation of the sensitivity of the results to the choice of rainfall verification data could be

carried out. A further limitation is the simple point-based metric used to evaluate the precipitation forecasts, and in the future other metrics that remove the point-by-point sensitivity could be applied, and/or the catchment-total rainfall calculated when analysing the forecasts for specific rivers.

This study focused on a particular case, Hurricane Iota, and the resultant flooding that occurred in Honduras and Nicaragua. Even in the same region, no two cyclones are the same in terms of their track, speed, size and wider environmental conditions, and so care should be taken in applying the understanding gained in this study to future storms. Although the findings of the case study provide important food for thought for research and forecast practitioners, the wider applicability of these results in other geographical contexts is uncertain because flood severity also depends on the local landscape and catchment characteristics. The wider application of the presented methods across a range of cases would help to inform the robustness and wider applicability of the findings of this case and how the predictability links in the TC flood forecast chain vary with different geographical and forecast contexts.

7.4.3 *Additional outcomes and impact of the research*

This paper has been submitted for publication and I have just submitted a response to the reviewers' comments, but there are already wider outcomes of the work as a result of the collaborations built during the work. The work was carried out in close collaboration with ECMWF, so even during the development stage it inspired a greater focus on TC flood forecasting at ECWMF. For example, in a detailed report on TC activities at ECMWF compiled in 2021 (Magnusson et al., 2021), a specific section relating to TC river flooding was included inspired by initial results that I had given at an ECMWF seminar, and I was asked to contribute a summary of my research in this report.

When I started my PhD in 2018, my PhD, and the work of this paper, was the first time that anyone had looked in any detail at the use of GloFAS in TC flood forecasting. In March 2019, when Cyclone Idai impacted Mozambique with devastating flooding impacts, researchers at the Universities of Reading and Bristol began to collaborate with ECMWF to assist in providing flood forecast information in support of the UK government's Foreign, Commonwealth and Development Office (FCDO). This further developed into the production of flood forecast bulletins by the Red Cross and Red Crescent Climate Centre, FATHOM, and HR Wallingford. The flood forecast bulletin work is described in more detail in Section 7.5.3, along with its associated publications. I have acted as an advisor to the flood bulletin team since 2019, both during the forecasting process and in the post-

event evaluation, inputting the knowledge about TC flood forecasting that I gained along my PhD journey into this important humanitarian endeavour.

The REPRESA project, previously mentioned in Section 7.3.3, also has a component looking at early flood warnings in the three southern African countries most exposed to cyclones in the tropics: Madagascar, Mozambique and Malawi. The national meteorological services of the three countries will work with researchers, humanitarian response actors and vulnerable communities to co-develop new early warning products aimed at enhancing responses to changing flood risks. Again I have been asked to advise on this process to contribute my knowledge gained on the predictability links in the TC flood forecast chain. This may include contributing some of the python code that I have developed during the PhD to help both with analysing flood forecasts and creating prototype products to assist in flood early warning in TC cases.

7.5 Additional research towards the aim of this thesis

In addition to the four main papers included in this thesis, there have been several other published works that I have either led or collaborated on that are also very relevant in their contribution to the aim of this thesis, which is to improve the pull-through of ensemble forecast information into operational TC forecasting and early flood warning. These are included below, divided into three main topics.

7.5.1 *WMO activities on ensemble pull-through in TC forecasting*

Section 7.1.3 explained how via the mechanism of the WMO IWTC recommendation process, the work of Chapter 3 (Titley et al., 2019) has directly led to an increased focus on the international stage into improving the pull-through of ensemble forecast data into operational TC forecasting and the creation of the WMO GDPFS pilot project called TC-PFP. The first phase of TC-PFP was from 2020-2022 and I presented a summary of this phase to IWTC-10 in Bali in December 2022 (Titley, 2022), which included linking back to the motivation from the study contained in Chapter 3. One of the key conclusions of the TC-PFP phase 1 efforts was how important it is to urgently improve the accessibility and standardisation of ensemble forecast track data (increasingly referred to as vortex parameter files in order to make it clear that many parameters rather than just location data should be included) (Dunion et al., 2023). As a result of this, IWTC-10 made the following recommendation:

> "Encourage modeling centers to encode vortex parameter files in a standardized format (e.g. BUFR, CXML) and exchange them in a timely and consistent manner, such as GTS. [WMO]" **(WMO, 2022)**

This led to the formation in June 2023 of a new WMO Task Team on Tropical Low/Cyclone Vortex Parameters (TT-TCPara) which aims to: i) Provide recommendations on which vortex parameters

should be designated as core (mandatory) and which should be recommended (optional); ii) Establish detailed technical requirements on tropical low/cyclone vortex parameter files and propose updates to the Manual on the GDPFS (WMO-No. 485); iii) Support the development of dissemination format and metadata requirement of tropical low/cyclone vortex parameter files. I am a member of this task team, as following my research in Chapter 3 and the subsequent work of the TC-PFP project, I believe improving the operational accessibility of ensemble forecast data is vital to facilitate increased pull through of ensemble forecasts into operational TC forecasting.

The work of Conroy et al. (2023) and the IWTC-10 sub-topic report from which this was published, has already been introduced in Sections 7.1.3 and 7.2.3. I am a co-author on this paper, and it is included in its published form in Appendix A5 due to its relevance to the aims of this thesis. A summary of this paper and my contribution to it is given below:

Appendix A5
Published version of Conroy et al. (2023): *'Track forecast: Operational capability and new techniques - Summary from the Tenth International Workshop on Tropical Cyclones (IWTC-10)'*. Tropical Cyclone Research and Review, 12 (1), pp.64–80.

Summary and relevance to this thesis
The paper summarises operational TC track forecasting capabilities and new techniques that have been developed since IWTC-9. This includes the development of multi-model ensemble-based depictions of track uncertainty, such as the 'Forecast Confidence Cone' developed by ABoM. This work is very relevant to the main thesis aim of unlocking the potential of ensemble forecasts in operational TC forecasting and highlights the progress that has been made since the publication of Titley et al. (2019) (Chapter 3).

My contribution
I contributed to several sections of the report including those about TC-PFP and the evaluation of multi-model ensemble track probability forecasts. I also provided TC track data to enable the research around vortex tracking algorithms, and contributed to the general writing and editing of the manuscript.

One final WMO project of which I am a member that is relevant to the aim of this thesis is the HIWeather Warning Value Chain project (Ebert et al., 2021). This project has created a Warning Value Chain case study template to enable researchers to analyse high-impact case studies in terms of their information flow through the warning chain from weather forecast to hazard and impact forecasts, through to the communication of warnings and associated responses. These will be compiled into a database to generate an easily accessible means for scientists and practitioners involved in researching, designing, and evaluating weather-related warning systems to review relevant previous experience and assess their efficacy using value chain approaches. I have contributed to several case study analyses including for Hurricanes Ida and Ian, Storm Eunice (an extra-tropical cyclone which impacted the UK), and Typhoon Noru, with a particular interest in how probabilistic ensemble-based information is passed down the forecast chain.

7.5.2 *WCSSP research in support of this thesis*

My Met Office work in support of WCSSP in India and SE Asia was briefly introduced in Section 7.1.3. There have been two larger pieces of research carried out under the WCSSP India project that helped to support and develop the ideas in this thesis. The first piece of work helped to develop the ideas that were illustrated in Chapter 6. It assessed the ability of ensemble river discharge forecasts from the Global Flood Awareness System (GloFAS) to forecast the fluvial flooding from five Indian TCs: Cyclones Bulbul, Amphan, Nisarga, Nivar and Burevi. It introduced a new score (the Titley Score) to assess the forecast discharge peaks across the area affected by the cyclone in terms of the amplitude and timing of the flood peaks. This specific score wasn't taken forward into Chapter 6; I decided that the established modified KGE score would be more appropriate in this paper as I wanted the paper's focus to be on the predictability links across the forecast chain rather than on the development of a new scoring method for river discharge. Many of the other ideas for Chapter 6 were formulated and developed during this research, and the report also provides a useful assessment of the ability of GloFAS to provide flood forecast guidance for five additional TC cases, so it is appropriate to include the report in full in Appendix A6. The main findings are summarised below.

Appendix A6
Met Office WCSSP India report: Titley, H.A. (2021). *M3.2.1.6: The predictability of fluvial flooding from recent Indian cyclones in the Global Flood Awareness System (GloFAS)* (WCSSP India project report).

Summary and relevance to this thesis
GloFAS forecasts were able to successfully distinguish between those TCs that led to relatively high levels of fluvial flooding and those where the flooding was much less severe. The track and forward speed of the storm seemed to be the key controls on predictability in several cases, for example in the Burevi case, where the stalling of the cyclone in the Gulf of Mannar was key to correctly forecasting the heavy precipitation accumulations. The overall forecast performance was better at longer lead times for the two more southerly events impacting Tamil Nadu and Andhra Pradesh (Nivar and Burevi), with good performance 6-8 days prior to the discharge peaks compared to only 2-3 days prior to the discharge peaks in the northern events impacting West Bengal (Bulbul and Amphan). Even when the landfall location is well forecast, the details of the precipitation forecast can influence the predictability of the flooding, highlighting the importance of further research to evaluate the ability of the driving ensemble precipitation forecasts to forecast the spatial distribution of rainfall in tropical cyclone cases.

This research is very relevant to the thesis as it was a stepping stone to help develop the ideas shown in Chapter 6 on assessing the predictability links across the flood forecast chain. It also contains useful information on the potential use of GloFAS to help forecast TC flooding by looking at the lead times at which flooding was predicted across a range of TC cases.

My contribution
I am the sole author of this report. I designed the study, carried out the research, and wrote the report.

The second main piece of work carried out within WCSSP in support of this thesis was around the feature-based verification of ensemble TC rainfall forecasts. Several of the studies in chapters 3 to

6 have discussed the importance of increasing the focus on understanding how well ensemble forecasts can predict rainfall in TC cases, and so I proposed and carried out research funded by WCSSP India to begin to investigate this.

When verifying rainfall, feature-based methods offer a greater insight into the ability of models to correctly forecast rainfall (Ebert et al., 2013), but there are only a handful of studies where these have been applied to TC rainfall forecasts (Marchok et al., 2007; Chen et al., 2018; Yu et al., 2020). While some feature-based methods have been adapted and applied to verification studies of ensemble forecast data (Gallus, 2010; Radanovics et al., 2018; Ji et al., 2020; Dube et al., 2022), to my knowledge there has not been any attempt to apply feature-based methods to ensemble forecasts from TCs in particular. This inspired the following research question: *Can the application of feature-based verification techniques help to improve our understanding of the ability of global ensemble forecast models to predict rainfall in TC cases?* To answer this research question I led collaborative research between Met Office and NCMRWF scientists as part of the WCSSP India project. The results of this research were published in a Met Office WCSSP India project report in October 2022. The study is included in full in Appendix A7, entitled "Exploring the use of feature-based verification techniques on ensemble forecasts of tropical cyclone rainfall". The findings are summarised below.

Appendix A7
Met Office WCSSP India report: Titley, H.A., North, R., Ashrit, R. and Dube, A. (2022). *M4.1.2: Exploring the use of feature-based verification techniques on ensemble forecasts of tropical cyclone rainfall* (WCSSP India project report).

Summary and relevance to this thesis
This study applied three object-based verification methods (SAL, MODE and CRA) to evaluate the forecasts of TC-related rainfall in four landfalling cyclones in India. Three ensemble forecast models were evaluated: MOGREPS-G, ECMWF ENS and NEPS-G. The three methods provided useful complementary information on the ability of the ensemble models to forecast TC rainfall. Across the cases the rainfall intensities and volumes are generally too low for all three models for both 24-hour and storm total rainfall. The error/spread relationship is consistently better for the rainfall object area than for the other attributes, with volume and intensity being generally under-dispersed. The rainfall object structures of ECMWF ENS are generally larger and flatter than the other two ensembles. The smaller and sharper objects in MOGREPS-G and NEPS-G are closer to the observations over the ocean. However, ECMWF ENS appears to better capture rainfall over land, which is crucial to forecasting flooding impacts.

This research is very relevant to the thesis aim as it helps to show the current strengths and weaknesses in the ability of global ensemble forecasts to predict TC rainfall, which is crucial for pulling through ensemble forecasts into flood forecasts via systems such as GloFAS.

My contribution
I designed the research, led and organised the collaboration, led the SAL component of the verification, and contributed to the MODE verification. I organised the writing and editing of the report.

7.5.3 *Emergency flood bulletins to support humanitarian decision-making in TC cases*

The emergency flood bulletin work in support of FCDO was briefly introduced in Section 7.4.3. When activated for a TC case where severe flooding is anticipated, a team of scientists from the Red Cross and Red Crescent Climate Centre, FATHOM, HR Wallingford, and ECMWF produce daily flood bulletins detailing the expected location and timing of TC landfall, the rivers most at risk of flooding, the predicted timing of the flood peak and recession, and inundation and population exposure information. The objective of the bulletins is to facilitate humanitarian decision-making, such as the distribution of resources ahead of flooding. The flood bulletins, modelling framework, and an evaluation of the forecasts and the bulletins themselves for the initial TC cases (Idai and Kenneth), are described in detail in Emerton et al. (2020). The pilot project has since been activated a number of times, including for TCs Iota, Amphan, Eloise, and Freddy, with a subsequent paper submitted by Speight et al. (2023) on the flood bulletins for Iota and Eloise. I have acted as an advisor to the flood bulletin team during the forecasting process and in the post-event evaluation. I am therefore a co-author on both of those papers, which are included in Appendix A8 and Appendix A9. See the boxes below for further details.

Appendix A8

Published version of Emerton et al. (2020): *Emergency flood bulletins for Cyclones Idai and Kenneth: A critical evaluation of the use of global flood forecasts for international humanitarian preparedness and response*. International Journal of Disaster Risk Reduction, 50, p.101811.

Summary and relevance to this thesis

This paper contains a critical evaluation of the collaborative production of emergency flood bulletins for Cyclones Idai and Kenneth in Mozambique in support of the international humanitarian community, using global river flood forecasts from GloFAS combined with flood inundation modelling and population exposure estimates to inform impact risk assessment. The bulletins supported critical actions such as sending an assessment team to the region most likely to be affected and considering the availability of equipment such as hygiene kits and tarpaulins ahead of the response. The forecast evaluation showed that the districts at greatest risk of flooding were successfully identified, and revealed the great potential value in using global ensemble-based operational forecasting models (such as GloFAS) for supporting the international humanitarian community to take actions for ahead of TC landfall. An important conclusion was the need to work more closely with the mandated national authorities responsible for disseminating forecast and warning information to support capacity development for a national, operational end-to-end multi-hazard early warning system.

This research is very relevant to the thesis aim as it provides a first analysis of the forecasts of an ensemble-based flood forecasting system (GloFAS) for a TC case. The analysis of the forecast components were only assessed for the ensemble as a whole e.g. looking at average track errors and ensemble mean precipitation, so it inspired my work to create a methodology to look in greater detail across the ensemble to analyse the predictability links in the flood forecast chain, as was carried out in Chapter 6.

My contribution

I acted as an advisor to the TC flood forecast bulletins, and helped with the write up of the forecast assessment (section 4) and the general editing of the manuscript.

Appendix A9
Submitted paper: Speight et al. (2023): *Recommendations to improve the interpretation of global flood forecasts to support international humanitarian operations for tropical cyclones*. Journal of Flood Risk Management, (Submitted).

Summary and relevance to this thesis
This paper captures the successes and challenges of using global ensemble forecasts to generate flood forecast bulletins to support humanitarian operations for two cyclones: Hurricane Iota in Central America (November 2020) and Cyclone Eloise in Mozambique (January 2021). The case studies show that global ensemble flood forecast models are a valuable resource for humanitarians. The access to probabilistic information and increased lead time with respect to local models is appreciated by decision makers with humanitarian colleagues excited by the expertise and insight this pilot project has enabled during TC flooding events.

The paper states the need for close research and operational collaborations to improve the linkages between expert knowledge and forecast interpretation. The work of Chapter 6 aims to respond to this need by providing useful information to help to interpret the flood forecasts for TC events and their sensitivities to uncertainties in each part of the TC flood forecast chain.

My contribution
I acted as an advisor to the TC flood forecast bulletins, and helped with the forecast evaluation sections and the general editing of the manuscript. I created Figures 5 and 9 which provide the basis for the subjective forecast evaluation for Hurricane Iota and Cyclone Eloise respectively.

7.6 Recommended next steps

The preceding sections of this discussion have shown how many of the specific recommendations from Chapters 3 to 6 are already being taken forward at NWP centres such as the Met Office and ECMWF, or in the wider scientific and forecasting community via international collaborations. Via my involvement in projects such as TC-PFP, TT-TCPara, and the REPRESA project, I will endeavour to ensure that the momentum of this work to pull through ensemble forecasts into operational TC forecasting and early flood warning will continue. Some specific recommendations are listed below, both for these projects and for other areas where further research or resources would be beneficial:

1. Timely, reliable, consistently formatted feeds of ensemble forecast vortex parameter data should be made available to all operational TC forecast centres (Chapter 3). This important task will be facilitated by the TT-TCPara group who have important work to do to agree the required parameters and file format for ensemble vortex parameter file distribution. Once this is in place, alongside training for NWP centres to encode the files and for operational forecast centres to decode these files, a key hurdle identified in Chapter 3 that prevents wider pull through of ensemble forecast data into operational TC forecasting will have been overcome. The newly formed WMO Advisory Group on Tropical Cyclones (AG-TC) made up of members from each of the operational TC forecast centres (WMO, 2023), who aim to support of the delivery of globally consistent TC-related services, will play an important role alongside the TC-PFP project in sharing and promoting best practices to exploit this data in operational TC forecasts and warnings.

2. Increased funding should be made available for research from a range of social science disciplines, providing expertise in communicating complex information and in graphic and product design, to guide how to optimally present probabilistic forecast information in TC forecasts and warnings to a range of users including the public to optimise understanding and behavioural responses. More generally, there is already a useful bank of research into how to effectively communicate probabilistic weather forecasts to improve public understanding (Ripberger et al., 2022b) and for hurricanes specifically the U.S. Hurricane Forecast Improvement Project (HFIP) has carried out some useful work in this area e.g. Ripberger et al. (2022a). Further investment could focus on building an international community of practice invested in researching these issues across those regions impacted by TCs, to examine whether regional or cultural differences affect the optimal way to present probabilistic forecasts and warnings.

3. Now that ECMWF have become the first centre to bring the ensemble to the same 9km resolution as their deterministic forecast, there should be increased focus on evaluating the potential benefit of pulling through ensemble forecasts for TC intensity and structure into operational TC forecasts and warnings. The previous poor performance of global ensembles in forecasting TC intensity (Magnusson et al., 2019) is one reason why the pull through of ensemble forecast data into operational TC forecasting has been slow (Chapter 3), but there are promising initial results for TC intensity with the upgraded ensemble (Lentze, 2023). Further research as to how to optimally combine information from multiple ensemble forecast models with different skill and resolution would be beneficial, including how to best utilise global ensembles alongside any available higher resolution ensemble forecast systems and deterministic forecasts. The TC-PFP project is ideally situated to take forward this work as it is just starting a second phase focussing on probabilistic forecasts for TC intensity and structure forecasting.

4. All operational TC forecast centres should adapt their TC track warnings to include the dynamic situation-based uncertainty information that ensembles can provide. These should be created using multi-model ensemble forecasts, given the additional skill and value shown over individual ensemble models in Chapter 4. The recent work developing ‘forecast confidence cones’ at ABoM (Conroy et al., 2023) serves as a current example of best practice that could be emulated in other regions, both in its methodology and its use of positive terminology through the use of ‘confidence’ rather than ‘uncertainty’ areas.

5. Feature-based TC rainfall verification, as introduced in Appendix A6, should be further developed, and added to model evaluation protocols so that the changes to the ability of ensemble forecasts to forecast TC rainfall can be considered alongside TC track and intensity forecasts when assessing model upgrades. This will ensure that model updates are providing

improvements to the hazard forecasts that are so important to correctly predicting TC impacts. A more thorough understanding of the ability of each ensemble forecast model to predict aspects of TC rainfall such as rainfall area, volume, and intensity, can also guide the creation of multi-model ensemble-based TC rainfall forecasts and warnings.

6. National Meteorological and Hydrological Services in countries impacted by TCs should be encouraged to use ensemble-based flood forecasts to improve flood early warning, which will help humanitarian preparedness activities. A wider and more routine assessment of the lead times at which ensemble flood forecasting systems such as GloFAS can provide useful early warning of river flooding in TC cases is recommended, to guide and support this activity. The REPRESA project, which started in June 2023, offers exciting opportunities to take forward the learning of this thesis related to TC flooding, from medium-range forecast timescales through to climate timescales. At the medium-range timescales REPRESA will work with national forecasting centres in Mozambique, Madagascar and Malawi to implement sustainable agency-led flood forecasts using the ensemble-based approach pioneered in the flood forecast bulletins (Emerton et al., 2020; Speight et al., 2023). At climate timescales, new high-resolution regional climate ensembles offer a great opportunity to evaluate the potential changes in compound flood forecast risk (including both river flooding and coastal flooding from storm surges) in Southern Africa and Madagascar under a warming climate. The knowledge of key drivers of flood severity (Chapter 5) and the predictability links across the flood forecast chain (Chapter 6) will be important in guiding the analysis of the changing fluvial flood hazard.

7. Increased focus should be given to translation speed throughout the TC forecast process, from model evaluation and development, through to warning design and communication, given its key role in modulating flood severity (Chapter 5). This thesis found that river flooding can be severe in all intensity categories of TC (Chapter 5), with weak predictability links between intensity errors and overall flood forecast skill (Chapter 6), and stronger correlations between the track errors after landfall and flood forecast skill compared to the correlation between landfall error and flood forecast skill (Chapter 6). Thus it is important that TC forecast centres emphasise more than just the landfall location and the intensity category of an approaching TC in their forecast messaging, as the track after landfall, speed of the storm, precipitation patterns, and catchment characteristics are more influential in controlling the impacts from river flooding.

8. Global-scale flood forecasting systems should incorporate multi-model ensemble precipitation forecasts to establish if it leads to increased skill in forecasting floods from TCs. Chapter 4 showed the additional skill gained by using multi-model ensembles for TC track forecasting, and given the close links between TC track error and river discharge forecast skill shown in Chapter

6, it would be anticipated that using multi-model ensemble forecasts as the driving data in flood forecasting systems such as GloFAS would lead to improved discharge forecast skill. The ensemble-based methodology presented in Chapter 6 to investigate predictability links in the flood forecast chain should be repeated across a larger number of cases to see how different forecast and geographical contexts lead to variations in the predictability and the strength of the links between each component of the flood forecast chain. If applied to flood forecasts driven by multi-model ensemble forecasts, this methodology would allow for a comparison of how consistent these predictability links are across the different driving models.

9. TC-specific products or layers within the GloFAS web interface should be created to provide improved forecast guidance for river flooding in TC cases and improve the uptake of GloFAS forecasts by the TC forecasting community. Even the option of simple layers showing the driving ensemble member tracks coloured by intensity would provide useful contextual information to the flood forecasts. More sophisticated dynamical layers could be developed that would allow you to follow particular scenarios through the forecast from the TC track through to the flood forecast. An example of a possible product was given in Figure 6-11 in Chapter 6, highlighting the forecast TC locations of those ensemble members that are forecasting a significant flood event at a particular river point. This idea could be further developed to help to provide planning scenarios as part of TC flood forecast bulletins. You could click on particular TC track forecasts and highlight the river discharge forecasts on key river hydrographs, or vice versa (i.e. clicking on those members with the highest peak river discharge and highlighting their TC tracks). This could help to support multi-hazard early warning systems and scenario planning activities, as the TC tracks of those members forecasting significant flooding would reveal whether a region forecast to experience river flooding may also experience other hazards such as from strong winds and storm surge.

This discussion has presented the scientific knowledge contributions, limitations, and wider outcomes of each of the four main chapters of this thesis, followed by a discussion of additional related research, and a set of recommendations for potential next steps. The research has provided evidence of how ensemble forecasts can improve TC forecasts and warnings, knowledge of TC flood drivers and predictability, and an increased focus internationally on unlocking the potential of ensemble forecasts in TC forecasting. This will hopefully mean that the legacy of the work will last beyond the confines of this thesis.

Chapter 8 Conclusions

Ensemble forecasting is now widely recognised as a key future path for improving the forecast skill and utility of weather forecasts and warnings. This thesis has presented a variety of research aiming to unlock the potential of ensemble forecasts to help improve operational TC forecasting and early flood warnings.

A survey of operational TC forecast centres (Chapter 3) revealed that the extent to which ensembles are used and pulled through into warnings varies from region to region, with several regions not yet including situation-based uncertainty information in their warnings despite the forecasters valuing the guidance ensembles can provide. Important hurdles were identified that need to be overcome to enable the wider and fuller use of ensemble forecasts, including improving consistency and timeliness of access to ensemble forecast data, improving ensemble intensity forecasts, creating hazard and impact-based TC ensemble products, and facilitating the sharing of best practice in how to create and communicate probabilistic TC products and warnings. The study revealed the huge potential to increase the use of ensemble forecasts in operational TC forecasts and warnings if these hurdles could be overcome, and gave recommendations that were taken forward via the IWTC process and specialist WMO groups formed after IWTC-9 such as TC-PFP, AG-TC and TT-TCPara.

An evaluation of multi-model ensemble TC track probability forecasts (Chapter 4) showed the benefit via improved forecast skill of using ensembles over traditional consensus techniques for forecasting the TC track, particularly at longer lead times. It also clearly demonstrated the added skill, reliability and value gained from using multi-model ensembles over any single ensemble forecast model, in every global TC basin. This study therefore incentivises the work to increase the use of ensemble forecasts in TC track warnings as it highlights the added information that is being denied to the public and emergency management communities by not including this dynamic ensemble-based probabilistic information. Following this study, verification of ensemble TC track probability forecasts continues to be updated, with ensemble TC verification expanding to also assess other TC forecast information such as intensity and rainfall.

A global analysis of TC cases and their associated flood hazard revealed that TC cases with slow-moving, large, and intense cyclones, affecting areas with wet antecedent conditions, have the highest likelihood of causing widespread flooding, with the risk of this being most likely where multiple drivers are present (Chapter 5). The knowledge gained in this study about the key meteorological and hydrological factors that lead to increased fluvial flood hazard can help improve flood early warning in TC cases and also give guidance to researchers as to where to focus forecast

improvements to maximise skill in TC flood forecasts. For example, slower-moving storms are revealed to also have an increased risk of longer and more extreme floods, so an important finding is that focus should be given to translation speed throughout the forecast process, from model evaluation and development, through to warning design and communication. The study found that river flooding can be severe in all intensity categories of TC, including those of tropical storm strength, emphasising the importance of focussing on hazard and impact rather than just the intensity category of an approaching TC.

Strong relationships between TC track, precipitation and overall flood forecast skill were found when predictability links across the flood forecast chain were assessed, with the sensitivity of discharge forecasts to the landfall location and intensity forecasts varying considerably at a catchment level, depending on the catchment's characteristics and location relative to the TC (Chapter 6). These predictability links across the flood forecast chain were assessed using novel ensemble-based methods developed to exploit the ensemble member forecasts from EC ENS and GloFAS, with these methods demonstrated for the case of Hurricane Iota which caused widespread and severe flooding in Central America. Only those catchments close to the landfall location had discharge forecasts that were sensitive to precise landfall and intensity predictions, with the rainfall from the wider storm circulation found to be crucial to flood impacts in most of the affected river basins. This highlights the importance of moving beyond landfall location and the category of the storm when forecasting and communicating TC impacts. Rainfall forecasts varied considerably even when the forecast landfall locations are similar, emphasizing how important it is to increase the focus on evaluating and understanding the skill and reliability of ensemble TC rainfall forecasts in order to improve TC flood forecasting. This study added a different angle on more traditional methods of verifying the global ensemble forecast performance for TCs, by looking at the forecasts specifically through a hazard-based hydrological lens. The research provided a first in depth look at the ability of GloFAS to forecast river flooding in a TC case. Once the forecast confidence increased for the TC genesis and track of Iota (around 5 days ahead of landfall) a clear signal was apparent for a significant flood event in the overall region that was affected, showing the potential for GloFAS to provide early flood warning information for humanitarian preparedness activities.

Organisation type	Recommendations
NWP centres	• Disseminate timely, reliable, consistently formatted feeds of ensemble forecast vortex parameter data to all operational TC forecast centres, following guidelines which will be set out by TT-TCPara task team. • Include verification of translation speed, structure, and TC rainfall alongside TC track and intensity forecasts when assessing ensemble model upgrades to ensure flood-relevant metrics are considered.
Research community	• Build an international community of practice for social science research focussed on advancing ensemble-based TC warning design to optimise understanding and behavioural response. • Engage with WMO initiatives such as TC-PFP to target research to improve the pull-through of ensemble forecast data into operational TC forecasting e.g. optimal combination of several information sources (global ensemble forecast models of different skill and resolution, any available regional ensembles, and hybrid or statistical forecasts) • Use knowledge of the key drivers of flood severity to guide an analysis of the changing fluvial flood hazard in ensemble climate projections.
Operational TC forecast centres (RSMCs/TCWCs)	• Set up operational decoding of new ensemble vortex parameter files, once these are available from NWP centres following TT-TCPara process. • Adapt TC track warnings to include the dynamic multi-model ensemble-based uncertainty information (where this has not already been done). • Expand the pull-through of multi-model ensemble forecasts into TC intensity, rainfall, and hazard forecasts and warnings, using TC-PFP and AG-TC groups to share best practice. • Promote flood-relevant information such as translation speed, track after landfall, and flood hazard and impact forecasts in TC messaging, moving away from traditional focus on landfall location and intensity category.
National Meteorological and Hydrological Services (NMHS)	• For NMHSs in countries impacted by TCs, use ensemble-based flood forecasts to improve flood early warning, working closely with partner agencies as appropriate depending on the national forecast context, and including local knowledge of catchment characteristics and management. • Promote flood-relevant information such as translation speed, track after landfall, and flood hazard and impact forecasts in TC messaging, moving away from traditional focus on landfall location and intensity category.
Global scale flood forecasting system developers	• Incorporate multi-model ensemble precipitation forecasts as driving data as this research indicates it is likely to improve TC flood forecast skill. • Create TC-specific layers within flood forecasting web interfaces (e.g. GloFAS map viewer), to increase forecast utility and take up of the forecasts within the TC forecasting community. • More routinely evaluate the performance and predictability of ensemble flood forecasts in TC cases using the methods developed in this research.
Humanitarian agencies	• Move towards using forecast information sources that include dynamic ensemble-based forecast information, and/or encourage development of existing sources that don't currently do this. • Work collaboratively with NHMSs to create sustainable agency-led ensemble-based flood early warnings using ensemble flood forecasts. • Focus on flood-relevant information such as translation speed, track after landfall, and flood hazard and impact forecasts in TC messaging, moving away from traditional focus on landfall location and intensity category.

Table 8-1. Recommendations for next steps to continue to unlock the potential of ensemble forecasts in operational TC forecasting and flood early warning, grouped according to the main communities to which they apply.

Recommended next steps to take this thesis topic forward, as discussed in Section 7.6, are summarised in Table 8.1, grouped according to the main community group to which they apply. Looking to the future, NWP centres are increasingly moving towards purely ensemble-based forecast systems, with ECMWF being the first centre to bring the ensemble to the same resolution as the deterministic as of June 2023. So the momentum towards making optimum use of the full probabilistic information that they provide in TC forecasting should continue to grow in the coming years. The work of this thesis has hopefully provided several useful steps along the way to unlocking the potential of ensemble forecasts in operational TC forecasting and early flood warning. It has provided evidence of how ensemble forecasts can improve TC forecasts and warnings, and improved knowledge of TC flood drivers and predictability. The project also leaves a legacy of increased focus on this important topic within international collaborations via its links to WMO policy, and the wider forecasting and research communities.

References

Adams, T.E. (2016) “Chapter 10 - Flood Forecasting in the United States NOAA/National Weather Service.” In Adams, T.E. and Pagano, T.C. (eds.) *Flood Forecasting*. Boston: Academic Press. pp. 249–310. doi:10.1016/B978-0-12-801884-2.00010-4.

Adams, T.E. and Pagano, T.C. (2016) *Flood Forecasting: A Global Perspective*. Academic Press. (Google-Books-ID: ttyoBAAAQBAJ).

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J. and Pappenberger, F. (2013) GloFAS - global ensemble streamflow forecasting and flood early warning. *Hydrology and Earth System Sciences*, 17 (3): 1161–1175. doi:10.5194/hess-17-1161-2013.

Alfieri, L., Zsoter, E., Harrigan, S., Aga Hirpa, F., Lavaysse, C., Prudhomme, C. and Salamon, P. (2019) Range-dependent thresholds for global flood early warning. *Journal of Hydrology X*, 4: 100034. doi:10.1016/j.hydroa.2019.100034.

Bachmann, K. and Torn, R.D. (2021) Validation of HWRF-Based Probabilistic TC Wind and Precipitation Forecasts. *Weather and Forecasting*, 36 (6): 2057–2070. doi:10.1175/WAF-D-21-0070.1.

Bagtasa, G. (2017) Contribution of Tropical Cyclones to Rainfall in the Philippines. *Journal of Climate*, 30 (10): 3621–3633. doi:10.1175/JCLI-D-16-0150.1.

Bagtasa, G. (2021) Analog forecasting of tropical cyclone rainfall in the Philippines. *Weather and Climate Extremes*, 32: 100323. doi:10.1016/j.wace.2021.100323.

Balsamo, G., Pappenberger, F., Dutra, E., Viterbo, P. and van den Hurk, B. (2011) A revised land hydrology in the ECMWF model: a step towards daily water flux prediction in a fully-closed water cycle. *Hydrological Processes*, 25 (7): 1046–1054. doi:10.1002/hyp.7808.

Bao, X., Davidson, N.E., Yu, H., Hankinson, M.C.N., Sun, Z., Rikus, L.J., Liu, J., Yu, Z. and Wu, D. (2015) Diagnostics for an Extreme Rain Event near Shanghai during the Landfall of Typhoon Fitow (2013). *Monthly Weather Review*, 143 (9): 3377–3405. doi:10.1175/MWR-D-14-00241.1.

Bates, B., Kundzewicz, Z.W., Wu, S., Burkett, V., Doell, P., Gwary, D., Hanson, C., Heij, B., Jiménez, B., Kaser, G., Kitoh, A., Kovats, S., Kumar, P., Magadza, C.H.D., Martino, D., Mata, L., Medany, M., Miller, K. and Arnell, N. (2008) *Climate Change and Water. Technical Paper of the Intergovernmental Panel on Climate Change*.

Beck, H.E., Pan, M., Roy, T., Weedon, G.P., Pappenberger, F., van Dijk, A.I.J.M., Huffman, G.J., Adler, R.F. and Wood, E.F. (2019) Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrology and Earth System Sciences*, 23 (1): 207–224. doi:10.5194/hess-23-207-2019.

Bell, M.M. (2019) “Extreme Precipitation from Tropical Cyclones.” In *The Multiscale Global Monsoon System*. World Scientific Series on Asia-Pacific Weather and Climate Volume 11. WORLD SCIENTIFIC. pp. 219–230. doi:10.1142/9789811216602_0018.

Berghuijs, W.R., Harrigan, S., Molnar, P., Slater, L.J. and Kirchner, J.W. (2019) The Relative Importance of Different Flood-Generating Mechanisms Across Europe. *Water Resources Research*, 55 (6): 4582–4593. doi:10.1029/2019WR024841.

Beven, K.J. (2012) *Rainfall-Runoff Modelling: The Primer (2nd Edition)*. John Wiley & Sons, Ltd. doi:10.1002/9781119951001.ch1.

Bidlot, J.-R., Prates, F., Ribas, R., Mueller-Quintino, A., Crepulja, M. and Vitart, F. (2020) *Enhancing tropical cyclone wind forecasts*. Available at: https://www.ecmwf.int/en/newsletter/164/meteorology/enhancing-tropical-cyclone-wind-forecasts (Accessed: 21 June 2023).

Bonnardot, F., Quetelard, H., Jumaux, G., Leroux, M.-D. and Bessafi, M. (2019) Probabilistic forecasts of tropical cyclone tracks and intensities in the southwest Indian Ocean basin. *Quarterly Journal of the Royal Meteorological Society*, 145 (719): 675–686. doi:10.1002/qj.3459.

Bousquet, O., Barbary, D., Bielli, S., Kebir, S., Raynaud, L., Malardel, S. and Faure, G. (2020) An evaluation of tropical cyclone forecast in the Southwest Indian Ocean basin with AROME-Indian Ocean convection-permitting numerical weather predicting system. *Atmospheric Science Letters*, 21 (3): e950. doi:10.1002/asl.950.

Boussetta, S., Balsamo, G., Arduini, G., Dutra, E., McNorton, J., Choulga, M., Agustí-Panareda, A., Beljaars, A., Wedi, N., Munõz-Sabater, J., de Rosnay, P., Sandu, I., Hadade, I., Carver, G., Mazzetti, C., Prudhomme, C., Yamazaki, D. and Zsoter, E. (2021) ECLand: The ECMWF Land Surface Modelling System. *Atmosphere*, 12 (6): 723. doi:10.3390/atmos12060723.

Bowler, N.E., Arribas, A., Mylne, K.R., Robertson, K.B. and Beare, S.E. (2008) The MOGREPS short-range ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 134 (632): 703–722. doi:10.1002/qj.234.

Bowler, N.E., Clayton, A.M., Jardak, M., Jermey, P.M., Lorenc, A.C., Wlasak, M.A., Barker, D.M., Inverarity, G.W. and Swinbank, R. (2017) The effect of improved ensemble covariances on hybrid variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143 (703): 785–797. doi:10.1002/qj.2964.

Brammer, A., Grogan, D., Albany, U. and Schumacher, A. (2021a) *Extending the Tropical Cyclone Genesis Index to Global Ensemble Forecasts (2021 HFIP Annual Meeting presentation)*. Available at: https://hfip.org/sites/default/files/events/269/210-brammer-enstcgi.pdf.

Brammer, A., Schumacher, A., Musgrave, K. and DeMaria, M. (2021b) *Development of a Multi-model Global Ensemble Based Tropical Cyclone Wind Speed Probability Model (HFIP Annual Meeting presentation)*. Available at: https://hfip.org/sites/default/files/events/269/1200-brammer-ensemblewsp.pdf.

Brown, J.D., Demargne, J., Seo, D.-J. and Liu, Y. (2010) The Ensemble Verification System (EVS): A software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations. *Environmental Modelling & Software*, 25 (7): 854–872. doi:10.1016/j.envsoft.2010.01.009.

Buizza, R. (2008) The value of probabilistic prediction. *Atmospheric Science Letters*, 9 (2): 36–42. doi:10.1002/asl.170.

Buizza, R. (2019) Introduction to the special issue on “25 years of ensemble forecasting.” *Quarterly Journal of the Royal Meteorological Society*, 145 (S1): 1–11. doi:10.1002/qj.3370.

Burke, P.C., Lamers, A., Carbin, G., Erickson, M.J., Klein, M., Chenard, M., McNatt, J. and Wood, L. (2023) The Excessive Rainfall Outlook at the Weather Prediction Center: Operational Definition,

Construction, and Real-Time Collaboration. *Bulletin of the American Meteorological Society*, 104 (3): E542–E562. doi:10.1175/BAMS-D-21-0281.1.

Cangialosi, J.P. (2019) 2018 Hurricane Season. *National Hurricane Center Forecast Verification Report*. Available at: https://www.nhc.noaa.gov/verification/pdfs/Verification_2018.pdf (Accessed: 22 March 2023).

Cangialosi, J.P., Blake, E., DeMaria, M., Penny, A., Latto, A., Rappaport, E. and Tallapragada, V. (2020) Recent Progress in Tropical Cyclone Intensity Forecasting at the National Hurricane Center. *Weather and Forecasting*, 35 (5): 1913–1922. doi:10.1175/WAF-D-20-0059.1.

Chen, F. and Huang, H. (2019) Comparisons of Gauge, TMPA and IMERG Products for Monsoon and Tropical Cyclone Precipitation in Southern China. *Pure and Applied Geophysics*, 176 (4): 1767–1784. doi:10.1007/s00024-018-2038-z.

Chen, J.-H., Lin, S.-J., Magnusson, L., Bender, M., Chen, X., Zhou, L., Xiang, B., Rees, S., Morin, M. and Harris, L. (2019) Advancements in Hurricane Prediction With NOAA's Next-Generation Forecast System. *Geophysical Research Letters*, 46 (8): 4495–4501. doi:10.1029/2019GL082410.

Chen, S.S., Knaff, J.A. and Marks, F.D. (2006) Effects of Vertical Wind Shear and Storm Motion on Tropical Cyclone Rainfall Asymmetries Deduced from TRMM. *Monthly Weather Review*, 134 (11): 3190–3208. doi:10.1175/MWR3245.1.

Chen, Y., Ebert, E.E., Davidson, N.E. and Walsh, K.J.E. (2018) Application of Contiguous Rain Area (CRA) Methods to Tropical Cyclone Rainfall Forecast Verification. *Earth and Space Science*, 5 (11): 736–752. doi:10.1029/2018EA000412.

Cheung, K., Yu, Z., Elsberry, R.L., Bell, M., Jiang, H., Lee, T.C., Lu, K.-C., Oikawa, Y., Qi, L., Rogers, R.F. and Tsuboki, K. (2018) Recent Advances in Research and Forecasting of Tropical Cyclone Rainfall. *Tropical Cyclone Research and Review*, 7 (2): 106–127. doi:10.6057/2018TCRR02.03.

Clark, M.P. and Hay, L.E. (2004) Use of Medium-Range Numerical Weather Prediction Model Output to Produce Forecasts of Streamflow. *Journal of Hydrometeorology*, 5 (1): 15–32. doi:10.1175/1525-7541(2004)005<0015:UOMNWP>2.0.CO;2.

Cloke, H.L. and Pappenberger, F. (2009) Ensemble flood forecasting: A review. *Journal of Hydrology*, 375 (3): 613–626. doi:10.1016/j.jhydrol.2009.06.005.

Cloke, H.L., Pappenberger, F., Jan van Andel, S., Thielen, J. and Ramos, M. (2013a) Hydrological ensemble prediction systems. *Hydrological Processes*, 27 (1): 1–4. doi:10.1002/hyp.9679.

Cloke, H.L., Pappenberger, F., Thielen, J. and Thiemig, V. (2013b) "Operational European Flood Forecasting." In *Environmental Modelling*. John Wiley & Sons, Ltd. pp. 415–434. doi:10.1002/9781118351475.ch25.

Conroy, A., Titley, H., Rivett, R., Feng, X., Methven, J., Hodges, K., Brammer, A., Burton, A., Chakraborty, P., Chen, G., Cowan, L., Dunion, J. and Sarkar, A. (2023) Track forecast: Operational capability and new techniques - Summary from the Tenth International Workshop on Tropical Cyclones (IWTC-10). *Tropical Cyclone Research and Review*, 12 (1): 64–80. doi:10.1016/j.tcrr.2023.05.002.

Coughlan de Perez, E., van den Hurk, B., van Aalst, M.K., Amuron, I., Bamanya, D., Hauser, T., Jongma, B., Lopez, A., Mason, S., Mendler de Suarez, J., Pappenberger, F., Rueth, A., Stephens, E.,

Suarez, P., Wagemaker, J. and Zsoter, E. (2016) Action-based flood forecasting for triggering humanitarian action. *Hydrology and Earth System Sciences*, 20 (9): 3549–3560. doi:10.5194/hess-20-3549-2016.

Czajkowski, J. and Kennedy, E. (2010) Fatal tradeoff? Toward a better understanding of the costs of not evacuating from a hurricane in landfall counties. *Population and Environment*, 31 (1): 121–149. doi:10.1007/s11111-009-0097-x.

Czajkowski, J., Villarini, G., Michel-Kerjan, E. and Smith, J.A. (2013) Determining tropical cyclone inland flooding loss on a large scale through a new flood peak ratio-based methodology. *Environmental Research Letters*, 8 (4): 044056. doi:10.1088/1748-9326/8/4/044056.

Czajkowski, J., Villarini, G., Montgomery, M., Michel-Kerjan, E. and Goska, R. (2017) Assessing Current and Future Freshwater Flood Risk from North Atlantic Tropical Cyclones via Insurance Claims. *Scientific Reports*, 7 (1): 41609. doi:10.1038/srep41609.

Davis, C., Brown, B. and Bullock, R. (2006) Object-Based Verification of Precipitation Forecasts. Part I: Methodology and Application to Mesoscale Rain Areas. *Monthly Weather Review*, 134 (7): 1772–1784. doi:10.1175/MWR3145.1.

Davis, C.A., Brown, B.G., Bullock, R. and Halley-Gotway, J. (2009) The Method for Object-Based Diagnostic Evaluation (MODE) Applied to Numerical Forecasts from the 2005 NSSL/SPC Spring Program. *Weather and Forecasting*, 24 (5): 1252–1267. doi:10.1175/2009WAF2222241.1.

DeHart, J.C. and Houze, R.A. (2017) Orographic Modification of Precipitation Processes in Hurricane Karl (2010). *Monthly Weather Review*, 145 (10): 4171–4186. doi:10.1175/MWR-D-17-0014.1.

Demargne, J., Brown, J., Liu, Y., Seo, D.-J., Wu, L., Toth, Z. and Zhu, Y. (2010) Diagnostic verification of hydrometeorological and hydrologic ensembles. *Atmospheric Science Letters*, 11 (2): 114–122. doi:10.1002/asl.261.

DeMaria, M., Knaff, J.A. and Connell, B.H. (2001) A Tropical Cyclone Genesis Parameter for the Tropical Atlantic. *Weather and Forecasting*, 16 (2): 219–233. doi:10.1175/1520-0434(2001)016<0219:ATCGPF>2.0.CO;2.

den Toom, M., Verkade, J., Weerts, A., and Schotmeijer, G.J. (2020) Development of the Deltares global fluvial flood forecast system. *EGU General Assembly 2020*, Online, 4–8 May 2020, EGU2020-22344, https://doi.org/10.5194/egusphere-egu2020-22344, 2020.

Deng, D., Davidson, N.E., Hu, L., Tory, K.J., Hankinson, M.C.N. and Gao, S. (2017) Potential Vorticity Perspective of Vortex Structure Changes of Tropical Cyclone Bilis (2006) during a Heavy Rain Event following Landfall. *Monthly Weather Review*, 145 (5): 1875–1895. doi:10.1175/MWR-D-16-0276.1.

Dowell, D.C., Alexander, C.R., James, E.P., Weygandt, S.S., Benjamin, S.G., Manikin, G.S., Blake, B.T., Brown, J.M., Olson, J.B., Hu, M., Smirnova, T.G., Ladwig, T., Kenyon, J.S., Ahmadov, R., Turner, D.D., Duda, J.D. and Alcott, T.I. (2022) The High-Resolution Rapid Refresh (HRRR): An Hourly Updating Convection-Allowing Forecast Model. Part I: Motivation and System Description. *Weather and Forecasting*, 37 (8): 1371–1395. doi:10.1175/WAF-D-21-0151.1.

Du, Y., Qi, L. and Cao, X. (2016) Selective ensemble-mean technique for tropical cyclone track forecast by using time-lagged ensemble and multi-centre ensemble in the western North Pacific.

Quarterly Journal of the Royal Meteorological Society, 142 (699): 2452–2462. doi:10.1002/qj.2838.

Dube, A., Karunasagar, S., Ashrit, R. and Mitra, A.K. (2022) Spatial verification of ensemble rainfall forecasts over India. *Atmospheric Research*, 273: 106169. doi:10.1016/j.atmosres.2022.106169.

Dunion, J., Davis, C., Titley, H.A., Greatrex, H., Yamaguchi, M., Methven, J., Ashrit, R., Wang, Z., Yu, H., Fontan, A.-C., Brammer, A., Kucas, M., Ford, M., Papin, P.P., Prates, F., Kruczkiewwicz, A., Chakraborty, P., Burton, A., DeMaria, M., Mooney, C., Torn, R. and Vigh, J. (2023) Recommendations for improved tropical cyclone formation and position probabilistic forecast products: Outcomes from Phase 1 of the WMO/WWRP TC-PFP project (Submitted). *Tropical Cyclone Research and Review*.

Dupont, T., Plu, M., Caroff, P. and Faure, G. (2011) Verification of Ensemble-Based Uncertainty Circles around Tropical Cyclone Track Forecasts. *Weather and Forecasting*, 26 (5): 664–676. doi:10.1175/WAF-D-11-00007.1.

Ebert, E.E. and Coauthors (2021) *Value Chain Approaches to Evaluate the End-to-End Warning Chain: WMO HIWeather Warning Value Chain Project Plan v1.2*. Available at: https://filecloud.wmo.int/share/s/PXkVpHZ6TE2BCOxCkjqfJQ (Accessed: 23 June 2023).

Ebert, E.E. and Gallus, W.A. (2009) Toward Better Understanding of the Contiguous Rain Area (CRA) Method for Spatial Forecast Verification. *Weather and Forecasting*, 24 (5): 1401–1415. doi:10.1175/2009WAF2222252.1.

Ebert, E.E. and McBride, J.L. (2000) Verification of precipitation in weather systems: determination of systematic errors. *Journal of Hydrology*, 239 (1): 179–202. doi:10.1016/S0022-1694(00)00343-7.

Ebert, E.E., Turk, M., Kusselson, S.J., Yang, J., Seybold, M., Keehn, P.R. and Kuligowski, R.J. (2011) Ensemble Tropical Rainfall Potential (eTRaP) Forecasts. *Weather and Forecasting*, 26 (2): 213–224. doi:10.1175/2010WAF2222443.1.

ECMWF (2020a) *IFS Documentation CY47R1 - Part IV: Physical Processes*. Available at: https://doi.org/10.21957/cpmkqvhja (Accessed: 25 April 2023).

ECMWF (2020b) *IFS Documentation CY47R1 - Part V: Ensemble Prediction System*. Available at: 10.21957/d7e3hrb (Accessed: 19 June 2023).

ECMWF (2022) *Access to archive datasets*. Available at: https://www.ecmwf.int/en/forecasts/access-forecasts/access-archive-datasets (Accessed: 24 May 2022).

ECMWF (2023) *Implementation of IFS Cycle 48r1 - Forecast User - ECMWF Confluence Wiki*. Available at: https://confluence.ecmwf.int/pages/viewpage.action?pageId=246483247 (Accessed: 9 March 2023).

Elliott, G. and Yamaguchi, M. (2014) *Advances in Forecasting Motion*. 8th International Workshop on Tropical Cyclones Topic 1. Jeju, Republic of Korea. Available at: https://wmoomm.sharepoint.com/sites/wmocpdb/eve_activityarea/Forms/AllItems.aspx?id=%2Fsites%2Fwmocpdb%2Feve%5Factivityarea%2FWorld%20Weather%20Research%20Programme%20%28WWRP%29%5Fed35626e%2D1373%2De911%2Da965%2D000d3a396ff4%2FTopic1%5FAdvancesinForecastingMotion%2Epdf&parent=%2Fsites%2Fwmocpdb%2Feve%5Factivityarea%2FWorl

d%20Weather%20Research%20Programme%20%28WWRP%29%5Fed35626e%2D1373%2De911%2Da965%2D000d3a396ff4&p=true&ga=1 (Accessed: 22 March 2023).

Emanuel, K. (2018) 100 Years of Progress in Tropical Cyclone Research. *Meteorological Monographs*, 59 (1): 15.1-15.68. doi:10.1175/AMSMONOGRAPHS-D-18-0016.1.

Emerton, R., Cloke, H., Ficchi, A., Hawker, L., de Wit, S., Speight, L., Prudhomme, C., Rundell, P., West, R., Neal, J., Cuna, J., Harrigan, S., Titley, H., Magnusson, L., Pappenberger, F., Klingaman, N. and Stephens, E. (2020) Emergency flood bulletins for Cyclones Idai and Kenneth: A critical evaluation of the use of global flood forecasts for international humanitarian preparedness and response. *International Journal of Disaster Risk Reduction*, 50: 101811. doi:10.1016/j.ijdrr.2020.101811.

Emerton, R.E., Stephens, E.M., Pappenberger, F., Pagano, T.C., Weerts, A.H., Wood, A.W., Salamon, P., Brown, J.D., Hjerdt, N., Donnelly, C., Baugh, C.A. and Cloke, H.L. (2016) Continental and global scale flood forecasting systems. *WIREs Water*, 3 (3): 391–418. doi:10.1002/wat2.1137.

Erickson, M.J., Albright, B. and Nelson, J.A. (2021) Verifying and Redefining the Weather Prediction Center's Excessive Rainfall Outlook Forecast Product. *Weather and Forecasting*, 36 (1): 325–340. doi:10.1175/WAF-D-20-0020.1.

Faure, G., Chambon, P. and Brousseau, P. (2020) Operational Implementation of the AROME Model in the Tropics: Multiscale Validation of Rainfall Forecasts. *Weather and Forecasting*, 35 (2): 691–710. doi:10.1175/WAF-D-19-0204.1.

Ficchì, A. and Stephens, L. (2019) Climate Variability Alters Flood Timing Across Africa. *Geophysical Research Letters*, 46 (15): 8809–8819. doi:10.1029/2019GL081988.

Franchini, M., Bernini, A., Barbetta, S. and Moramarco, T. (2011) Forecasting discharges at the downstream end of a river reach through two simple Muskingum based procedures. *Journal of Hydrology*, 399 (3): 335–352. doi:10.1016/j.jhydrol.2011.01.009.

Gallus, W.A. (2010) Application of Object-Based Verification Techniques to Ensemble Precipitation Forecasts. *Weather and Forecasting*, 25 (1): 144–158. doi:10.1175/2009WAF2222274.1.

GDACS (2023) *GDACS - Global Disaster Alert and Coordination System*. Available at: https://www.gdacs.org/Knowledge/models_TC.aspx (Accessed: 23 March 2023).

Gilleland, E., Ahijevych, D., Brown, B.G., Casati, B. and Ebert, E.E. (2009) Intercomparison of Spatial Forecast Verification Methods. *Weather and Forecasting*, 24 (5): 1416–1430. doi:10.1175/2009WAF2222269.1.

Gilleland, E., Ahijevych, D.A., Brown, B.G. and Ebert, E.E. (2010) Verifying Forecasts Spatially. *Bulletin of the American Meteorological Society*, 91 (10): 1365–1376. doi:10.1175/2010BAMS2819.1.

Gneiting, T. and Raftery, A.E. (2007) Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102 (477): 359–378. doi:10.1198/016214506000001437.

Goerss, J.S. (2000) Tropical Cyclone Track Forecasts Using an Ensemble of Dynamical Models. *Monthly Weather Review*, 128 (4): 1187–1193. doi:10.1175/1520-0493(2000)128<1187:TCTFUA>2.0.CO;2.

Goerss, J.S. (2007) Prediction of Consensus Tropical Cyclone Track Forecast Error. *Monthly Weather Review*, 135 (5): 1985–1993. doi:10.1175/MWR3390.1.

Goerss, J.S. and Sampson, C.R. (2014) Prediction of Consensus Tropical Cyclone Intensity Forecast Error. *Weather and Forecasting*, 29 (3): 750–762. doi:10.1175/WAF-D-13-00058.1.

Golding, B., Sun, J., Riemer, M., Yussouf, N., Titley, H., Robbins, J., Ebert, B., Pagano, T., Lewis, H., Dashwood, C., Boyce, G. and Peace, M. (2022) "Connecting Weather and Hazard: A Partnership of Physical Scientists in Connected Disciplines." In Golding, B. (ed.) *Towards the "Perfect" Weather Warning: Bridging Disciplinary Gaps through Partnership and Communication*. Cham: Springer International Publishing. pp. 149–200. doi:10.1007/978-3-030-98989-7_6.

Gray, W.M. (1968) GLOBAL VIEW OF THE ORIGIN OF TROPICAL DISTURBANCES AND STORMS. *Monthly Weather Review*, 96 (10): 669–700. doi:10.1175/1520-0493(1968)096<0669:GVOTOO>2.0.CO;2.

Guo, L., Klingaman, N.P., Vidale, P.L., Turner, A.G., Demory, M.-E. and Cobb, A. (2017) Contribution of Tropical Cyclones to Atmospheric Moisture Transport and Rainfall over East Asia. *Journal of Climate*, 30 (10): 3853–3865. doi:10.1175/JCLI-D-16-0308.1.

Guo, Q. and Matyas, C.J. (2016) Comparing the spatial extent of Atlantic basin tropical cyclone wind and rain fields prior to land interaction. *Physical Geography*, 37 (1): 5–25. doi:10.1080/02723646.2016.1142929.

Gupta, H.V., Kling, H., Yilmaz, K.K. and Martinez, G.F. (2009) Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377 (1): 80–91. doi:10.1016/j.jhydrol.2009.08.003.

Guzman, O. and Jiang, H. (2021) Global increase in tropical cyclone rain rate. *Nature Communications*, 12 (1): 5344. doi:10.1038/s41467-021-25685-2.

Hagedorn, R., Buizza, R., Hamill, T.M., Leutbecher, M. and Palmer, T.N. (2012) Comparing TIGGE multimodel forecasts with reforecast-calibrated ECMWF ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 138 (668): 1814–1827. doi:10.1002/qj.1895.

Hagedorn, R., Doblas-Reyes, F.J. and Palmer, T.N. (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concept. *Tellus A*, 57 (3): 219–233. doi:10.1111/j.1600-0870.2005.00103.x.

Hakim, G.J. and Torn, R.D. (2008) "Ensemble Synoptic Analysis." In Bosart, L.F. and Bluestein, H.B. (eds.) *Synoptic—Dynamic Meteorology and Weather Analysis and Forecasting: A Tribute to Fred Sanders*. Meteorological Monographs. Boston, MA: American Meteorological Society. pp. 147–161. doi:10.1007/978-0-933876-68-2_7.

Hamill, T.M. (2012) Verification of TIGGE Multimodel and ECMWF Reforecast-Calibrated Probabilistic Precipitation Forecasts over the Contiguous United States*. *Monthly Weather Review*, 140 (7): 2232–2252. doi:10.1175/MWR-D-11-00220.1.

Hamill, T.M., Brennan, M.J., Brown, B., DeMaria, M., Rappaport, E.N. and Toth, Z. (2012) NOAA's Future Ensemble-Based Hurricane Forecast Products. *Bulletin of the American Meteorological Society*, 93 (2): 209–220. doi:10.1175/2011BAMS3106.1.

Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., Barnard, C., Cloke, H. and Pappenberger, F. (2020) GloFAS-ERA5 operational global river discharge reanalysis 1979–present. *Earth System Science Data*, 12 (3): 2043–2060. doi:10.5194/essd-12-2043-2020.

Harrigan, S., Zsoter, E., Barnard, C., Wetterhall, F., Ferrario, I., Mazzetti, C., Alfieri, L., Salamon, P. and Prudhomme, C. (2019) *River discharge and related historical data from the Global Flood Awareness System. v2.1*. Available at: 10.24381/cds.a4fdd6b9 (Accessed: 8 October 2020).

Harrigan, S., Zsoter, E., Barnard, C., Wetterhall, F., Ferrario, I., Mazzetti, C., Alfieri, L., Salamon, P. and Prudhomme, C. (2021) *River discharge and related historical data from the Global Flood Awareness System. v2.1*. Available at: 10.24381/cds.ff1aef77 (Accessed: 23 March 2023).

Harrigan, S., Zsoter, E., Cloke, H., Salamon, P. and Prudhomme, C. (2023) Daily ensemble river discharge reforecasts and real-time forecasts from the operational Global Flood Awareness System. *Hydrology and Earth System Sciences*, 27 (1): 1–19. doi:10.5194/hess-27-1-2023.

He, Y., Wetterhall, F., Cloke, H.L., Pappenberger, F., Wilson, M., Freer, J. and McGregor, G. (2009) Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions. *Meteorological Applications*, 16 (1): 91–101. doi:10.1002/met.132.

Heming, J.T. (2017) Tropical cyclone tracking and verification techniques for Met Office numerical weather prediction models. *Meteorological Applications*, 24 (1): 1–8. doi:10.1002/met.1599.

Heming, J.T., Prates, F., Bender, M.A., Bowyer, R., Cangialosi, J., Caroff, P., Coleman, T., Doyle, J.D., Dube, A., Faure, G., Fraser, J., Howell, B.C., Igarashi, Y., McTaggart-Cowan, R., Mohapatra, M., Moskaitis, J.R., Murtha, J., Rivett, R., Sharma, M., Short, C.J., Singh, A.A., Tallapragada, V., Titley, H.A. and Xiao, Y. (2019) Review of Recent Progress in Tropical Cyclone Track Forecasting and Expression of Uncertainties. *Tropical Cyclone Research and Review*, 8 (4): 181–218. doi:10.1016/j.tcrr.2020.01.001.

Hernández Ayala, J.J. and Matyas, C.J. (2016) Tropical cyclone rainfall over Puerto Rico and its relations to environmental and storm-specific factors. *International Journal of Climatology*, 36 (5): 2223–2237. doi:10.1002/joc.4490.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. and Thépaut, J.-N. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146 (730): 1999–2049. doi:10.1002/qj.3803.

Hersbach, H. and Coauthors (2018a) *ERA5 hourly data on single levels from 1979 to present.* doi:10.24381/cds.adbb2d47.

Hersbach, H. and Coauthors (2018b) *Operational global reanalysis: progress, future directions and synergies with NWP*. ERA Report Series no. 27. Reading, UK: ECMWF.

Hewson, T.D. and Titley, H.A. (2010) Objective identification, typing and tracking of the complete life-cycles of cyclonic features at high spatial resolution. *Meteorological Applications*, 17 (3): 355–381. doi:10.1002/met.204.

Hirpa, F.A., Salamon, P., Beck, H.E., Lorini, V., Alfieri, L., Zsoter, E. and Dadson, S.J. (2018) Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data. *Journal of Hydrology*, 566: 595–606. doi:10.1016/j.jhydrol.2018.09.052.

Hodges, K.I. (1995) Feature Tracking on the Unit Sphere. *Monthly Weather Review*, 123 (12): 3458–3465. doi:10.1175/1520-0493(1995)123<3458:FTOTUS>2.0.CO;2.

Hong, J.-S., Fong, C.-T., Hsiao, L.-F., Yu, Y.-C. and Tzeng, C.-Y. (2015) Ensemble Typhoon Quantitative Precipitation Forecasts Model in Taiwan. *Weather and Forecasting*, 30 (1): 217–237. doi:10.1175/WAF-D-14-00037.1.

Hossain, S., Cloke, H.L., Ficchì, A., Gupta, H., Speight, L., Hassan, A. and Stephens, E. (2023) A decision-led evaluation approach for flood forecasting system developments: An application to the Global Flood Awareness System in Bangladesh (submitted). *Journal of Flood Risk Management*.

Hou, A.Y., Kakar, R.K., Neeck, S., Azarbarzin, A.A., Kummerow, C.D., Kojima, M., Oki, R., Nakamura, K. and Iguchi, T. (2014) The Global Precipitation Measurement Mission. *Bulletin of the American Meteorological Society*, 95 (5): 701–722. doi:10.1175/BAMS-D-13-00164.1.

Houtekamer, P.L., Mitchell, H.L. and Deng, X. (2009) Model Error Representation in an Operational Ensemble Kalman Filter. *Monthly Weather Review*, 137 (7): 2126–2143. doi:10.1175/2008MWR2737.1.

Houze, R.A. (2010) Clouds in Tropical Cyclones. *Monthly Weather Review*, 138 (2): 293–344. doi:10.1175/2009MWR2989.1.

Hsiao, L.-F., Yang, M.-J., Lee, C.-S., Kuo, H.-C., Shih, D.-S., Tsai, C.-C., Wang, C.-J., Chang, L.-Y., Chen, D.Y.-C., Feng, L., Hong, J.-S., Fong, C.-T., Chen, D.-S., Yeh, T.-C., Huang, C.-Y., Guo, W.-D. and Lin, G.-F. (2013) Ensemble forecasting of typhoon rainfall and floods over a mountainous watershed in Taiwan. *Journal of Hydrology*, 506: 55–68. doi:10.1016/j.jhydrol.2013.08.046.

Hu, C.-C. and Wu, C.-C. (2020) Ensemble Sensitivity Analysis of Tropical Cyclone Intensification Rate during the Development Stage. *Journal of the Atmospheric Sciences*, 77 (10): 3387–3405. doi:10.1175/JAS-D-19-0196.1.

Hu, P., Zhang, Q., Shi, P., Chen, B. and Fang, J. (2018) Flood-induced mortality across the globe: Spatiotemporal pattern and influencing factors. *Science of The Total Environment*, 643: 171–182. doi:10.1016/j.scitotenv.2018.06.197.

Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E.J., Sorooshian, S., Tan, J. and Xie, P. (2019) *NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG)*. Available at: https://gpm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V06.pdf.

IFRC (2020) *Revised emergency appeal: Central America – hurricane Eta and Iota, Red Cross & Red Crescent International*. Available at: https://reliefweb.int/report/honduras/central-america-hurricane-eta-iota-revised-emergency-appeal-n-mdr43007-revision-no-1 (Accessed: 26 August 2022).

IPCC (2012) *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change*. ; Cambridge University Press: Cambridge, UK; New York, NY, USA.

Jain, S.K., Mani, P., Jain, S.K., Prakash, P., Singh, V.P., Tullos, D., Kumar, S., Agarwal, S.P. and Dimri, A.P. (2018) A Brief review of flood forecasting techniques and their applications. *International Journal of River Basin Management*, 16 (3): 329–344. doi:10.1080/15715124.2017.1411920.

Ji, L., Zhi, X., Simmer, C., Zhu, S. and Ji, Y. (2020) Multimodel Ensemble Forecasts of Precipitation Based on an Object-Based Diagnostic Evaluation. *Monthly Weather Review*, 148 (6): 2591–2606. doi:10.1175/MWR-D-19-0266.1.

Jia, L., Ren, F., Ding, C., Jia, Z., Wang, M., Chen, Y. and Feng, T. (2022) Improvement of the Ensemble Methods in the Dynamical–Statistical–Analog Ensemble Forecast Model for Landfalling Typhoon Precipitation. *Journal of the Meteorological Society of Japan. Ser. II*, 100 (3): 575–592. doi:10.2151/jmsj.2022-029.

Jia-jie, X., Hui, Y. and Pei-yan, C. (2021) Evaluation of Tropical Cyclone Intensity Forecasts from Five Global Ensemble Prediction Systems During 2015-2019. *Journal of Tropical Meteorology*, 27 (3): 218–231. doi:10.46267/j.1006-8775.2021.020.

Jiang, H., Liu, C. and Zipser, E.J. (2011) A TRMM-Based Tropical Cyclone Cloud and Precipitation Feature Database. *Journal of Applied Meteorology and Climatology*, 50 (6): 1255–1274. doi:10.1175/2011JAMC2662.1.

Jiang, Q., Li, W., Fan, Z., HE, X., Sun, W., Chen, S., Wen, J., Gao, J. and Wang, J. (2021) Evaluation of the ERA5 reanalysis precipitation dataset over Chinese Mainland. *Journal of Hydrology*, 595: 125660. doi:10.1016/j.jhydrol.2020.125660.

Jiang, X., Ren, F., Li, Y., Qiu, W., Ma, Z. and Cai, Q. (2018) Characteristics and Preliminary Causes of Tropical Cyclone Extreme Rainfall Events over Hainan Island. *Advances in Atmospheric Sciences*, 35 (5): 580–591. doi:10.1007/s00376-017-7051-0.

Johnson, C. and Swinbank, R. (2009) Medium-range multimodel ensemble combination and calibration. *Quarterly Journal of the Royal Meteorological Society*, 135 (640): 777–794. doi:10.1002/qj.383.

Jolliffe, I.T. and Stephenson, D.B. (2012) *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley & Sons. (Google-Books-ID: DCxsKQeaBH8C).

Jones, S., Golding, B. and Members of the HIWeather Task Teams (2014) *HIWeather: A research activity on High Impact Weather within the World Weather Research Programme*. Geneva: WMO. Available at: https://library.wmo.int/index.php?lvl=notice_display&id=19856#.ZEE2yHbMJD8 (Accessed: 20 April 2023).

Jurlina, T., Baugh, C., Pappenberger, F. and Prudhomme, C. (2020) Flood hazard risk forecasting index (FHRFI) for urban areas: The Hurricane Harvey case study. *Meteorological Applications*, 27 (1): e1845. doi:10.1002/met.1845.

JWGFVR (2013) *Verification methods for tropical cyclone forecasts*. WWRP 2013-7. WMO. Available at: https://filecloud.wmo.int/share/s/Fotf-7H1RLyK1lkSC_onBA (Accessed: 23 March 2023).

Khouakhi, A., Villarini, G. and Vecchi, G.A. (2017) Contribution of Tropical Cyclones to Rainfall at the Global Scale. *Journal of Climate*, 30 (1): 359–372. doi:10.1175/JCLI-D-16-0298.1.

Kidd, C., Becker, A., Huffman, G.J., Muller, C.L., Joe, P., Skofronick-Jackson, G. and Kirschbaum, D.B. (2017) So, How Much of the Earth's Surface Is Covered by Rain Gauges? *Bulletin of the American Meteorological Society*, 98 (1): 69–78. doi:10.1175/BAMS-D-14-00283.1.

Kidder, S.Q., Knaff, J.A., Kusselson, S.J., Turk, M., Ferraro, R.R. and Kuligowski, R.J. (2005) The Tropical Rainfall Potential (TRaP) Technique. Part I: Description and Examples. *Weather and Forecasting*, 20 (4): 456–464. doi:10.1175/WAF860.1.

Kim, H.-J., Moon, I.-J. and Kim, M. (2020a) Statistical prediction of typhoon-induced accumulated rainfall over the Korean Peninsula based on storm and rainfall data. *Meteorological Applications*, 27 (1): e1853. doi:10.1002/met.1853.

Kim, J.-S., Chen, A., Lee, J., Moon, I.-J. and Moon, Y.-I. (2020b) Statistical Prediction of Typhoon-Induced Rainfall over China Using Historical Rainfall, Tracks, and Intensity of Typhoon in the Western North Pacific. *Remote Sensing*, 12 (24): 4133. doi:10.3390/rs12244133.

Kling, H., Fuchs, M. and Paulin, M. (2012) Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, 424–425: 264–277. doi:10.1016/j.jhydrol.2012.01.011.

Knaff, J.A., DeMaria, M., Sampson, C.R. and Gross, J.M. (2003) Statistical, 5-Day Tropical Cyclone Intensity Forecasts Derived from Climatology and Persistence. *Weather and Forecasting*, 18 (1): 80–92. doi:10.1175/1520-0434(2003)018<0080:SDTCIF>2.0.CO;2.

Knapp, K.R., Kruk, M.C., Levinson, D.H., Diamond, H.J. and Neumann, C.J. (2010) The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying Tropical Cyclone Data. *Bulletin of the American Meteorological Society*, 91 (3): 363–376. doi:10.1175/2009BAMS2755.1.

Knapp, K.R., Kruk, M.C., Levinson, D.H., Diamond, H.J. and Neumann, C.J. (2018) *International Best Track Archive for Climate Stewardship (IBTrACS) Project, Version 4*. Available at: https://doi.org/10.25921/82ty-9e16 (Accessed: 28 September 2020).

Knoben, W.J.M., Freer, J.E. and Woods, R.A. (2019) Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, 23 (10): 4323–4331. doi:10.5194/hess-23-4323-2019.

Knutson, T., Camargo, S.J., Chan, J.C.L., Emanuel, K., Ho, C.-H., Kossin, J., Mohapatra, M., Satoh, M., Sugi, M., Walsh, K. and Wu, L. (2020) Tropical Cyclones and Climate Change Assessment: Part II: Projected Response to Anthropogenic Warming. *Bulletin of the American Meteorological Society*, 101 (3): E303–E322. doi:10.1175/BAMS-D-18-0194.1.

Knutson, T.R., McBride, J.L., Chan, J., Emanuel, K., Holland, G., Landsea, C., Held, I., Kossin, J.P., Srivastava, A.K. and Sugi, M. (2010) Tropical cyclones and climate change. *Nature Geoscience*, 3 (3): 157–163. doi:10.1038/ngeo779.

Knutson, T.R., Sirutis, J.J., Vecchi, G.A., Garner, S., Zhao, M., Kim, H.-S., Bender, M., Tuleya, R.E., Held, I.M. and Villarini, G. (2013) Dynamical Downscaling Projections of Twenty-First-Century Atlantic Hurricane Activity: CMIP3 and CMIP5 Model-Based Scenarios. *Journal of Climate*, 26 (17): 6591–6617. doi:10.1175/JCLI-D-12-00539.1.

Knutson, T.R., Sirutis, J.J., Zhao, M., Tuleya, R.E., Bender, M., Vecchi, G.A., Villarini, G. and Chavas, D. (2015) Global Projections of Intense Tropical Cyclone Activity for the Late Twenty-First Century

from Dynamical Downscaling of CMIP5/RCP4.5 Scenarios. *Journal of Climate*, 28 (18): 7203–7224. doi:10.1175/JCLI-D-15-0129.1.

Ko, M.-C., Marks, F.D., Alaka, G.J. and Gopalakrishnan, S.G. (2020) Evaluation of Hurricane Harvey (2017) Rainfall in Deterministic and Probabilistic HWRF Forecasts. *Atmosphere*, 11 (6): 666. doi:10.3390/atmos11060666.

Kossin, J.P. (2018) A global slowdown of tropical-cyclone translation speed. *Nature*, 558 (7708): 104–107. doi:10.1038/s41586-018-0158-3.

Lalaurette, F. (2003) Early detection of abnormal weather conditions using a probabilistic extreme forecast index. *Quarterly Journal of the Royal Meteorological Society*, 129 (594): 3037–3057. doi:10.1256/qj.02.152.

Lamers, A., Devi, S., Galvez, J.M., Yu, Z., Kriat, T., Cardos, S., Grant, D. and Moron, L. (2022) *Forecasting Tropical Cyclone Rainfall and Flooding Hazards and Impacts, Sub-topic 5.1 report, Tenth International workshop on tropical cyclones (IWTC-10), Bali, Indonesia*. Tenth International workshop on tropical cyclones (IWTC-10). Available at: https://wmoomm.sharepoint.com/sites/Services/Shared%20Documents/Forms/AllItems.aspx?id=%2Fsites%2FServices%2FShared%20Documents%2FDPS%5FDRR%20and%20Public%20Services%2FTCP%2FIWTC%2DX%2Freports%2FTopic%205%20%2D%20TC%20hazards%2FIWTC%2D10%5FTopic5%2E1%5Frainfall%5Freport%2Epdf&parent=%2Fsites%2FServices%2FShared%20Documents%2FDPS%5FDRR%20and%20Public%20Services%2FTCP%2FIWTC%2DX%2Freports%2FTopic%205%20%2D%20TC%20hazards&p=true&ga=1 (Accessed: 14 March 2023).

Landsea, C.W. and Cangialosi, J.P. (2018) Have We Reached the Limits of Predictability for Tropical Cyclone Track Forecasting? *Bulletin of the American Meteorological Society*, 99 (11): 2237–2243. doi:10.1175/BAMS-D-17-0136.1.

Lang, S.T.K., Leutbecher, M. and Jones, S.C. (2012) Impact of perturbation methods in the ECMWF ensemble prediction system on tropical cyclone forecasts. *Quarterly Journal of the Royal Meteorological Society*, 138 (669): 2030–2046. doi:10.1002/qj.1942.

Lavers, D.A., Harrigan, S., Andersson, E., Richardson, D.S., Prudhomme, C. and Pappenberger, F. (2019) A vision for improving global flood forecasting. *Environmental Research Letters*, 14 (12): 121002. doi:10.1088/1748-9326/ab52b2.

Lentze, G. (2023) *Model upgrade increases skill and unifies medium-range resolutions*. Available at: https://www.ecmwf.int/en/about/media-centre/news/2023/model-upgrade-increases-skill-and-unifies-medium-range-resolutions (Accessed: 27 June 2023).

Leonardo, N.M. and Colle, B.A. (2017) Verification of Multimodel Ensemble Forecasts of North Atlantic Tropical Cyclones. *Weather and Forecasting*, 32 (6): 2083–2101. doi:10.1175/WAF-D-17-0058.1.

Leutbecher, M., Lock, S.-J., Ollinaho, P., Lang, S.T.K., Balsamo, G., Bechtold, P., Bonavita, M., Christensen, H.M., Diamantakis, M., Dutra, E., English, S., Fisher, M., Forbes, R.M., Goddard, J., Haiden, T., Hogan, R.J., Juricke, S., Lawrence, H., MacLeod, D., Magnusson, L., Malardel, S., Massart, S., Sandu, I., Smolarkiewicz, P.K., Subramanian, A., Vitart, F., Wedi, N. and Weisheimer, A. (2017) Stochastic representations of model uncertainties at ECMWF: state of the art and future vision. *Quarterly Journal of the Royal Meteorological Society*, 143 (707): 2315–2339. doi:10.1002/qj.3094.

Li, J. and Liu, J. (2022) The Analysis of the Asymmetric Precipitation Caused by the No.10 Tropical Cyclone “Damrey” in 2012 Based on Observation and Numerical Simulation. *Journal of Geoscience and Environment Protection*, 10 (8): 145–157. doi:10.4236/gep.2022.108011.

Lin, P., Pan, M., Beck, H.E., Yang, Y., Yamazaki, D., Frasson, R., David, C.H., Durand, M., Pavelsky, T.M., Allen, G.H., Gleason, C.J. and Wood, E.F. (2019) Global Reconstruction of Naturalized River Flows at 2.94 Million Reaches. *Water Resources Research*, 55 (8): 6499–6516. doi:10.1029/2019WR025287.

Liu, Y.-Y., Li, L., Liu, Y.-S., Chan, P.-W., Zhang, W.-H. and Zhang, L. (2021) Estimation of precipitation induced by tropical cyclones based on machine-learning-enhanced analogue identification of numerical prediction. *Meteorological Applications*, 28 (2): e1978. doi:10.1002/met.1978.

Lonfat, M., Marks, F.D. and Chen, S.S. (2004) Precipitation Distribution in Tropical Cyclones Using the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager: A Global Perspective. *Monthly Weather Review*, 132 (7): 1645–1660. doi:10.1175/1520-0493(2004)132<1645:PDITCU>2.0.CO;2.

Lonfat, M., Rogers, R., Marchok, T. and Marks, F.D. (2007) A Parametric Model for Predicting Hurricane Rainfall. *Monthly Weather Review*, 135 (9): 3086–3097. doi:10.1175/MWR3433.1.

Lorenz, E.N. (1963) Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, 20 (2): 130–141. doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

Lorenz, E.N. (1969) Atmospheric Predictability as Revealed by Naturally Occurring Analogues. *Journal of the Atmospheric Sciences*, 26 (4): 636–646. doi:10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2.

Luitel, B., Villarini, G. and Vecchi, G.A. (2018) Verification of the skill of numerical weather prediction models in forecasting rainfall from U.S. landfalling tropical cyclones. *Journal of Hydrology*, 556: 1026–1037. doi:10.1016/j.jhydrol.2016.09.019.

Lynch, S.L. and Schumacher, R.S. (2014) Ensemble-Based Analysis of the May 2010 Extreme Rainfall in Tennessee and Kentucky. *Monthly Weather Review*, 142 (1): 222–239. doi:10.1175/MWR-D-13-00020.1.

MacLeod, D., Easton-Calabria, E., de Perez, E.C. and Jaime, C. (2021) Verification of forecasts for extreme rainfall, tropical cyclones, flood and storm surge over Myanmar and the Philippines. *Weather and Climate Extremes*, 33: 100325. doi:10.1016/j.wace.2021.100325.

Magnusson, L., Bidlot, J.-R., Bonavita, M., Brown, A.R., Browne, P.A., Chiara, G.D., Dahoui, M., Lang, S.T.K., McNally, T., Mogensen, K.S., Pappenberger, F., Prates, F., Rabier, F., Richardson, D.S., Vitart, F. and Malardel, S. (2019) ECMWF Activities for Improved Hurricane Forecasts. *Bulletin of the American Meteorological Society*, 100 (3): 445–458. doi:10.1175/BAMS-D-18-0044.1.

Magnusson, L., Majumdar, S., Emerton, R., Richardson, D.S. and et al. (2021) *Tropical cyclone activities at ECMWF*. ECMWF Technical Memoranda 888. Reading, UK: ECMWF. Available at: 10.21957/zzxzzygwv.

Majumdar, S.J. and Finocchio, P.M. (2010) On the Ability of Global Ensemble Prediction Systems to Predict Tropical Cyclone Track Probabilities. *Weather and Forecasting*, 25 (2): 659–680. doi:10.1175/2009WAF2222327.1.

Mamgain, A., Sarkar, A. and Rajagopal, E.N. (2020) Medium-range global ensemble prediction system at 12 km horizontal resolution and its preliminary validation. *Meteorological Applications*, 27 (1): e1867. doi:10.1002/met.1867.

Marchok, T., Rogers, R. and Tuleya, R. (2007) Validation Schemes for Tropical Cyclone Quantitative Precipitation Forecasts: Evaluation of Operational Models for U.S. Landfalling Cases. *Weather and Forecasting*, 22 (4): 726–746. doi:10.1175/WAF1024.1.

Marks, F.D., Kappler, G. and DeMaria, M. (2002) "Development of a Tropical Cyclone Rainfall Climatology and Persistence (R-CLIPER) Model." In *Preprints*. San Diego, CA, 2002.

Matsueda, M. and Nakazawa, T. (2015) Early warning products for severe weather events derived from operational medium-range ensemble forecasts. *Meteorological Applications*, 22 (2): 213–222. doi:10.1002/met.1444.

Merrill, R.T. (1984) A Comparison of Large and Small Tropical Cyclones. *Monthly Weather Review*, 112 (7): 1408–1418. doi:10.1175/1520-0493(1984)112<1408:ACOLAS>2.0.CO;2.

Meyer, R.J., Baker, J., Broad, K., Czajkowski, J. and Orlove, B. (2014) The Dynamics of Hurricane Risk Perception: Real-Time Evidence from the 2012 Atlantic Hurricane Season. *Bulletin of the American Meteorological Society*, 95 (9): 1389–1404. doi:10.1175/BAMS-D-12-00218.1.

Mitheu, F., Tarnavsky, E., Ficchì, A., Stephens, E., Cornforth, R. and Petty, C. (2023) The utility of impact data in flood forecast verification for anticipatory actions: Case studies from Uganda and Kenya. *Journal of Flood Risk Management*, n/a (n/a): e12911. doi:10.1111/jfr3.12911.

Mittermaier, M., Roberts, N. and Thompson, S.A. (2013) A long-term assessment of precipitation forecast skill using the Fractions Skill Score. *Meteorological Applications*, 20 (2): 176–186. doi:10.1002/met.296.

Molteni, F., Buizza, R., Palmer, T.N. and Petroliagis, T. (1996) The ECMWF Ensemble Prediction System: Methodology and validation. *Quarterly Journal of the Royal Meteorological Society*, 122 (529): 73–119. doi:10.1002/qj.49712252905.

Moon, I.-J., Kim, S.-H. and Chan, J.C.L. (2019) Climate change and tropical cyclone trend. *Nature*, 570 (7759): E3–E5. doi:10.1038/s41586-019-1222-3.

Moore, R.J. (2002) "Aspects of uncertainty, reliability, and risk in flood forecasting systems incorporating weather radar." In Bogardi, J.J. and Kundzewicz, Z.W. (eds.) *Risk, Reliability, Uncertainty, and Robustness of Water Resource Systems*. International Hydrology Series. Cambridge: Cambridge University Press. pp. 30–40. doi:10.1017/CBO9780511546006.004.

Mylne, K., Steele, E., Brown, H., Bunney, C., Gill, P., Saulter, A. and Standen, J. (2021) *Using probabilistic forecasts to effectively enhance operational marine industry decision-making*. EMS2021-237. Copernicus Meetings. doi:10.5194/ems2021-237.

NASA earth observatory (2023) *Freddy Delivers Another Blow*. Available at: https://earthobservatory.nasa.gov/images/151111/freddy-delivers-another-blow (Accessed: 23 March 2023).

Nash, J.E. and Sutcliffe, J.V. (1970) River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10 (3): 282–290. doi:10.1016/0022-1694(70)90255-6.

NCEP and Coauthors (2008) *THORPEX Interactive Grand Global Ensemble (TIGGE) Model Tropical Cyclone Track Data*. doi:doi.org/10.5065/D6GH9GSZ.

Neal, R. (2023) *The use of ECMWF ensembles within a new version of the Met Office's probabilistic weather pattern forecasting tool (Video recording of presentation)*. Available at: https://vimeo.com/833631009/5d00119850 (Accessed: 22 June 2023).

Neal, R., Cox, D., Millard, J., Fenwick, K. and Kelly, J. (2022) *Overview of the new seamless blended multi-model version of Decider*. Met Office project report. Exeter, UK.

Neal, R., Fereday, D., Crocker, R. and Comer, R.E. (2016) A flexible approach to defining weather patterns and their application in weather forecasting over Europe. *Meteorological Applications*, 23 (3): 389–400. doi:10.1002/met.1563.

NHC (2022) *Hurricane Ian Graphics Archive*. Available at: https://www.nhc.noaa.gov/archive/2022/IAN_graphics.php?product=5day_cone_with_line_and_wind (Accessed: 19 June 2023).

NHC (2023a) *Definition of the NHC Track Forecast Cone*. Available at: https://www.nhc.noaa.gov/aboutcone.shtml (Accessed: 22 March 2023).

NHC (2023b) *Hurricane Preparedness - Hazards*. Available at: https://www.nhc.noaa.gov/prepare/hazards.php#rain (Accessed: 13 March 2023).

Nogueira, M. (2020) Inter-comparison of ERA-5, ERA-interim and GPCP rainfall over the last 40 years: Process-based analysis of systematic and random differences. *Journal of Hydrology*, 583: 124632. doi:10.1016/j.jhydrol.2020.124632.

NWS (2023) *NWS JetStream - Tropical Cyclone Hazards*. Available at: https://www.weather.gov/jetstream/tc_hazards (Accessed: 13 March 2023).

Okiria, E., Okazawa, H., Noda, K., Kobayashi, Y., Suzuki, S. and Yamazaki, Y. (2022) A Comparative Evaluation of Lumped and Semi-Distributed Conceptual Hydrological Models: Does Model Complexity Enhance Hydrograph Prediction? *Hydrology*, 9 (5): 89. doi:10.3390/hydrology9050089.

Omranian, E., Sharif, H.O. and Tavakoly, A.A. (2018) How Well Can Global Precipitation Measurement (GPM) Capture Hurricanes? Case Study: Hurricane Harvey. *Remote Sensing*, 10 (7): 1150. doi:10.3390/rs10071150.

Owens, R. and Hewson, T.D. (2018) *ECMWF Forecast User Guide*. Available at: 10.21957/m1cs7h (Accessed: 19 June 2023).

Palmer, T. (2019) The ECMWF ensemble prediction system: Looking back (more than) 25 years and projecting forward 25 years. *Quarterly Journal of the Royal Meteorological Society*, 145 (S1): 12–24. doi:10.1002/qj.3383.

Palmer, T.N. and Tibaldi, S. (1988) On the Prediction of Forecast Skill. *Monthly Weather Review*, 116 (12): 2453–2480. doi:10.1175/1520-0493(1988)116<2453:OTPOFS>2.0.CO;2.

Pappenberger, F., Pagano, T.C., Brown, J.D., Alfieri, L., Lavers, D.A., Berthet, L., Bressand, F., Cloke, H.L., Cranston, M., Danhelka, J., Demargne, J., Demuth, N., de Saint-Aubin, C., Feikema, P.M., Fresch, M.A., Garçon, R., Gelfan, A., He, Y., Hu, Y.-Z., Janet, B., Jurdy, N., Javelle, P., Kuchment, L., Laborda, Y., Langsholt, E., Le Lay, M., Li, Z.J., Mannessiez, F., Marchandise, A., Marty, R., Meißner,

D., Manful, D., Organde, D., Pourret, V., Rademacher, S., Ramos, M.-H., Reinbold, D., Tibaldi, S., Silvano, P., Salamon, P., Shin, D., Sorbet, C., Sprokkereef, E., Thiemig, V., Tuteja, N.K., van Andel, S.J., Verkade, J.S., Vehviläinen, B., Vogelbacher, A., Wetterhall, F., Zappa, M., Van der Zwan, R.E. and Thielen-del Pozo, J. (2016) "Hydrological Ensemble Prediction Systems Around the Globe." In Duan, Q., Pappenberger, F., Thielen, J., Wood, A., Cloke, H.L. and Schaake, J.C. (eds.) *Handbook of Hydrometeorological Ensemble Forecasting*. Berlin, Heidelberg: Springer. pp. 1–35. doi:10.1007/978-3-642-40457-3_47-1.

Park, Y.-Y., Buizza, R. and Leutbecher, M. (2008) TIGGE: Preliminary results on comparing and combining ensembles. *Quarterly Journal of the Royal Meteorological Society*, 134 (637): 2029–2050. doi:10.1002/qj.334.

Pfost, R.L. (2000) Operational tropical cyclone quantitative precipitation forecasting. *National Weather Digest*, 24 (1–2): 61–66.

Prat, O.P. and Nelson, B.R. (2013) Mapping the world's tropical cyclone rainfall contribution over land using the TRMM Multi-satellite Precipitation Analysis. *Water Resources Research*, 49 (11): 7236–7254. doi:10.1002/wrcr.20527.

Prat, O.P. and Nelson, B.R. (2016) On the Link between Tropical Cyclones and Daily Rainfall Extremes Derived from Global Satellite Observations. *Journal of Climate*, 29 (17): 6127–6135. doi:10.1175/JCLI-D-16-0289.1.

Qiu, W., Ren, F., Wu, L., Chen, L. and Ding, C. (2019) Characteristics of tropical cyclone extreme precipitation and its preliminary causes in Southeast China. *Meteorology and Atmospheric Physics*, 131 (3): 613–626. doi:10.1007/s00703-018-0594-5.

Racoma, B.A.B., Holloway, C.E., Schiemann, R.K.H., Feng, X. and Bagtasa, G. (2023) The Effect of the Cordillera Mountain Range on Tropical Cyclone Rainfall in the Northern Philippines. *Atmosphere*, 14 (4): 643. doi:10.3390/atmos14040643.

Racoma, B.A.B., Klingaman, N.P., Holloway, C.E., Schiemann, R.K.H. and Bagtasa, G. (2022) Tropical cyclone characteristics associated with extreme precipitation in the northern Philippines. *International Journal of Climatology*, 42 (6): 3290–3307. doi:10.1002/joc.7416.

Radanovics, S., Vidal, J.-P. and Sauquet, E. (2018) Spatial Verification of Ensemble Precipitation: An Ensemble Version of SAL. *Weather and Forecasting*, 33 (4): 1001–1020. doi:10.1175/WAF-D-17-0162.1.

Ramsay, H. (2017) "The Global Climatology of Tropical Cyclones." In *Oxford Research Encyclopedia of Natural Hazard Science*. doi:10.1093/acrefore/9780199389407.013.79.

Rappaport, E.N. (2000) Loss of Life in the United States Associated with Recent Atlantic Tropical Cyclones. *Bulletin of the American Meteorological Society*, 81 (9): 2065–2074. doi:10.1175/1520-0477(2000)081<2065:LOLITU>2.3.CO;2.

Rappaport, E.N. (2014) Fatalities in the United States from Atlantic Tropical Cyclones: New Data and Interpretation. *Bulletin of the American Meteorological Society*, 95 (3): 341–346. doi:10.1175/BAMS-D-12-00074.1.

Rezapour, M. and Baldock, T.E. (2014) Classification of Hurricane Hazards: The Importance of Rainfall. *Weather and Forecasting*, 29 (6): 1319–1331. doi:10.1175/WAF-D-14-00014.1.

Richardson, D.S. (2000) Skill and relative economic value of the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 126 (563): 649–667. doi:10.1002/qj.49712656313.

Richardson, D.S., Cloke, H.L. and Pappenberger, F. (2020) Evaluation of the Consistency of ECMWF Ensemble Forecasts. *Geophysical Research Letters*, 47 (11): e2020GL087934. doi:10.1029/2020GL087934.

Rios Gaona, M.F., Villarini, G., Zhang, W. and Vecchi, G.A. (2018) The added value of IMERG in characterizing rainfall in tropical cyclones. *Atmospheric Research*, 209: 95–102. doi:10.1016/j.atmosres.2018.03.008.

Ripberger, J., Allan, J., Gupta, K., Bitterman, A., Silva, C., Jenkins-Smith, H. and Cokely, E. (2022a) *There's a Chance of What? Assessing Numeracy Skills of Forecasters, Emergency Managers, and Publics to Improve TC Product Uncertainty Communication, IDSS, and Training*. Available at: https://repository.library.noaa.gov/view/noaa/47959 (Accessed: 28 June 2023).

Ripberger, J., Bell, A., Fox, A., Forney, A., Livingston, W., Gaddie, C., Silva, C. and Jenkins-Smith, H. (2022b) Communicating Probability Information in Weather Forecasts: Findings and Recommendations from a Living Systematic Review of the Research Literature. *Weather, Climate, and Society*, 14 (2): 481–498. doi:10.1175/WCAS-D-21-0034.1.

Robbins, J.C. and Titley, H.A. (2018) Evaluating high-impact precipitation forecasts from the Met Office Global Hazard Map (GHM) using a global impact database. *Meteorological Applications*, 25 (4): 548–560. doi:10.1002/met.1720.

Roberts, N.M. and Lean, H.W. (2008) Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. *Monthly Weather Review*, 136 (1): 78–97. doi:10.1175/2007MWR2123.1.

Rogers, R., Courtney, J. and Wood, K. (2023) The world meteorological organization Tenth International Workshop on tropical cyclones (IWTC-10): A summary. *Tropical Cyclone Research and Review*, 12 (1): 1–9. doi:10.1016/j.tcrr.2023.04.001.

Rogers, R., Marks, F. and Marchok, T. (2009) "Tropical Cyclone Rainfall." In *Encycl. Hydrol. Sci.* doi:10.1002/0470848944.hsa030.

Roy, C. and Kovordányi, R. (2012) Tropical cyclone track forecasting techniques — A review. *Atmospheric Research*, 104–105: 40–69. doi:10.1016/j.atmosres.2011.09.012.

Saha, K.K., Hasan, M.M. and Quazi, A. (2015) Forecasting tropical cyclone-induced rainfall in coastal Australia: implications for effective flood management. *Australasian Journal of Environmental Management*, 22 (4): 446–457. doi:10.1080/14486563.2015.1028109.

Saleh, F., Ramaswamy, V., Wang, Y., Georgas, N., Blumberg, A. and Pullen, J. (2017) A multi-scale ensemble-based framework for forecasting compound coastal-riverine flooding: The Hackensack-Passaic watershed and Newark Bay. *Advances in Water Resources*, 110: 371–386. doi:10.1016/j.advwatres.2017.10.026.

Sampson, C.R., Franklin, J.L., Knaff, J.A. and DeMaria, M. (2008) Experiments with a Simple Tropical Cyclone Intensity Consensus. *Weather and Forecasting*, 23 (2): 304–312. doi:10.1175/2007WAF2007028.1.

Schaake, J.C., Hamill, T.M., Buizza, R. and Clark, M. (2007) HEPEX: The Hydrological Ensemble Prediction Experiment. *Bulletin of the American Meteorological Society*, 88 (10): 1541–1548. doi:10.1175/BAMS-88-10-1541.

Scoccimarro, E., Villarini, G., Gualdi, S., Navarra, A., Vecchi, G., Walsh, K. and Zhao, M. (2017) "Tropical Cyclone Rainfall Changes in a Warmer Climate." In Collins, J.M. and Walsh, K. (eds.) *Hurricanes and Climate Change: Volume 3*. Cham: Springer International Publishing. pp. 243–255. doi:10.1007/978-3-319-47594-3_10.

Sene, K. (2010) "Hydrological Forecasting." In Sene, K. (ed.) *Hydrometeorology: Forecasting and Applications*. Dordrecht: Springer Netherlands. pp. 101–140. doi:10.1007/978-90-481-3403-8_4.

Senkbeil, J., Collins, J. and Reed, J. (2019) Evacuee Perception of Geophysical Hazards for Hurricane Irma. *Weather, Climate, and Society*, 11 (1): 217–227. doi:10.1175/WCAS-D-18-0019.1.

Shen, Y., Du, Y. and Chen, G. (2020) Ensemble Sensitivity Analysis of Heavy Rainfall Associated With Three MCSs Coexisting Over Southern China. *Journal of Geophysical Research: Atmospheres*, 125 (2): e2019JD031266. doi:10.1029/2019JD031266.

Shi, C., Wu, J. and Qi, L. (2017) Analysis on precipitation difference of two typhoons with similar tracks. *J. Marine Meteor*, 37: 36–45.

Simon, A., Penny, A.B., DeMaria, M., Franklin, J.L., Pasch, R.J., Rappaport, E.N. and Zelinsky, D.A. (2018) A Description of the Real-Time HFIP Corrected Consensus Approach (HCCA) for Tropical Cyclone Track and Intensity Guidance. *Weather and Forecasting*, 33 (1): 37–57. doi:10.1175/WAF-D-17-0068.1.

Sivillo, J.K., Ahlquist, J.E. and Toth, Z. (1997) An Ensemble Forecasting Primer. *Weather and Forecasting*, 12 (4): 809–818. doi:10.1175/1520-0434(1997)012<0809:AEFP>2.0.CO;2.

Song, J.Y., Alipour, A., Moftakhari, H.R. and Moradkhani, H. (2020) Toward a more effective hurricane hazard communication. *Environmental Research Letters*, 15 (6): 064012. doi:10.1088/1748-9326/ab875f.

Speight, L., Stephens, E.M., Hawker, L., Baugh, C.A., Neal, J., Cloke, H.L., Grey, S., Titley, H.A., Marsden, K., Simner, A., Ficchì, A., Prudhomme, C., Archer, J., Bazo, J., Dambo, J., Dolan, S., Hunn, A.L., Moschini, F., Savage, J., Smith, J., Towner, J. and Wanzala, M. (2023) Recommendations to improve the interpretation of global flood forecasts to support international humanitarian operations for tropical cyclones. *Journal of Flood Risk Management*, (Submitted).

Stein, R.M., Dueñas-Osorio, L. and Subramanian, D. (2010) Who Evacuates When Hurricanes Approach? The Role of Risk, Information, and Location*. *Social Science Quarterly*, 91 (3): 816–834. doi:10.1111/j.1540-6237.2010.00721.x.

Stephens, E., Day, J.J., Pappenberger, F. and Cloke, H. (2015) Precipitation and floodiness. *Geophysical Research Letters*, 42 (23): 10,316-10,323. doi:10.1002/2015GL066779.

Stewart, S.R. (2021) *National Hurricane Center tropical cyclone report: Hurricane Iota (AL312020)*. NOAA. Available at: https://www.nhc.noaa.gov/data/tcr/AL312020_Iota.pdf (Accessed: 26 August 2022).

Storer, L.N., Gill, P.G. and Williams, P.D. (2020) Multi-diagnostic multi-model ensemble forecasts of aviation turbulence. *Meteorological Applications*, 27 (1): e1885. doi:10.1002/met.1885.

Sturdevant-Rees, P., Smith, J.A., Morrison, J. and Baeck, M.L. (2001) Tropical storms and the flood hydrology of the central Appalachians. *Water Resources Research*, 37 (8): 2143–2168. doi:10.1029/2000WR900310.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S. and Hsu, K.-L. (2018) A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons. *Reviews of Geophysics*, 56 (1): 79–107. doi:10.1002/2017RG000574.

Swinbank, R., Kyouda, M., Buchanan, P., Froude, L., Hamill, T.M., Hewson, T.D., Keller, J.H., Matsueda, M., Methven, J., Pappenberger, F., Scheuerer, M., Titley, H.A., Wilson, L. and Yamaguchi, M. (2016) The TIGGE Project and Its Achievements. *Bulletin of the American Meteorological Society*, 97 (1): 49–67. doi:10.1175/BAMS-D-13-00191.1.

Takemi, T. and Unuma, T. (2020) Environmental Factors for the Development of Heavy Rainfall in the Eastern Part of Japan during Typhoon Hagibis (2019). *Sola*, 16: 30–36. doi:10.2151/sola.2020-006.

Tallapragada, V., Bernardet, L., Biswas, M.K., Gopalakrishnan, S., Kwon, Y., Liu, Q., Marchok, T., Sheinin, D., Tong, M. and Trahan, S. (2014) Hurricane Weather Research and Forecasting (HWRF) model: 2013 scientific documentation. *HWRF Development Testbed Center Tech. Rep*, 99.

Tang, L., Wang, Y., Yu, Z. and Wang, L. (2021) The Asymmetric Precipitation Evolution in Weak Landfalling Tropical Cyclone Rumbia (2018) Over East China. *Frontiers in Earth Science*, 9. doi:10.3389/feart.2021.747271.

Tarek, M., Brissette, F.P. and Arsenault, R. (2020) Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrology and Earth System Sciences*, 24 (5): 2527–2544. doi:10.5194/hess-24-2527-2020.

Taylor, J.W. and Buizza, R. (2003) Using weather ensemble predictions in electricity demand forecasting. *International Journal of Forecasting*, 19 (1): 57–70. doi:10.1016/S0169-2070(01)00123-6.

Titley, H.A. (2022) *WMO/WWRP Project: Tropical Cyclone Probabilistic Forecast Products (TC-PFP) Project (Video recording from IWTC-X day 3 video, Bali, Indonesia)*. Available at: https://www.youtube.com/embed/lCH_CRG0ivs?start=7164&end=8348 (Accessed: 21 June 2023).

Titley, H.A., Bowyer, R.L. and Cloke, H.L. (2020) A global evaluation of multi-model ensemble tropical cyclone track probability forecasts. *Quarterly Journal of the Royal Meteorological Society*, 146 (726): 531–545. doi:10.1002/qj.3712.

Titley, H.A., Cloke, H.L., Harrigan, S., Pappenberger, F., Prudhomme, C., Robbins, J.C., Stephens, E.M. and Zsoter, E. (2021) Key factors influencing the severity of fluvial flood hazard from tropical cyclones. *Journal of Hydrometeorology*, 22 (7): 1801–1817. doi:10.1175/JHM-D-20-0250.1.

Titley, H.A., Yamaguchi, M. and Magnusson, L. (2019) Current and potential use of ensemble forecasts in operational TC forecasting: results from a global forecaster survey. *Tropical Cyclone Research and Review*, 8 (3): 166–180. doi:10.1016/j.tcrr.2019.10.005.

Todini, E. (2007) Hydrological catchment modelling: past, present and future. *Hydrology and Earth System Sciences*, 11 (1): 468–482. doi:10.5194/hess-11-468-2007.

Torn, R.D. and DeMaria, M. (2021) Validation of Ensemble-Based Probabilistic Tropical Cyclone Intensity Change. *Atmosphere*, 12 (3): 373. doi:10.3390/atmos12030373.

Torn, R.D., Elless, T.J., Papin, P.P. and Davis, C.A. (2018) Tropical Cyclone Track Sensitivity in Deformation Steering Flow. *Monthly Weather Review*, 146 (10): 3183–3201. doi:10.1175/MWR-D-18-0153.1.

Toth, Z. and Kalnay, E. (1997) Ensemble Forecasting at NCEP and the Breeding Method. *Monthly Weather Review*, 125 (12): 3297–3319. doi:10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2.

Touma, D., Stevenson, S., Camargo, S.J., Horton, D.E. and Diffenbaugh, N.S. (2019) Variations in the Intensity and Spatial Extent of Tropical Cyclone Precipitation. *Geophysical Research Letters*, 46 (23): 13992–14002. doi:10.1029/2019GL083452.

Towner, J., Cloke, H.L., Zsoter, E., Flamig, Z., Hoch, J.M., Bazo, J., Coughlan de Perez, E. and Stephens, E.M. (2019) Assessing the performance of global hydrological models for capturing peak river flows in the Amazon basin. *Hydrology and Earth System Sciences*, 23 (7): 3057–3080. doi:10.5194/hess-23-3057-2019.

Towner, J., Ficchí, A., Cloke, H.L., Bazo, J., Coughlan de Perez, E. and Stephens, E.M. (2021) Influence of ENSO and tropical Atlantic climate variability on flood characteristics in the Amazon basin. *Hydrology and Earth System Sciences*, 25 (7): 3875–3895. doi:10.5194/hess-25-3875-2021.

Tracton, M.S. and Kalnay, E. (1993) Operational Ensemble Prediction at the National Meteorological Center: Practical Aspects. *Weather and Forecasting*, 8 (3): 379–398. doi:10.1175/1520-0434(1993)008<0379:OEPATN>2.0.CO;2.

University of Bristol (2023) *Resilience and Preparedness to Tropical Cyclones across Southern Africa (REPRESA) project press release*. Available at: https://www.bristol.ac.uk/news/2023/july/represa.html (Accessed: 5 July 2023).

UN-OCHA (2020a) *Central America & Mexico: 2020 Hurricane Season - Situation Report No. 5 (As of 8:00am EST 26 November 2020) - Honduras | ReliefWeb*. Available at: https://reliefweb.int/report/honduras/central-america-mexico-2020-hurricane-season-situation-report-no-5-800am-est-26 (Accessed: 26 August 2022).

UN-OCHA (2020b) *Latin America & The Caribbean: 2020 Hurricane Season Situation Report No. 4 (As of 5:00pm EST 20 November 2020) - Honduras | ReliefWeb*. Available at: https://reliefweb.int/report/honduras/latin-america-caribbean-2020-hurricane-season-situation-report-no-4-500pm-est-20 (Accessed: 26 August 2022).

Van der Grijn, G., Paulsen, J., Lalaurette, F. and Leutbecher, M. (2005) Early medium-range forecasts of tropical cyclones. *ECMWF newsletter*, 102: 7–14.

Van Der Knijff, J.M., Younis, J. and De Roo, A.P.J. (2010) LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation. *International Journal of Geographical Information Science*, 24 (2): 189–212. doi:10.1080/13658810802549154.

Villarini, G. and Denniston, R.F. (2016) Contribution of tropical cyclones to extreme rainfall in Australia. *International Journal of Climatology*, 36 (2): 1019–1025. doi:10.1002/joc.4393.

Villarini, G., Goska, R., Smith, J.A. and Vecchi, G.A. (2014a) North Atlantic Tropical Cyclones and U.S. Flooding. *Bulletin of the American Meteorological Society*, 95 (9): 1381–1388. doi:10.1175/BAMS-D-13-00060.1.

Villarini, G., Lavers, D.A., Scoccimarro, E., Zhao, M., Wehner, M.F., Vecchi, G.A., Knutson, T.R. and Reed, K.A. (2014b) Sensitivity of Tropical Cyclone Rainfall to Idealized Global-Scale Forcings. *Journal of Climate*, 27 (12): 4622–4641. doi:10.1175/JCLI-D-13-00780.1.

Villarini, G., Smith, J.A., Baeck, M.L., Marchok, T. and Vecchi, G.A. (2011) Characterization of rainfall distribution and flooding associated with U.S. landfalling tropical cyclones: Analyses of Hurricanes Frances, Ivan, and Jeanne (2004). *Journal of Geophysical Research: Atmospheres*, 116 (D23). doi:10.1029/2011JD016175.

Vitart, F. and Stockdale, T.N. (2001) Seasonal Forecasting of Tropical Storms Using Coupled GCM Integrations. *Monthly Weather Review*, 129 (10): 2521–2537. doi:10.1175/1520-0493(2001)129<2521:SFOTSU>2.0.CO;2.

Wang, C.-C., Lin, B.-X., Chen, C.-T. and Lo, S.-H. (2015) Quantifying the Effects of Long-Term Climate Change on Tropical Cyclone Rainfall Using a Cloud-Resolving Model: Examples of Two Landfall Typhoons in Taiwan. *Journal of Climate*, 28 (1): 66–85. doi:10.1175/JCLI-D-14-00044.1.

Wang, W., Zhang, Z., Cangialosi, J.P., Brennan, M., Cowan, L., Clegg, P., Takuya, H., Masaaki, I., Das, A.K., Mohapatra, M., Sharma, M., Knaff, J.A., Kaplan, J., Birchard, T., Doyle, J.D., Heming, J., Moskaitis, J., Komaromi, W.A., Ma, S., Sampson, C., Wu, L. and Blake, E. (2023) A review of recent advances (2018–2021) on tropical cyclone intensity change from operational perspectives, part 2: Forecasts by operational centers. *Tropical Cyclone Research and Review*, 12 (1): 50–63. doi:10.1016/j.tcrr.2023.05.003.

Wang, Z., Zhao, J., Huang, H. and Wang, X. (2022) A Review on the Application of Machine Learning Methods in Tropical Cyclone Forecasting. *Frontiers in Earth Science*, 10. doi:10.3389/feart.2022.902596.

Weber, H.C., Lok, C.C.F., Davidson, N.E. and Xiao, Y. (2014) Objective Estimation of the Radius of the Outermost Closed Isobar in Tropical Cyclones. *Tropical Cyclone Research and Review*, 3 (1): 1–21. doi:10.6057/2014TCRR01.01.

Wernli, H., Hofmann, C. and Zimmer, M. (2009) Spatial Forecast Verification Methods Intercomparison Project: Application of the SAL Technique. *Weather and Forecasting*, 24 (6): 1472–1484. doi:10.1175/2009WAF2222271.1.

Wernli, H., Paulat, M., Hagen, M. and Frei, C. (2008) SAL—A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts. *Monthly Weather Review*, 136 (11): 4470–4487. doi:10.1175/2008MWR2415.1.

Whitehead, J.C., Edwards, B., Van Willigen, M., Maiolo, J.R., Wilson, K. and Smith, K.T. (2000) Heading for higher ground: factors affecting real and hypothetical hurricane evacuation behavior. *Global Environmental Change Part B: Environmental Hazards*, 2 (4): 133–142. doi:10.1016/S1464-2867(01)00013-4.

Wilks, D.S. (2011) *Statistical methods in the atmospheric sciences*. Academic Press. Available at: https://scholar.google.com/scholar_lookup?hl=en&publication_year=2011&pages=%00empty%00&author=D.S.+Wilks&isbn=%00null%00&title=Statistical+Methods+in+the+Atmospheric+Sciences (Downloaded: 25 April 2023).

WMO (2011) *Manual on flood forecasting and warning*. WMO 1072. Geneva: WMO. Available at: https://library.wmo.int/index.php?lvl=notice_display&id=5841.

WMO (2018) *Ninth International Workshop on Tropical Cyclones (IWTC-9)*. Honolulu, Hawaii, USA: WMO. Available at: https://library.wmo.int/index.php?lvl=notice_display&id=22038#.ZEE5aHbMJD8 (Accessed: 20 April 2023).

WMO (2020) *Tropical Cyclone-Probabilistic Forecast Products (TC-PFP), WMO website*. Available at: https://community.wmo.int/en/wwrp-tropical-cyclone-probabilistic-forecast-products (Accessed: 21 June 2023).

WMO (2022) *Tenth International Workshop on Tropical Cyclones (IWTC-10) | World Meteorological Organization*. Available at: https://community.wmo.int/en/meetings/tenth-international-workshop-tropical-cyclones-iwtc-10 (Accessed: 23 June 2023).

WMO (2023) *Advisory Group on Tropical Cyclones (AG-TC)*. Available at: https://community.wmo.int/en/advisory-group-tropical-cyclones-ag-tc (Accessed: 27 June 2023).

Wright, D.B., Knutson, T.R. and Smith, J.A. (2015) Regional climate model projections of rainfall from U.S. landfalling tropical cyclones. *Climate Dynamics*, 45 (11): 3365–3379. doi:10.1007/s00382-015-2544-y.

Wu, C.-C., Chen, S.-G., Lin, S.-C., Yen, T.-H. and Chen, T.-C. (2013) Uncertainty and Predictability of Tropical Cyclone Rainfall Based on Ensemble Simulations of Typhoon Sinlaku (2008). *Monthly Weather Review*, 141 (10): 3517–3538. doi:10.1175/MWR-D-12-00282.1.

Wu, W., Emerton, R., Duan, Q., Wood, A.W., Wetterhall, F. and Robertson, D.E. (2020) Ensemble flood forecasting: Current status and future opportunities. *WIREs Water*, 7 (3): e1432. doi:10.1002/wat2.1432.

Yamaguchi, M., Chan, J.C.L., Moon, I.-J., Yoshida, K. and Mizuta, R. (2020) Global warming changes tropical cyclone translation speed. *Nature Communications*, 11 (1): 47. doi:10.1038/s41467-019-13902-y.

Yamaguchi, M., Ishida, J., Sato, H. and Nakagawa, M. (2017) WGNE Intercomparison of Tropical Cyclone Forecasts by Operational NWP Models: A Quarter Century and Beyond. *Bulletin of the American Meteorological Society*, 98 (11): 2337–2349. doi:10.1175/BAMS-D-16-0133.1.

Yamaguchi, M. and Koide, N. (2017) Tropical Cyclone Genesis Guidance Using the Early Stage Dvorak Analysis and Global Ensembles. *Weather and Forecasting*, 32 (6): 2133–2141. doi:10.1175/WAF-D-17-0056.1.

Yamaguchi, M., Nakazawa, T. and Hoshino, S. (2012) On the relative benefits of a multi-centre grand ensemble for tropical cyclone track prediction in the western North Pacific. *Quarterly Journal of the Royal Meteorological Society*, 138 (669): 2019–2029. doi:10.1002/qj.1937.

Yamaguchi, M., Nakazawa, T. and Hoshino, S. (2014) North Western Pacific Tropical Cyclone Ensemble Forecast Project. *Tropical Cyclone Research and Review*, 3 (3): 193–201. doi:10.6057/2014TCRR03.05.

Yamaguchi, M., Sakai, R., Kyoda, M., Komori, T. and Kadowaki, T. (2009) Typhoon Ensemble Prediction System Developed at the Japan Meteorological Agency. *Monthly Weather Review*, 137 (8): 2592–2604. doi:10.1175/2009MWR2697.1.

Yamaguchi, M., Vitart, F., Lang, S.T.K., Magnusson, L., Elsberry, R.L., Elliott, G., Kyouda, M. and Nakazawa, T. (2015) Global Distribution of the Skill of Tropical Cyclone Activity Forecasts on Short- to Medium-Range Time Scales. *Weather and Forecasting*, 30 (6): 1695–1709. doi:10.1175/WAF-D-14-00136.1.

Yang, L., Villarini, G., Zeng, Z., Smith, J., Liu, M., Li, X., Wang, L. and Hou, A. (2020) Riverine Flooding and Landfalling Tropical Cyclones Over China. *Earth's Future*, 8 (3): e2019EF001451. doi:10.1029/2019EF001451.

Young, M.V. and Grahame, N.S. (2023) The history of UK weather forecasting: the changing role of the central guidance forecaster. Part 2: the birth of operational numerical weather prediction. *Weather*, 78 (1): 4–9. doi:10.1002/wea.4216.

Yu, H., Chen, G., Zhou, C., Wong, W.K., Yang, M., Xu, Y., Chen, P., Wan, R. and Hu, X. (2022a) Are We Reaching the Limit of Tropical Cyclone Track Predictability in the Western North Pacific? *Bulletin of the American Meteorological Society*, 103 (2): E410–E428. doi:10.1175/BAMS-D-20-0308.1.

Yu, J., Liu, J., Baek, J.-W., Fong, C. and Fu, M. (2022b) Impact-based forecasting for improving the capacity of typhoon-related disaster risk reduction in typhoon committee region. *Tropical Cyclone Research and Review*, 11 (3): 163–173. doi:10.1016/j.tcrr.2022.09.003.

Yu, W., Nakakita, E., Kim, S. and Yamaguchi, K. (2016) Impact Assessment of Uncertainty Propagation of Ensemble NWP Rainfall to Flood Forecasting with Catchment Scale. *Advances in Meteorology*, 2016: e1384302. doi:10.1155/2016/1384302.

Yu, Z., Chen, Y.J., Ebert, B., Davidson, N.E., Xiao, Y., Yu, H. and Duan, Y. (2020) Benchmark rainfall verification of landfall tropical cyclone forecasts by operational ACCESS-TC over China. *Meteorological Applications*, 27 (1): e1842. doi:10.1002/met.1842.

Yu, Z., Wang, Y. and Xu, H. (2015) Observed Rainfall Asymmetry in Tropical Cyclones Making Landfall over China. *Journal of Applied Meteorology and Climatology*, 54 (1): 117–136.

Yu, Z., Wang, Y., Xu, H., Davidson, N., Chen, Y., Chen, Y. and Yu, H. (2017) On the Relationship between Intensity and Rainfall Distribution in Tropical Cyclones Making Landfall over China. *Journal of Applied Meteorology and Climatology*, 56 (10): 2883–2901. doi:10.1175/JAMC-D-16-0334.1.

Yu, Z., Wang, Y., Yu, H. and Duan, Y. (2022c) The Relationship Between the Inner-Core Size and the Rainfall Distribution in Landfalling Tropical Cyclones Over China. *Geophysical Research Letters*, 49 (8): e2021GL097576. doi:10.1029/2021GL097576.

Yuan, S., Zhu, L. and Quiring, S.M. (2021) Comparison of Two Multisatellite Algorithms for Estimation of Tropical Cyclone Precipitation in the United States and Mexico: TMPA and IMERG. *Journal of Hydrometeorology*, 22 (4): 923–939. doi:10.1175/JHM-D-19-0296.1.

Zappa, M., Jaun, S., Germann, U., Walser, A. and Fundel, F. (2011) Superposition of three sources of uncertainties in operational flood forecasting chains. *Atmospheric Research*, 100 (2): 246–262. doi:10.1016/j.atmosres.2010.12.005.

Zhang, M. and Meng, Z. (2018) Impact of Synoptic-Scale Factors on Rainfall Forecast in Different Stages of a Persistent Heavy Rainfall Event in South China. *Journal of Geophysical Research: Atmospheres*, 123 (7): 3574–3593. doi:10.1002/2017JD028155.

Zhang, W., Villarini, G., Vecchi, G.A. and Smith, J.A. (2018) Urbanization exacerbated the rainfall and flooding caused by hurricane Harvey in Houston. *Nature*, 563 (7731): 384–388. doi:10.1038/s41586-018-0676-z.

Zhang, X. and Yu, H. (2017) A Probabilistic Tropical Cyclone Track Forecast Scheme Based on the Selective Consensus of Ensemble Prediction Systems. *Weather and Forecasting*, 32 (6): 2143–2157. doi:10.1175/WAF-D-17-0071.1.

Zhou, F. and Toth, Z. (2020) On the Prospects for Improved Tropical Cyclone Track Forecasts. *Bulletin of the American Meteorological Society*, 101 (12): E2058–E2077. doi:10.1175/BAMS-D-19-0166.1.

Zhou, X., Zhu, Y., Hou, D., Fu, B., Li, W., Guan, H., Sinsky, E., Kolczynski, W., Xue, X., Luo, Y., Peng, J., Yang, B., Tallapragada, V. and Pegion, P. (2022a) The Development of the NCEP Global Ensemble Forecast System Version 12. *Weather and Forecasting*, 37 (6): 1069–1084. doi:10.1175/WAF-D-21-0112.1.

Zhou, X., Zhu, Y., Hou, D., Fu, B., Li, W., Guan, H., Sinsky, E., Kolczynski, W., Xue, X., Luo, Y., Peng, J., Yang, B., Tallapragada, V. and Pegion, P. (2022b) The Development of the NCEP Global Ensemble Forecast System Version 12. *Weather and Forecasting*, 37 (6): 1069–1084. doi:10.1175/WAF-D-21-0112.1.

Zhou, Y., Matyas, C., Li, H. and Tang, J. (2018) Conditions associated with rain field size for tropical cyclones landfalling over the Eastern United States. *Atmospheric Research*, 214: 375–385. doi:10.1016/j.atmosres.2018.08.019.

Zhou, Y. and Matyas, C.J. (2017) Spatial characteristics of storm-total rainfall swaths associated with tropical cyclones over the Eastern United States. *International Journal of Climatology*, 37 (S1): 557–569. doi:10.1002/joc.5021.

Zsoter, E., Harrigan, S., Barnard, C., Wetterhall, F., Salamon, P. and Prudhomme, C. (2019) *River discharge and related forecasted data from the Global Flood Awareness System. v2.1.* Available at: 10.24381/cds.ff1aef77 (Accessed: 23 March 2023).

Zsoter, E., Prudhomme, C., Harrigan, S., de Rosnay, P., Munõz-Sabater, J. and Stephens, E. (2020) *Trends in the GloFAS-ERA5 river discharge reanalysis*. number 871. Available at: https://www.ecmwf.int/en/elibrary/81194-trends-glofas-era5-river-discharge-reanalysis (Accessed: 23 March 2023).

Appendices

These appendices contain the typeset versions of each of the published chapters presented in this thesis, alongside additional important publications during this PhD. All author contribution statements (provided in the respective chapters or in the summary boxes for the appendices) have been approved by Professor Hannah Cloke, supervisor.

Hannah L. Cloke

A1: Published Article: Current and potential use of ensemble forecasts in operational TC forecasting: results from a global forecaster survey

This appendix contains the published version of chapter 3 of this thesis, with the following reference:

Titley, H.A., Yamaguchi, M. and Magnusson, L. (2019) Current and potential use of ensemble forecasts in operational TC forecasting: results from a global forecaster survey. *Tropical Cyclone Research and Review*, 8 (3): 166–180. doi:10.1016/j.tcrr.2019.10.005.

CURRENT AND POTENTIAL USE OF ENSEMBLE FORECASTS IN OPERATIONAL TC FORECASTING: RESULTS FROM A GLOBAL FORECASTER SURVEY

H. A. TITLEY[a,b], M. YAMAGUCHI[c], AND L. MAGNUSSON[d]

[a]*Met Office, Exeter, U.K.*
[b]*University of Reading, Reading, U.K.*
[c]*Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan*
[d]*European Centre for Medium-Range Weather Forecasts, Reading, U.K.*

ABSTRACT

In order to understand the current and potential use of ensemble forecasts in operational tropical cyclone (TC) forecasting, a questionnaire on the use of dynamic ensembles was conducted at operational TC forecast centers across the world, in association with the World Meteorological Organisation (WMO) High-Impact Weather Project (HIWeather). The results of the survey are presented, and show that ensemble forecasts are used by nearly all respondents, particularly in TC track and genesis forecasting, with several examples of where ensemble forecasts have been pulled through successfully into the operational TC forecasting process. There is still, however, a notable difference between the high proportion of operational TC forecasters who use and value ensemble forecast information, and the slower pull-through into operational forecast warnings and products of the probabilistic guidance and uncertainty information that ensembles can provide. Those areas of research and development that would help TC forecasters to make increased use of ensemble forecast information in the future include improved access to ensemble forecast data, verification and visualizations, the development of hazard and impact-based products, an improvement in the skill of the ensembles (particularly for intensity and structure), and improved guidance on how to use ensembles and optimally combine forecasts from all available models. A change in operational working practices towards using probabilistic information, and providing and communicating dynamic uncertainty information in operational forecasts and warnings, is also recommended.

Keywords: tropical cyclones, operational forecaster survey, ensemble forecasts, probabilistic forecasts, uncertainty

1. Introduction and aims of the questionnaire

For over 25 years, dynamical ensemble model forecasts have been routinely produced by several global numerical weather prediction (NWP) modelling centers (Palmer, 2018). Instead of making a single forecast of the most likely weather, an ensemble of forecasts is produced, taking into account uncertainty in the initial conditions, and imperfections in the model formulation, and aiming to give an indication of the range of possible future states of the atmosphere. Buizza (2008) described the two main advantages of ensemble forecasts over deterministic forecasts: "first, the fact that an ensemble system predicts not only the most likely scenario but also the probability of occurrence of any event, and second the fact that an ensemble system can provide more consistent (i.e. less changeable) successive forecasts."

Dynamic ensemble forecasting of tropical cyclones has an important role to play in capturing the situation-dependent uncertainty in the forecasts for existing storms (e.g., Dupont et al., 2011; Yamaguchi et al., 2012; Du et al., 2016; Leonardo and Colle, 2017; Zhang and Yu, 2018), and in providing probabilistic information about tropical cyclone genesis (e.g., Yamaguchi et al., 2015; Yamaguchi and Koide, 2017). Many global NWP centers produce tropical cyclone track forecasts from their ensemble forecast models, develop tropical cyclone forecast products, verify the forecasts, and share these via The International Grand Global Ensemble (TIGGE) cyclone exchange program (Swinbank et al., 2016). There is, however, great potential to increase the use of this probabilistic ensemble forecast information in operational TC forecasts and warnings.

The World Meteorological Organisation (WMO)'s 10-

Corresponding author: H. A. Titley, helen.titley@metoffice.gov.uk

DOI: 10.6057/2019TCRR03.05

year high impact weather research project, HIWeather, aims to enable increased global resilience to severe weather by improving forecasts of severe weather and its impacts, and the communication of information to users, especially emergency managers (WMO, 2014). As a project within HIWeather, and in association with the World Weather Research Programme (WWRP) Predictability, Dynamics and Ensemble Forecasting (PDEF) working group, a group of international researchers started a collaboration aimed at enhancing collaboration amongst the research and operational community on the topic of ensemble forecasting and verification of TCs. The first activity was to design and circulate a questionnaire to all operational TC forecasting centers around the world asking about their use of dynamical ensemble forecast information.

While it is acknowledged that the use of multiple deterministic forecasts in consensus forecasting is currently a valuable cornerstone for track and intensity forecasting (e.g., Goerss, 2000, 2007; Sampson et al., 2008; Goerss and Sampson, 2014, Yamaguchi et al., 2017), in order to create a manageable and clear scope for the questionnaire, neither consensus, statistical, nor analog-based ensembles, were included in the definition of ensembles. The questionnaire defined "ensemble" to be a dynamically-based ensemble run with multiple members, using a combination of initial condition and model physics perturbations to represent the initial condition and model uncertainty, in order to focus on the current and potential use of dynamic ensembles in operational tropical cyclone forecasting.

The questionnaire aimed to report a baseline on the current use of ensembles at global operational TC forecast centers to the 9th International Workshop on Tropical Cyclones (IWTC-9), and help shape future research and development in order to maximise the pull through of the benefits of ensemble forecasts into operational tropical cyclone forecasts and warnings. The survey took the form of a Google Form, and was sent to key contacts at operational centers worldwide using an email distribution list provided by WMO, including Regional Specialised Met Centers (RSMCs), Tropical Cyclone Warning Centers (TCWCs), and the Typhoon Committee Secretariat, for onward circulation to their members. The forecasting centers were requested to circulate the survey to all forecasters at their center, and several follow-up reminders were also sent to ensure as wide a participation as possible. The more detailed objectives of the survey were as follows:

a) To document the current availability of ensemble forecasts in tropical cyclone forecast centers, and their use by operational forecasters.

b) To ascertain how uncertainty is represented in the operational warnings for track, intensity, genesis and hazards from their center, and whether or not this uncertainty information is taken from dynamical ensemble forecasts.

c) To obtain examples where probabilistic forecasts have been successfully integrated into operations, but also occasions where hurdles have prevented them from being fully utilised.

d) To collate forecaster feedback on where they would like to see future research and development focussed to enable them to make wider use of ensemble forecasts.

The list of questions in the survey is contained in the Appendix. The results of the questionnaire were presented at IWTC-9 in Hawaii in December 2018, as the main component of sub topic 6.3 titled "Current and Potential Use of Ensemble Forecasts in Operational TC Forecasting" (Yamaguchi, Titley and Magnusson, 2018). This paper contains a more detailed write-up of the results of the questionnaire. Section 2 contains the results of the questionnaire, organised into sub-sections according to the four main objectives as stated above; and Section 3 gives conclusions and recommendations going forward.

2. Results of the questionnaire

2.1 Current availability and use of ensemble forecasts

It was important to ensure that the questionnaire results took into account the views and experiences of forecasters in all areas of the world, and from a variety of types of forecast centers. A total of 60 questionnaire responses were received from 25 different centers around the world, including RSMCs, TCWCs, national meteorological services, and the Joint Typhoon Warning Center (see Figure 1). Please note that no question was compulsory so there are cases where the total number of respondents does not add up to 60. Figure 2 shows the tropical cyclone basins that were selected to be of interest to the respondents, and indicates that there is a good balance globally, with at least 14 respondents having forecast interest in each of the global basins (with many respondents having forecast interest in more than one basin). The majority of respondents were operational forecasters, but there were also inputs from managers, researchers, and others such as technical developers.

Dynamic ensemble forecasts were used by nearly all respondents (95%). This is maybe not a surprise, as although all forecasters were encouraged to fill in the questionnaire regardless of their use of ensembles, it is to be expected that there would be a higher degree of motivation to complete the survey if they were interested in ensemble forecasting. Nonetheless it shows that a large proportion of operational TC forecasters use ensemble forecasts. Those who didn't were from National Met Services in developing countries, where ensemble forecasts are not readily accessible. Figure 3 shows the breakdown of how important respondents felt the use of ensemble forecasts was in each of the main areas of tropical cyclone forecasting. Ensemble forecasts are seen to be particularly important in track and genesis forecasting. For hazard forecasting and impact-based forecasting, although a handful of respondents did not use ensemble forecasts, the majority did, but the importance of ensemble forecasts was less than in track and genesis forecasts. In

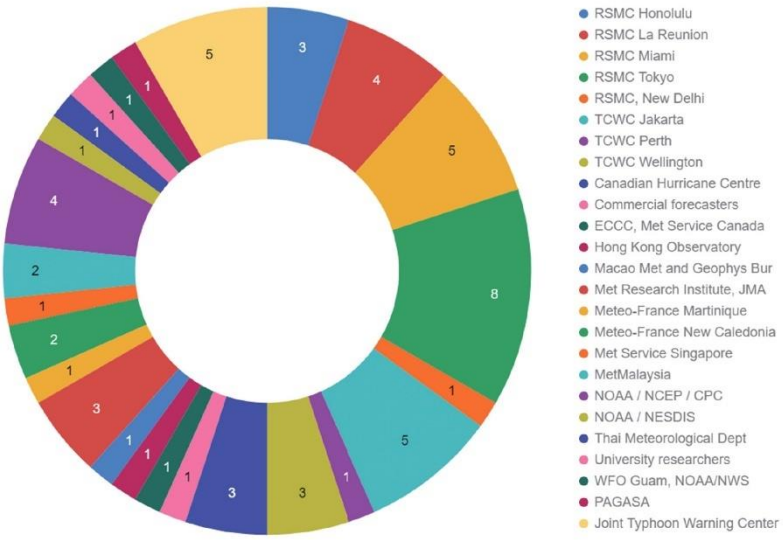


FIG. 1. Organisation of respondents to the questionnaire (including RSMC/TCWC where applicable).

intensity forecasting, the categories were very evenly split, with a significant proportion of respondents not using ensemble forecasts. The areas where ensemble forecasts were used least were in forecasting tropical cyclone size/structure, and in extra-tropical transition forecasting (although this result may be influenced by not all of the forecasters being involved in predicting this process).

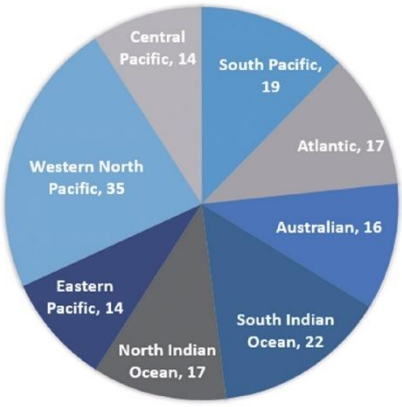


FIG. 2. The tropical cyclone basins that were of interest to the respondents.

The ensemble forecast information was accessed in a variety of different ways. Over half of respondents were able to use ensemble-based products created at their center, but even larger numbers looked at online products from other centers or products sent by global NWP centers, either in combination with those produced in-center, or, in many cases, being reliant on these external sources of ensemble forecast information. Ensemble data was also accessed from the WMO Global Telecommunications System (GTS) and from the TIGGE archive. The most frequently used ensemble models worldwide are the European Centre for Medium-Range Weather Forecasts Ensemble (ECMWF ENS) and the National Centers for Environmental Prediction Global Ensemble Forecast System (NCEP GEFS) ensembles, followed by the UK Met Office Global and Regional Ensemble Prediction System (MOGREPS-G), Japan Meteorological Agency's global ensemble and the Canadian global ensemble (see Figure 4). Again these models are used primarily for their track and genesis forecasts, although a significant number do also use NCEP GEFS and ECMWF ENS for intensity and hazard forecasting. A range of other ensembles are used by a smaller number of respondents.

Over 90% of respondents used multi-model ensemble

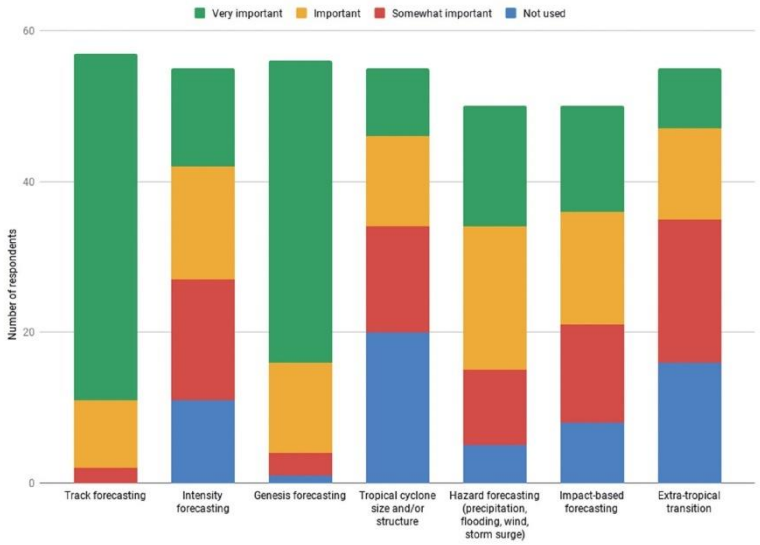


FIG. 3. The questionnaire asked "How important would you say ensemble forecasts are in your forecasting of tropical cyclones?", in each of the above categories.

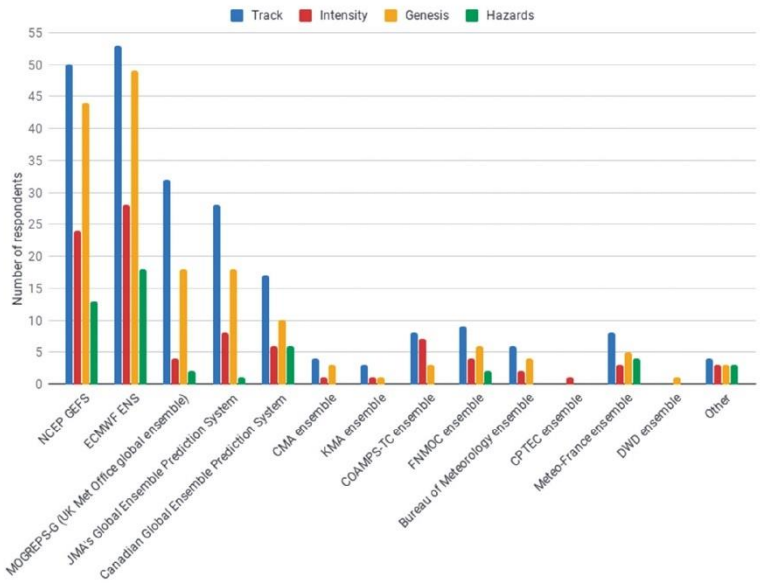


FIG. 4. "For each of the main areas of TC forecasting, which ensemble forecast models do you use?"

forecasts (multiple dynamical ensembles), with around half using a full probabilistic multi-model ensemble combination and half mainly by comparing one ensemble with another. Calibrated ensemble forecasts were less-well used, with 40% of respondents using them for track forecasting, and between 20 and 25% for intensity and genesis forecasting. Figure 5 shows that the product type used most regularly was the ensemble mean track, followed by the tracks from the ensemble members (plumes), which were used more regularly than the strike probabilities created from them. However, the probabilistic information in the genesis products is well-used. It was noticeable that many respondents were keen to have future access to additional ensemble-based products such as landfall probability, storm surge, river discharge, flood risk and impact-based products.

When asked if there was a particular product or ensemble that they would like to be able to see the answers were very varied, including i) an operational ensemble-based cone of probability, ii) more useful intensity guidance including probability of intensity change and bias corrected intensity forecasts from multiple ensembles, iii) hazard forecasts e.g. high-impact winds and precipitation, storm surge, river discharge, flash flood risk/inundation, iv) impact-based products, v) box-whisker plots of TC tracks, vi) ensemble clustering with probabilities.

Figures 6 and 7 show the differences in how forecasters use ensemble forecasts in two areas of tropical cyclone forecasting: tracks and intensity. Only 4 respondents said that they rarely used ensemble forecasts for track forecasting, but this rose to half of respondents for intensity. However, well over half of people for track, and a third for intensity, said that they mainly used ensemble forecasts to compare the ensemble mean forecasts to the deterministic forecasts. This was more than those who used the full probabilistic forecast information provided by the ensemble, indicating that at present the ensemble information may be being used in a deterministic-focussed environment, where it is harder to make full use of the probabilistic forecast information provided by the ensemble forecasts. For track, three quarters of respondents found the spread useful to provide situation-dependent uncertainty, falling to one third for intensity. Hardly any respondents ticked that they "don't find the ensemble forecasts very helpful when the spread is very large", indicating that they are finding cases with a large spread to be correctly highlighting the increased uncertainty in that forecast.

2.2 *How is uncertainty represented in operational warnings?*

Tables 1-4 detail how forecast uncertainty is represented in operational forecasts and warnings issued from each separate center (where this information was available from respondents). Table 1 shows the responses from each center for their operational track forecasts. There is no standard forecast length or frequency, with forecast lengths of 3, 5 or 7 days, and frequencies of 3, 6, 12 or 24 hours. Uncertainty is represented in all forecasts, but this can be via a circle, a

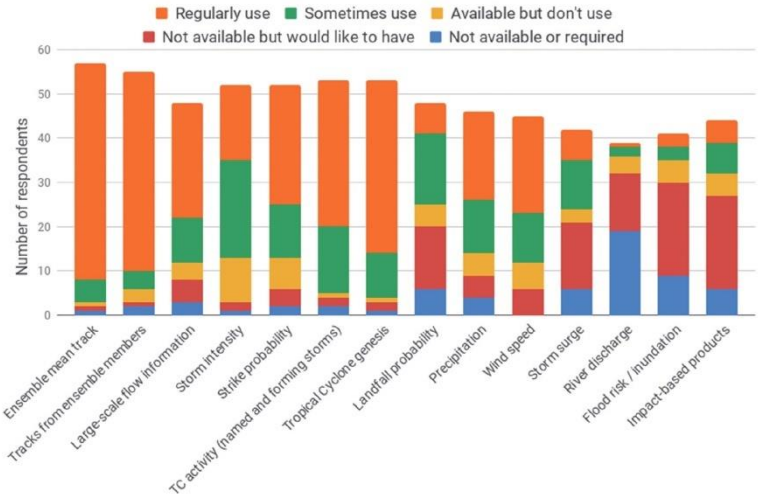


FIG. 5. "Which ensemble-based or probabilistic products do you use?"

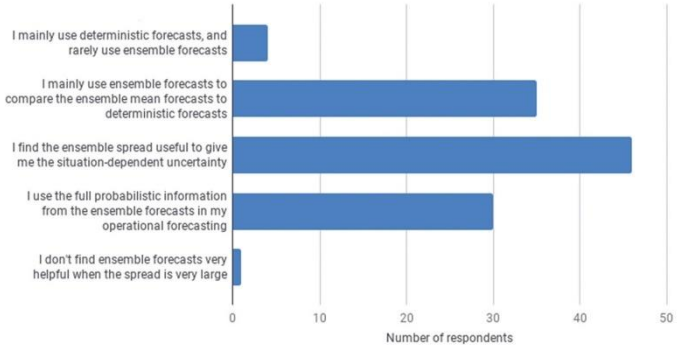


FIG. 6. "Which of the following statements describe your use of ensemble forecast information in track forecasting?"

cone, or an error swath. The size of each circle/cone/error swath is generally set so that a storm would be expected to fall within it in for a given percentage of cases, but this percentage varies too from center to center, from 66% to 80%. Out of the RSMCs, only two include ensemble spread in this calculation (RSMC La Réunion and Tokyo). More commonly used are historical forecast error statistics. Out of the other warning centers only TCWC Wellington and Météo France New Caledonia include ensemble spread when calculating the uncertainty in their operational forecasts. Thus there appears to be a clear discrepancy between the number of operational TC forecasters who use and value ensemble forecast information, as shown in the previous section, and the pull-through into operational forecast products of the probabilistic guidance that ensembles can provide.

Table 2 shows the responses from each center about their operational intensity forecasts. Forecast lengths and frequencies again vary between 3, 5, and 7 days, at 3, 6, 12 or 24 hourly frequency. Only RSMC Honolulu and RSMC Miami said that uncertainty was included, with this being in the form of wind speed probabilities from a statistical model in both cases. Dynamical ensemble forecasts are however often included when calculating the uncertainty in operational TC genesis forecasts (Table 3), often in combination with consensus forecasting, climatological information, and forecaster judgement. Forecast lengths vary between 1 and 7 days. RSMC Tokyo issue with the highest frequency (3 hours), with many forecast centers issuing their genesis forecasts once per day. The most used way of representing uncertainty is to place forecasts into probability categories such as high/medium/low.

Table 4 shows that the expression of uncertainty and the use of ensembles in operational hazard forecasts varies considerably, with products often taking the form of named or color-coded categories assigned based on probabilities

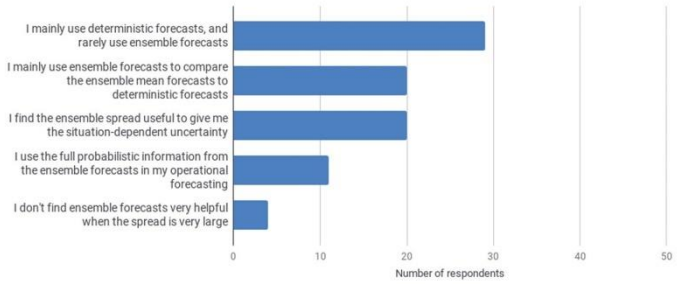


FIG. 7. "Which of the following statements describe your use of ensemble forecast information in intensity forecasting?"

TABLE 1. Operational track forecasts from each tropical cyclone forecasting center.

Center name	Forecast lengths / frequency	How is uncertainty represented (and what proportion of cases are expected to stay within this)?	How is it calculated?
RSMC Tokyo	5 days / 8 times a day	Circle (70%)	Historical forecast error statistics (up to 72 hours) and ensemble spread (after 96 hours, Yamaguchi et al. 2009)
RSMC Honolulu	5 days / 6 hourly	Cone (67%)	Historical forecast error (5 year)
RSMC La Reunion	5 days / 6 hourly	Cone (75%)	Based on both ensemble spread and historical forecast error, (Dupont et al. 2011)
RSMC Miami	5 days publicly, 7 days internally / 6 hourly (more frequent when necessary)	Cone (67%)	Previous 5 years of track error for that basin
RSMC New Delhi	5 days / 6 hourly (and 3 hourly updates for intervening)	Cone (72%)	Historical forecast error statistics
TCWC Jakarta	3 days / 6 or 12 hourly	Cone (80%)	Consensus spread. Patterns obtained from wind model forecasts, and sometimes the form is modified manually to make the shape smoother.
TCWC Perth	7 days / 1, 3, or 6 hourly	Cone	Situation dependent, usually either consensus spread or climatological uncertainty but can be manually manipulated
TCWC Wellington	24 hours, but up to 5 days if a threat to NZ / 6 hourly	Cone (70%)	Calculated based on consensus spread, ensemble spread and climatological uncertainty
Joint Typhoon Warning Center (JWTC)	5 days / 6 hourly	Error swath	Calculated by adding the JTWC 5-year running mean forecast track error to the forecast 34-knot wind radii at each forecast time
Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA)	5 days / 6 hourly	Cone / circle (70%)	Historical forecast error statistics
Weather Forecast Office (WFO) Guam	5 days / 6 hourly	Error swath (66%)	Standard deviation of track error
Thai Meteorological Department	3 days / 3, 6, 12 hourly dependent on situation	Cone	Historical forecast error statistics
Malaysian Meteorological Department (MetMalaysia)	7 days / 3 hourly if TC with in our region	Circle (80%)	Not given
Météo-France New Caledonia	3 days / 6 hourly	Cone (75%)	Choice between climatological uncertainty and ensemble-based uncertainty (ECMWF). We mostly use climatological uncertainty for operational reasons (faster).
Météo-France Martinique	5 days / 6 hourly	Cone (66%)	Climatological uncertainty
Hong Kong Observatory	5 days / 24 hourly	Cone/Circle (70%)	Historical forecast error statistics
Canadian Hurricane Center	5 days / 6 hourly	Cone (70%)	Historical forecast error

(statistical or dynamical).

2.3 *Integration of probabilistic forecasts into operations*

Respondents were asked for examples where probabilistic forecasts have been successfully integrated into operations, and also for occasions where hurdles have prevented them from being fully utilized. Numerous excellent examples were given of where probabilistic forecasts have been pulled through into operational forecasts and warnings, including the ensemble-based RSMC La Réunion operational track uncertainty cone (Dupont et al., 2011), and the multi-model ensemble TC track and activity forecast products made available from RSMC Tokyo to Typhoon Committee members based on the outcomes of the North Western Pacific Tropical Cyclone Ensemble Forecast Project

TABLE 2. Operational intensity forecasts from each tropical cyclone forecasting center.

Center name	Forecast lengths / frequency	Is uncertainty represented, and if so how?
RSMC Tokyo	3 days / 8 times a day (extended to 5 days in March 2019)	No
RSMC Honolulu	5 days / 6 hourly	Yes, wind speed probabilities for 34/50/65 knots – Monte Carlo simulation
RSMC La Reunion	5 days / 6 hourly	No
RSMC Miami	5 to 7 days / 6 hourly	Yes, wind speed probabilities
RSMC New Delhi	5 days / 6 hourly (and 3 hourly updates for intervening)	No
TCWC Jakarta	3 days / 6 hourly	No
TCWC Perth	7 days / 1, 3 or 6 hourly	No
TCWC Wellington	24 hours but up to 5 days if a threat to NZ / 6 hourly	No
JTWC	5 days / 6 hourly	No (only qualitatively in discussions)
PAGASA	Only forecast the TC by category / 6 hourly	No
WFO Guam	5 days / 6 hourly	No
Thai Meteorological Department	3 days / 3, 6, 12 hourly depending on situation	No
MetMalaysia	7 days / 6 hourly	No
Météo-France New Caledonia	3 days / 6 hourly	No
Météo-France Martinique	5 days / 6 hourly	No
Hong Kong Observatory	5 days / 24 hourly	No
Canadian Hurricane Center	5 days / 6 hourly	No

(NWP-TCEFP, Yamaguchi et al., 2012, 2014, 2015). Other examples were the successful integration of ensemble probabilities in the production of Tropical Cyclone Potential Bulletins at TCWC Wellington, and the calibrated genesis potential guidance from the CFS, GEFS, and ECMWF ensembles that are used directly in the creation of the Global Tropical Hazards and Benefits outlook created by the NWS Climate Prediction Center.

Just over half of respondents stated that they now placed more emphasis on using dynamical ensemble forecast information rather than using the traditional consensus approach using multiple deterministic forecasts. Of the consensus-favouring group, several said this was mainly due to habit, tradition, or because they had neither the desired access to, understanding of, nor the tools required to use the dynamical ensemble forecasts. Others felt the consensus approach was more skilful, particularly at shorter lead times and for intensity and structure forecasts, due to the higher resolution and the different model physics, biases, and parameterizations when combining multiple deterministic models. Several people felt that they would have liked to have answered that both were equally important, with the consensus forecasts better suited to the required deterministic components of the products and warnings, and ensembles being more useful for probabilistic forecasts and communicating uncertainty, while others requested research to compare the skill from each approach, and develop methods that combine and maximize the benefit from both consensus and ensemble forecasts.

The main hurdles experienced that had prevented ensemble forecasts from being fully utilized fell into four broad categories:

i) *Availability of ensemble data* e.g. lack of access to ensemble data at their center, lack of inclusion of ensembles in their key operational or visualization tools, and late availability time in an operational context (particularly from ECMWF ENS). Several respondents named TIGGE CXML as a great resource but would like to see it extended so that all centers provide forming information, and additional information such as intensity, structure and precipitation, which would increase research opportunities and lead to more research to operations (R2O) achievements in the future. Another respondent would like to see ensemble forecast information consistently available in Automated Tropical Cyclone Forecasting System (ATCF) format.

ii) *Performance of the ensemble forecasts* e.g. concerns that the resolution is not useful for intensity prediction, about unrealistic ensemble spread (respondents commented both on excessive spread with unrealistic solutions and on the forecasts often being under-spread), and that ensembles can struggle to represent the vortex and get the structure that is essential to assess the hazard risks.

iii) *Lack of familiarity with the interpretation of ensemble and probabilistic forecasts* by forecasters, and problems with synthesizing the enormous amount of data in a time-restricted operational environment.

iv) *Challenge of communicating uncertainty and probabilistic forecasts to customers and the public,* who are used to deterministic weather bulletins and warnings.

2.4 Priorities for future research and development

The final section of the questionnaire asked the respon-

TABLE 3. Operational genesis forecasts from each tropical cyclone forecasting center.

Center name	Forecast lengths / frequency	How is uncertainty represented?	How is the uncertainty calculated?
RSMC Tokyo	1 day / 8 times a day	Not represented	n/a
RSMC Honolulu	5 days / 6 hourly	High/medium/low categories with assigned probabilities	Subjectively, forecasters use climatology, multi-model consensus, ensembles, and satellite observation trends
RSMC La Reunion	5 days / daily	Probability categories (very low, low, moderate, high, very high)	Ensemble based and forecaster judgement
RSMC Miami	5 days / 6 hourly	Percentages to nearest 10% that are grouped into high/medium/low categories	Combination of ensembles, multi-model ensemble, and forecaster experience
RSMC New Delhi	5 days / daily	Nil, Low, Fair, Moderate, High corresponding to 00, 1-25, 26-50, 51-75, 76-100% of probability	Based on consensus derived from about 10 deterministic and probabilistic model guidance
TCWC Jakarta	3 days / 6 hourly	high/medium/low	Climatological, consensus
TCWC Perth	3-7 days / daily	High/moderate/low/very low	Assigned by forecaster
TCWC Wellington	5 days / daily	Low potential/moderate potential/ high potential	Uncertainty is based on ensemble (mostly ECMWF and MOGREPS tropical storm strike/ genesis probabilities) and forecaster diagnosis and prognosis of the environmental conditions susceptible to TC genesis
JTWC	14 days / 12 hourly	Low (<40%), Medium (40-60%), High (>60%) classification	Deterministic and ensemble model forecasts, MJO and statistical-dynamical tools
PAGASA	48 hours / daily	Using terms like "less likely", "likely" to develop	Ensemble based and other trends in TC parameters
WFO Guam	5 day / 12 hourly	Discussion, and use of JTWC high, medium and low	NWP, climatology and experience with synoptic data (including upper air forecasts)
Thai Meteorological Department	5 days /6 or 12 hourly depending on situation	Probability	Ensemble-based
Météo-France New Caledonia	7 days / daily	High/medium/low	Ensemble-based and multi-model
Météo-France Martinique	5 days (6 hourly), one week (12 hourly) and one month (Mon/Thu).	5 days (high/medium/low) Longer range (percentiles)	5 days: from NHC guidance Longer range: from ECMWF
Canadian Hurricane Center	Follow National Hurricane Center (NHC) guidance.	n/a	n/a

dents where they would like to see future research and development focus to maximise the utility of and benefit from ensemble forecasts in operational tropical cyclone forecasting. Figure 8 shows the scores from 1 (not important) to 5 (very important) given by the respondents when asked which of the statements would help them make better use of ensemble forecast information in the future. All of the ideas were popular, with the yellow, green and orange bars (representing the higher end of the scale) dominating on all of the statements. However, the top statements having the highest number of respondents selecting a "5" were: i) Improvement in the skill of the ensemble forecasts (intensity); ii) Greater access to ensemble forecast data; iii) Improvement in the skill of the ensemble forecasts (track); iv) Better understanding of best practice in ensemble-based forecasting of TCs; and v) The development of hazard-based ensemble products.

Hazard-based and impact-based products also featured strongly when asked for specific details of a particular product that they would like to be able to see. Other requests were an operational ensemble-based cone of probability, more useful intensity guidance including probability of intensity change and bias corrected intensity forecasts from multiple ensembles, and ensemble clustering with probabilities.

The most popular verification metric amongst the respondents for evaluating the skill and value in ensemble forecasts of tropical cyclones was ensemble spread, followed by the root mean square (RMS) error of the ensemble mean, reliability diagrams, and the % of forecasts falling within an ensemble defined ellipse/cone (see Figure 9). Several ensemble-based verification metrics such as Rela-

TABLE 4. Operational hazard forecasts from each tropical cyclone forecasting center.

Center name	In hazard forecasts e.g. precipitation, winds, storm surge, flooding, how is uncertainty represented?	How is the uncertainty calculated?
RSMC Tokyo	Hazard forecasts not currently included in our service	n/a
RSMC Honolulu	Wind speed probability graphics for 34/50/65 knots Adjective category and color coded hazard threat index: - Little to None - Elevated (changed from Low) - Moderate - High - Extreme	Probabilistic according to the hazard e.g. wind speed probabilities
RSMC La Reunion	No probabilistic data available for the public	n/a
RSMC Miami	In categories (representing probabilities), that vary depending on the hazard.	Generally ensemble based, either statistically or direct from ensemble output
RSMC New Delhi	Forecast for heavy precipitation is given in range (7-11 cm, 12-20 cm and more than 20 cm in 24 hrs) and likelihood of occurrence (likely: 26-50%, very likely: 51-75% and most likely: 76-100%). The winds, storm surges and flooding are also mentioned in ranges with above probability of occurrence.	Uncertainty is calculated based on likelihood of occurrence and confidence level of forecast based on various deterministic and ensemble models. Accordingly the colour codes of green, yellow, orange and red are used in increasing order of severity based on confidence level and likelihood of occurrence.
TCWC Perth	Probabilistic language (likely, possibly etc)	Assigned by forecaster
TCWC Wellington	High/medium/low risk in severe weather outlooks	Subjective assessment based on analysis of system and environment, deterministic forecasts, and ensemble outlooks.
JTWC	n/a	n/a
PAGASA	In description e.g. we say that the hazard is possible to happen	Based on daily observations and deterministic forecast from other models
WFO Guam	Descriptively	Forecast accuracy and type of system
Thai Meteorological Department	Show on map	Risk matrix
Météo-France New Caledonia	There is no graphical representation of hazard forecasts. They are mentioned in bulletins with key words ("probably", "risk", "certainly" etc).	Ensemble-based and multi-model
Météo-France Martinique	Not available for the public	ECMWF EPS and ARPEGE PEARP
Hong Kong Observatory	Uncertainty in the change of tropical cyclone warning signal, which is related to local wind forecast, is represented in five categories of probability for government departments	Based on the probabilistic relationship between local winds and position/intensity of tropical cyclone
Canadian Hurricane Center	Listing the worst case scenario and the sensitivity of forecast errors. Qualitatively in free form text discussion.	Empirically, plus local effects and criteria (i.e. known flooding levels for storm surge)

tive Operating Characteristic (ROC) diagrams, Brier Skill Score, Ranked Probability Skill Score, and in particular economic value plots, were not as well used or understood, and therefore ensemble-based verification may need to be communicated more clearly, or adapted to provide more user-relevant information. The respondents gave several ideas for more user-oriented verification information that they would like to see that are more relevant to their day-to-day forecasting experience. For example, only including members that had a similar track to the actual event in intensity verification, creating and verifying metrics that translate ensemble spread into categorical estimates of forecast confidence in comparison to the widely used historical forecast error statistics, and regularly updated or real-time performance metrics comparing the performance and skill for all types of models (deterministic, consensus and ensemble

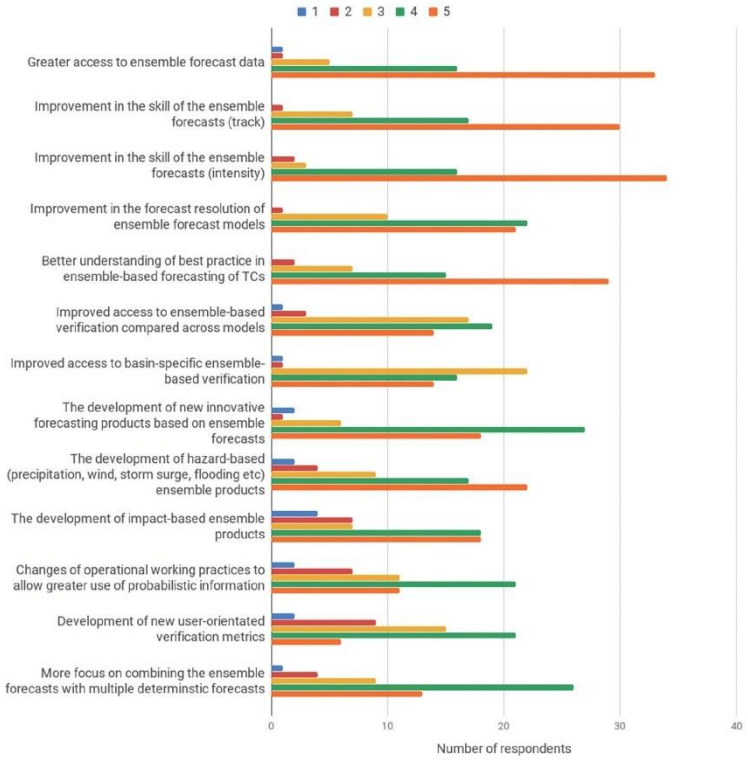


FIG. 8. The questionnaire asked which of the above statements would help them make better use of ensemble forecast information in the future, rating each idea from 1 (not important) to 5 (very important).

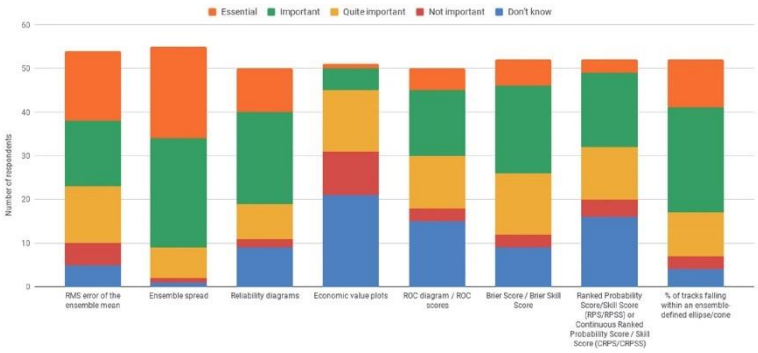


FIG. 9. "Please state how important you feel each of the following verification metrics are when evaluating the skill and value in ensemble forecasts."

models) for the current storm, basin, and season-to-date.

At the end of the questionnaire respondents were asked what specific developments they would like to see. The key themes coming through these responses are detailed below:

- Several responses urged a focus on how to ***communicate uncertainty to the general public***, and explain how the public can use probabilistic information, with a need to move away from the very deterministic-orientated information given in weather charts and bulletins. One respondent emphasised that a key challenge for the future was to implement changes to operational products and workflows that better utilise ensemble forecasts while balancing the communication of a clear and concise warning product.
- Others felt the ***development of new tools, products and visualizations*** to allow forecasters to rapidly and efficiently view and interpret ensemble data should be the main goal in the coming years, given the limited time available in operational context. One respondent stressed the importance of integrating ensemble data into the forecast platform, rather than relying on external sources such as internet products.
- ***Collaboration between all NWP centers and forecasting centers*** was requested, to share ensemble data in standard formats, products, verification packages and results, and expertise on interpreting ensemble forecast data and including ensemble information in forecast and warning products.
- ***Training of forecasters on the utilisation of ensembles*** was mentioned by several respondents as essential along with the need for a ***change in operational working practice towards using probabilistic information***. This need is not unique to tropical cyclone forecasting, as illustrated by a recent survey of meteorologists' understanding of forecast uncertainty and probabilistic forecast guidance at NOAA's Hazardous Weather Testbed (Wilson et al., 2019), which found substantial variability in the extent to which participants could think beyond a deterministic mindset and successfully synthesize and apply uncertainty guidance information.
- Several users requested a focus on the development of ***ensemble-based hazard-based forecasts*** e.g. storm surge, coastal inundation, inland flash flood and river flooding, especially where the risk is focussed remotely (away from the center of the cone). This should include research into evaluating and improving the ensemble's ability to resolve the details of the vortex in order to allow a better assessment of the predictability of the associated hazards.
- The ***development of impact-based ensemble products*** was seen as crucial by a number of respondents, with access to and/or collection of socio-economic data sources being one key issue here.
- ***Improvement in the skill of ensemble forecasts of tropical cyclone intensity and structure*** was seen as a key priority for some. Improved availability of bias-corrected ensemble forecasts, or methods to do so, was also requested.
- ***User-oriented and basin specific verification comparing ensemble guidance across models*** was seen as important to help forecasters gain confidence in using ensemble forecast information. One respondent said it would help if training were available on the performance and skill of all models (deterministic, consensus, and ensemble models) in their basin. Another emphasized the importance of research into how to optimally combine the guidance from all available models (deterministic and ensemble, global and regional).

The results of the questionnaire should help to provide guidance to NWP modelling centers, researchers working in this area, and managers in the operational forecasting centers, on where further development should focus to maximize the potential benefit of ensemble forecasting in operational TC forecasting.

3. Conclusions and recommendations

The questionnaire results show that dynamic ensemble forecasts are currently used by nearly all operational tropical cyclone forecasters and at TC forecast centers around the world, although there is a need to ensure that ensemble forecast information is more widely available to forecasters in developing countries. The survey showed that ensemble forecasts are seen to be particularly important in track and genesis forecasting, but of only limited use in intensity and structure forecasting due to the difficulty of adequately resolving and forecasting these in current global dynamical ensemble models. The responses to several questions point to the ensembles being introduced into a very deterministic-focussed environment, where the ensemble member tracks, or even just the ensemble mean track forecasts, are viewed alongside deterministic forecasts, rather than forecasters being able to utilise the full probabilistic uncertainty information that is available from ensemble forecasts. There are several examples of excellent pull-through of ensemble forecasts from research into the operational TC forecasting process. However, many forecast centers still use historical error statistics rather than dynamic situation-based uncertainty in determining the width of the uncertainty cone/circle/swath in their operational track warning products. The survey therefore reveals a notable difference between the high proportion of operational TC forecasters who use and value ensemble forecast information, and the slower pull-through into operational warnings and products of the probabilistic guidance and uncertainty information that ensembles can provide.

The survey highlighted several areas of research and development that would help TC forecasters to make increased use of ensemble forecast information in the future: a) an improvement in the skill of the ensemble forecasts (in particular for intensity); b) improved access to ensemble forecast data, tools, visualizations and user-orientated verification; c) better understanding of best practice in

ensemble-based forecasting of TCs; d) the development of hazard-based and impact-based ensemble products; and e) a change in operational working practices towards using probabilistic information, and providing and communicating dynamic uncertainty information in operational forecasts and warnings.

The specific recommendations put forward and accepted by IWTC-9 based on the results of the questionnaire were as follows:

• Researchers are encouraged to promote research activities to evaluate and improve probabilistic TC intensity and structure forecasts (e.g. Magnusson et al. 2018).

• Researchers are encouraged to promote research activities to demonstrate the usefulness of dynamical ensemble forecasts for hazard-based products such as for strong wind, precipitation, flooding and storm surge, and the use of ensembles in impact-based forecasting for TCs.

• Greater access to forecast data including deterministic/ensemble forecast data and global/regional forecast data would be beneficial. Research is encouraged to evaluate how best to combine the information from these models, in order to maximize forecast skill and reliability.

• The research community should recognize the value and importance of access to forecast data such as TIGGE and further promote the use of such data, including wider sharing of code for product generation and verification to facilitate the operational use of ensemble forecasts.

• There is a need for greater collaboration both amongst the operational forecast centers, and between operational forecast centers, NWP modelling centers, and researchers, in order to co-ordinate the development of new ensemble-based tools and methods, and develop and share best practice on the use of ensemble forecasts in operational TC forecasting.

• The operational forecasting community are encouraged to facilitate a change in operational working practice towards using ensemble-based uncertainty and probabilistic information in forecasts and warnings, including a focus on how best to communicate uncertainty to the general public.

Appendix: Survey questions

Introduction

- Organisation (including RSMC/TCWC where applicable)
- Basin of interest [North Indian Ocean, Western North Pacific, Central Pacific, Eastern Pacific, Atlantic, South Indian Ocean, Australian, South Pacific]
- Job role [Operational forecaster, Manager, Researcher, Other]
- Do you currently use ensemble forecast information [Yes, No]

Current use of ensemble forecast information

- How important would you say ensemble forecasts are in your forecasting of tropical cyclones (Not used, Somewhat important, Important, Very important)? [Impact-based forecasting, Track forecasting, Hazard forecasting (precipitation, flooding, wind, storm surge), Extra-tropical transition, Intensity forecasting, Tropical cyclone size and/or structure, Genesis forecasting]
- For each of the main areas of tropical cyclone forecasting (Track, Intensity, Genesis, Hazards), which ensemble forecast models do you use? [NCEP GEFS, ECMWF ENS, MOGREPS-G (UK Met Office global ensemble), JMA's Global Ensemble Prediction System, Canadian Global Ensemble Prediction System, CMA ensemble, KMA ensemble, COAMPS-TC ensemble, FNMOC ensemble, Bureau of Meteorology ensemble, CPTEC ensemble, Meteo-France ensemble, DWD ensemble, Other]
- Do you use multi-model ensemble forecast data (i.e. combination of multiple dynamical ensembles)? [Yes, in full probabilistic multi-model ensemble combination, Yes, but only by comparing one ensemble with another, No]
- Please tick to indicate if you use calibrated ensemble forecasts and for which type of forecast [Track, Intensity, Genesis]
- How do you access ensemble forecast information? [From products sent by global NWP centres, From products created in your centre, From online products from other centres, From the TIGGE archive, From WMO Global Telecommunications System (GTS), Other]
- Which ensemble-based or probabilistic products do you use (Not available or required, Not available but would like to have, Available but don't use, Sometimes use, Regularly use)? [Ensemble mean track, Tracks from ensemble members, Large-scale flow information, Storm intensity, Strike probability, Tropical cyclone activity (named and forming storms), Tropical Cyclone genesis, Landfall probability, Precipitation, Wind speed, Storm surge, River discharge, Flood risk / inundation, Impact-based products]
- If there is a product that you find particularly helpful please give the product and model details below.
- If there is a particular product or ensemble that you would like to be able to see please give the details below.
- Which of the following statements describe your use of ensemble forecast information in track forecasting? [I mainly use deterministic forecasts, and rarely use ensemble forecasts, I mainly use ensemble forecasts to compare the ensemble mean forecasts to deterministic forecasts, I find the ensemble spread useful to give me the situation-dependent uncertainty, I use the full probabilistic information from the ensemble forecasts in my operational forecasting, I don't find ensemble forecasts very helpful when the spread is very large]
- Which of the following statements describe your use of ensemble forecast information in intensity forecasting? [I mainly use deterministic forecasts, and rarely use ensemble forecasts, I mainly use ensemble forecasts to compare the ensemble mean forecasts to deterministic forecasts, I find the ensemble spread useful to give me the situation-depen-

dent uncertainty, I use the full probabilistic information from the ensemble forecasts in my operational forecasting, I don't find ensemble forecasts very helpful when the spread is very large]

- Please state how important you feel each of the following verification metrics are when evaluating the skill and value in ensemble forecasts (Don't know, Not important, Quite important, Important, Essential) [RMS error of the ensemble mean, Ensemble spread, Reliability diagrams, Economic value plots, ROC diagram / ROC scores, Brier Score / Brier Skill Score, Ranked Probability Score/Skill Score (RPS/RPSS) or Continuous Ranked Probability Score / Skill Score (CRPS/CRPSS), % of tracks falling within an ensemble-defined ellipse/cone, Other (please specify)]

- If you have an idea for a particular "user-orientated" verification metric from ensemble forecasts please detail below.

Communication of forecast uncertainty in operational forecasts and warnings

- Track: In operational track forecasts from your organisation, what forecast lengths are issued?

- Track: With what frequency are forecasts issued?

- Track: How is uncertainty represented? [Cone, Circle, Error swath, Don't know, Other]

- Track: How is the cone/circle/swath defined?

- Track: Please detail how the size of cone/circle/swath is calculated for each storm? e.g. historical forecast error statistics, climatological uncertainty, consensus spread, ensemble spread, ensemble-based uncertainty etc.

- Intensity: In operational intensity forecasts from your organisation, what forecast lengths are issued?

- Intensity: With what frequency are forecasts issued?

- Intensity: Is uncertainty represented? [Yes, No]. If yes, how is uncertainty in intensity forecasts calculated?

- Genesis: In genesis forecasts at your organisation, what forecast lengths are used?

- Genesis: With what frequency are genesis forecasts issued?

- Genesis: How is uncertainty represented in the genesis forecasts? e.g. high/medium/low or probability categories etc.

- Genesis: How is the uncertainty calculated in the genesis forecasts? e.g. climatological, ensemble-based, other methods etc.

- Hazard: In hazard forecasts from your organisation (e.g. precipitation, winds, storm surge, flooding), how is uncertainty represented?

- Hazard: In hazard forecasts from your organisation, how is uncertainty calculated?

Examples of the use of ensemble forecasts

- Please give an example of where probabilistic ensemble forecasts have been successfully integrated into the operational forecasting process.

- Please give an example of where hurdles have prevented probabilistic ensemble forecasts being successfully integrated into the operational forecasting process.

- Do you place more emphasis on the traditional consensus approach to tropical cyclone forecasting (using multiple deterministic forecasts), rather than the use of dynamical ensemble forecast information? [Yes, No]. If you answered yes, please explain why below.

Future use of ensemble forecast information

- Which of the following would help you to make better use of ensemble forecast information in the future (Please rate each idea from 1 (not important) to 5 (very important))? [Greater access to ensemble forecast data, Improvement in the skill of the ensemble forecasts (track), Improvement in the skill of the ensemble forecasts (intensity), Improvement in the forecast resolution of ensemble forecast models, Better understanding of best practice in ensemble-based forecasting of TCs, Improved access to ensemble-based verification compared across models, Improved access to basin-specific ensemble-based verification, The development of new innovative forecasting products based on ensemble forecasts, The development of hazard-based (precipitation, wind, storm surge, flooding etc.) ensemble products, The development of impact-based ensemble products, Changes of operational working practices to allow greater use of probabilistic information, Development of new user-orientated verification metrics, More focus on combining the ensemble forecasts with multiple deterministic forecasts]

- Please give further details of your thoughts when you were giving your answers above e.g. what specific developments would you like to see?

- Please use this space to put any additional feedback on any other topic relating to the ensemble forecasting and verification of tropical cyclones.

Acknowledgements

The authors would like to offer their sincere thanks Mike Brennan, Operational Co-Chair of IWTC-9, for his support and encouragement, and the 60 respondents to the questionnaire, to whom we are very grateful.

References

Buizza, R., 2008: The value of probabilistic prediction. *Atmosph. Sci. Lett.*, **9**, 36-42.

Du, Y.G., L.B. Qi and X.G. Cao, 2016: Selective ensemble-mean technique for tropical cyclone track forecast by using time-lagged ensemble and multi-centre ensemble in the western North Pacific. *Quart. J. Roy. Meteor. Soc.*, **142**, 2452-2462.

Dupont, T., M. Plu, P. Caroff and G. Faure, 2011: Verification of ensemble-based uncertainty circles around tropical cyclone track forecasts. *Weather and Forecasting*, **26**: 664–676.

Goerss, J., 2000: Tropical cyclone track forecasts using an ensemble of dynamical models. *Mon. Wea. Rev.*, **128**, 1187-1193.

Goerss, J., 2007: Prediction of consensus tropical cyclone track

forecast error. *Mon. Wea. Rev.*, **135**, 1985-1993.

Goerss, J.S. and C.R. Sampson, 2014: Prediction of consensus tropical cyclone intensity forecast error. *Wea. Forecasting*, **29**, 750-762.

Leonardo, N.M. and B.A. Colle, 2017: Verification of Multimodel Ensemble Forecasts of North Atlantic Tropical Cyclones. *Wea. Forecasting*, **32**, 2083–2101.

Magnusson, L., J.R. Bidlot, M. Bonavita, A. Brown, P. Browne, G. De Chiara, M. Dahoui, S.T.K. Lang, T. McNally, K.S. Mogensen, F. Pappenberger, F. Prates, F. Rabier, D.S. Richardson, F. Vitart and S. Malardel, 2019: ECMWF activities for improved hurricane forecasts. *Bull. Amer. Meteor. Soc.,* **100 (3)**, 445-458.

Palmer, T., 2018: The ECMWF ensemble prediction system: Looking back (more than) 25 years and projecting forward 25 years. *Q J R Meteorol Soc*, 1– 13.

Sampson, C.R., J.L. Franklin, J.A. Knaff and M. DeMaria, 2008: Experiments with a Simple Tropical Cyclone Intensity Consensus. *Wea*. Forecasting, **23**, 304-312.

Swinbank, R., M. Kyouda, P. Buchanan, L. Froude, T.M. Hamill, T.D. Hewson, J.H. Keller, M. Matsueda, J. Methven, F. Pappenberger, M. Scheuerer, H.A. Titley, L. Wilson and M. Yamaguchi, 2016: The TIGGE Project and Its Achievements. *Bull. Amer. Meteor. Soc.*, **97**, 49–67

Wilson, K.A., P.L. Heinselman, P.S. Skinner, J.J. Choate, and K.E. Klockow-McClain, 2019: Meteorologists' Interpretations of Storm-Scale Ensemble-Based Forecast Guidance. *Wea. Climate Soc.*, **11**, 337–354

World Meteorological Organisation (WMO), 2014: HIWeather Implementation Plan (available online at https://www.wmo.int/pages/prog/arep/wwrp/new/documents/HIW_IP_v1_4.pdf)

Yamaguchi, M., R. Sakai, M. Kyoda, T. Komori and T. Kadowaki, 2009: Typhoon Ensemble Prediction System developed at the Japan Meteorological Agency, *Mon. Wea. Rev.*, **137**, 2592-2604.

Yamaguchi, M., T. Nakazawa and S. Hoshino, 2012: On the relative benefits of a multi-centre grand ensemble for tropical cyclone track prediction in the western North Pacific. *Q.J.R. Meteorol. Soc.*, **138**, 2019-2029.

Yamaguchi, M., T. Nakazawa, and S. Hoshino, 2014: North Western Pacific Tropical Cyclone Ensemble Forecast Project. *Tropical Cyclone Research and Review*, **3**, 193-201.

Yamaguchi, M., F. Vitart, S.T. Lang, L. Magnusson, R.L. Elsberry, G. Elliott, M. Kyouda and T. Nakazawa, 2015: Global Distribution of the Skill of Tropical Cyclone Activity Forecasts on Short- to Medium-Range Time Scales. Weather and Forecasting, **30**, 1695–1709.

Yamaguchi, M., J. Ishida, H. Sato, and M. Nakagawa, 2017: WGNE Intercomparison of Tropical Cyclone Forecasts by Operational NWP Models: A Quarter Century and Beyond. *Bull. Amer. Meteor. Soc.*, **98**, 2337-2349.

Yamaguchi, M., H.A. Titley, and L. Magnusson, 2018: Current and potential use of ensemble forecasts in operational TC forecasting. Ninth International Workshop on Tropical Cyclones, Hawaii, World Meteorological Organization (available online at https://www.wmo.int/pages/prog/arep/wwrp/tmr/documents/IWTC-9_Subtopic_6-3.pdf)

Zhang X.P. and H. Yu, 2017: A probabilistic tropical cyclone track forecast scheme based on the selective consensus of ensemble prediction systems. *Weather and Forecasting*, **32**, 2143-2157.

A2: Published Article: A global evaluation of multi-model ensemble tropical cyclone track probability forecasts

This appendix contains the published version of chapter 4 of this thesis, with the following reference:

Titley, H.A., Bowyer, R.L. and Cloke, H.L. (2020) A global evaluation of multi-model ensemble tropical cyclone track probability forecasts. *Quarterly Journal of the Royal Meteorological Society*, 146 (726): 531–545. doi:10.1002/qj.3712.

Received: 18 July 2019 | Revised: 21 November 2019 | Accepted: 24 November 2019 | Published on: 1 January 2020

DOI: 10.1002/qj.3712

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

A global evaluation of multi-model ensemble tropical cyclone track probability forecasts

Helen A. Titley[1,2] | Rebecca L. Bowyer[1] | Hannah L. Cloke[2,3,4,5]

[1]Weather Science, Met Office, Exeter, UK

[2]Department of Geography and Environmental Science, University of Reading, Reading, UK

[3]Department of Meteorology, University of Reading, Reading, UK

[4]Department of Earth Sciences, Uppsala University, Uppsala, Sweden

[5]Centre of Natural Hazards and Disaster Science, CNDS, Uppsala, Sweden

Correspondence
Helen A. Titley, Met Office, FitzRoy Road, Exeter, EX1 3PB, UK.
Email: helen.titley@metoffice.gov.uk

Funding information
Newton-Bhabha Fund, Met Office WCSSP India Programme

Abstract

At the Met Office, dynamic ensemble forecasts from the Met Office Global and Regional Ensemble Prediction System (MOGREPS-G), the European Centre for Medium-Range Weather Forecasts Ensemble (ECMWF ENS) and National Centers for Environmental Prediction Global Ensemble Forecast System (NCEP GEFS) global ensemble forecast models are post-processed to identify and track tropical cyclones. The ensemble members from each model are also combined into a 108-member multi-model ensemble. Track probability forecasts are produced for named tropical cyclones showing the probability of a location being within 120 km of a named tropical cyclone at any point in the next 7 days, and also broken down into each 24-hour forecast period. This study presents the verification of these named-storm track probabilities over a two-year period across all global tropical cyclone basins, and compares the results from basin to basin. The combined multi-model ensemble is found to increase the skill and value of the track probability forecasts over the best-performing individual ensemble (ECMWF ENS), for both overall 7-day track probability forecasts and 24-hour track probabilities. Basin-based and storm-based verification illustrates that the best performing individual ensemble can change from basin to basin and from storm to storm, but that the multi-model ensemble adds skill in every basin, and is also able to match the best performing individual ensemble in terms of overall probabilistic forecast skill in several high-profile case-studies. This study helps to illustrate the potential value and skill to be gained if operational tropical cyclone forecasting can continue to migrate away from a deterministic-focused forecasting environment to one where the probabilistic situation-based uncertainty information provided by the dynamic multi-model ensembles can be incorporated into operational forecasts and warnings.

KEYWORDS

ensembles, probabilistic forecasting, tropical cyclones, uncertainty, verification

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2019 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

1477870x, 2020, 726, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3712 by Test, Wiley Online Library on [29/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1 | INTRODUCTION

In dynamic ensemble forecasting, instead of making a single forecast of the most likely weather conditions, a forecast model is run multiple times to produce an ensemble of forecasts. These ensemble forecasts take into account uncertainty in the initial conditions and imperfections in the model formulation, and aim to give an indication of the range of possible future states of the atmosphere. For over 25 years, these dynamical ensemble model forecasts have been routinely produced by several global numerical weather prediction modelling centres, including the European Centre for Medium-Range Weather Forecasts Ensemble (ECMWF ENS: Palmer, 2019), the National Centers for Environmental Prediction Global Ensemble Forecast System (NCEP GEFS: Toth and Kalnay, 1997) and the Met Office Global and Regional Ensemble Prediction System (MOGREPS-G: Bowler *et al.*, 2008).

These ensemble forecast models have an important role to play in tropical cyclone forecasting, through their ability to highlight the situation-dependent uncertainty and provide probabilistic forecast information to help inform decision makers. Consequently, many global modelling centres produce tropical cyclone track forecasts from their ensemble forecast models, develop ensemble tropical cyclone forecast products, verify these forecasts, and share the forecast track data via the TIGGE (The International Grand Global Ensemble) cyclone exchange programme (Swinbank *et al.*, 2016). Several studies have shown the benefits of dynamic ensemble forecasting for both existing tropical cyclones (e.g. Dupont *et al.*, 2011; Yamaguchi *et al.*, 2012; Du *et al.*, 2016; Leonardo and Colle, 2017; Zhang and Yu, 2017), and for providing probabilistic information about tropical cyclone genesis (e.g. Yamaguchi *et al.*, 2015; Yamaguchi and Koide, 2017). However, a recent survey among operational tropical cyclone forecasters (Titley *et al.*, 2019) showed that although ensemble forecasts are used and valued by almost all forecasters, the deterministic-focussed forecasting environment has often limited the extent to which the full probabilistic information provided by ensembles has been pulled through into operational tropical cyclone forecasts and warnings.

Objectively identifying the forecast track of each ensemble member is essential both for post-event model evaluation and for the generation of forecast guidance products in real time. Various tracking techniques are used by operational centres around the world (Vitart and Stockdale, 2001; Van der Grijn, 2002; Tallapragada *et al.*, 2013) and in research (Hodges, 1995). At the Met Office, a bivariate (850 hPa relative vorticity and mean sea-level pressure) tropical cyclone tracker known as MOTCTracker (Heming, 2017) is run in real time on the Met Office MOGREPS-G ensemble, and the tracks are made available to Regional Specialised Met Centres (RSMCs) and via the research TIGGE cyclone CXML archive.

Although ensemble forecasts are a good way of assessing forecast uncertainty, they are limited to the uncertainty captured by a specific modelling system, and there is a tendency for single-model ensembles to be under-spread, with the observations too often falling outside of the range of solutions. A multi-model ensemble approach, where dynamic ensemble systems from multiple centres are combined together into a grand multi-centre ensemble (thereafter called multi-model ensemble), can help address this shortcoming and provide a more complete representation of the uncertainty in the model structure, also potentially reducing the errors. The rationale behind multi-model ensemble forecasting was summarised by Hagedorn *et al.* (2005) who stated that "the key to the success of the multi-model concept lies in combining independent and skilful models, each with its own strengths and weaknesses". Several studies have demonstrated that probabilistic forecast skill and reliability can be improved through the use of multi-model ensembles, including Park *et al.* (2008), Johnson and Swinbank (2009), Hagedorn *et al.* (2012), Hamill (2012) and Matsueda and Nakazawa (2015).

The application of a grand combined multi-model ensemble approach in tropical cyclone forecasting is a natural extension of the "consensus" forecasting approach that has been a valuable cornerstone of tropical cyclone track and intensity forecasting for many years, where traditionally three or more deterministic forecasts have been combined or averaged into a "consensus" forecast and also used to give a prediction of likely forecast error (e.g. Goerss, 2000; 2007; Sampson *et al.*, 2008; Goerss and Sampson, 2014; Yamaguchi *et al.*, 2017). The process of combining together several ensemble forecast models into a multi-model ensemble combines the strengths of the consensus and ensemble approaches, by pulling through the full probabilistic forecast information from several dynamic ensembles and model formulations into multi-model ensemble probability forecasts.

At the Met Office, in order to produce real-time multi-model ensemble tropical cyclone products, MOTC-Tracker is also run in real time on the direct input data from the ECMWF ENS and NCEP GEFS ensembles. The ensemble forecast tracks from the ensemble members from these models are combined with the Met Office MOGREPS-G ensemble tracks, to create a 108-member multi-model ensemble. The three corresponding deterministic models (the Met Office Unified Model, the ECMWF IFS, and the NCEP GFS model) are also tracked. A range of products, including track and intensity forecasts for both named and developing tropical cyclones, are produced and used by forecasters in the Met Office Global Guidance Unit. An example of the track and track

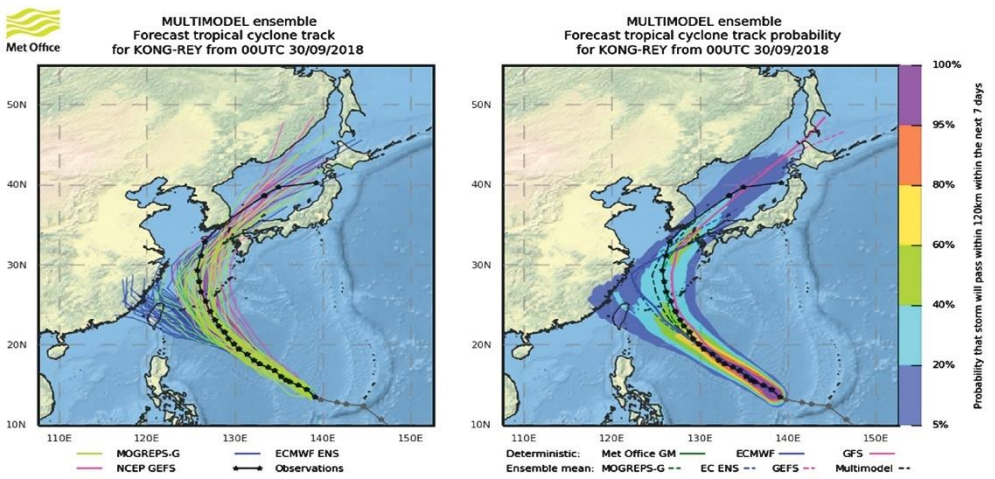


FIGURE 1 Multi-model ensemble forecasts for Typhoon Kong-rey from 0000 UTC 30 September 2018: Tracks coloured according to model (left), and multi-model ensemble track probability with deterministic (solid) and ensemble mean (dashed) tracks (right)

probability products produced for named tropical cyclones from the multi-model ensemble is shown in Figure 1 for Typhoon Kong-rey. The products are also distributed to several operational tropical cyclone forecasting centres, including the RSMCs in Miami and La Réunion. The tracks and track probabilities from each of the ensembles and the multi-model ensemble are also displayed in the Met Office Global Hazard Map (Robbins and Titley, 2018), where they can be viewed alongside the associated wind and rain hazard information, and overlain on vulnerability and exposure fields.

To fully assess the skill and value of the multi-model ensemble tropical cyclone forecasts produced at the Met Office, a framework to produce objective verification of named tropical cyclone track probability forecasts has been developed, the results of which are presented in this article. Previous studies evaluating named tropical cyclone track probability forecasts using multi-model ensembles have focussed on only one or two Northern Hemisphere basins, e.g. Majumdar and Finocchio, 2010 (North Atlantic and Northwest Pacific); Yamaguchi *et al.*, 2012 (Northwest Pacific); and Leonardo and Colle, 2017 (North Atlantic). In this article the verification results are analysed for all named tropical cyclones in all tropical cyclone basins around the globe during 2017 and 2018, and then split into each basin in order to compare the absolute and relative performance of the ensembles across basins. Storm-based verification results have also been calculated for some high-profile tropical cyclones in order to compare the results from storm to storm.

The key questions addressed by this article are:

- How do forecast performance and characteristics vary between the global ensembles, and with lead time?
- Is there benefit in terms of probabilistic forecast skill, reliability and value, from combining the three global ensembles into a multi-model ensemble?
- Does using the full probability forecast information via the multi-model ensemble add skill compared to a "consensus" forecast of the parent deterministic models?
- When comparing basin to basin, does forecast performance vary, both overall compared to a reference consensus forecast, and relatively between the different ensembles?
- When comparing storm to storm, how does forecast performance vary, and what does it tell us about the benefit of multi-model ensembles?

The verification framework and methodology are described in Section 2. Section 3 presents the results, split into three sections: (a) 7-day ("overall") track probability forecast results, (b) 24-hour track probability forecast results, and (c) storm-specific verification results for two high-profile North Atlantic hurricanes. Section 4 provides a discussion, focussed around the implications of the results for operational forecasters, decision-makers, and model developers, while Section 5 states some key conclusions and ideas for future work.

1477870x, 2020, 726, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3712 by Test, Wiley Online Library on [29/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2 | VERIFICATION METHODOLOGY

The three ensembles tracked in real time at the Met Office using MOTCTracker (Heming, 2017), and included in the multi-model ensemble products, are MOGREPS-G (24 members until 11 July 2017, then 36 members thereafter), ECMWF ENS (51 members) and NCEP GEFS (21 members). The ensemble forecast tracks from the three ensembles are combined to create a 108-member multi-model ensemble (96 members until 11 July 2017), with each individual member from each ensemble equally weighted when creating the track probability forecasts. The track probability forecasts are calculated for named tropical cyclones from each of the three global ensembles, and the combined multi-model ensemble, both for the next 7 days (with no lead time component, to directly verify the forecast product shown in Figure 1 – henceforth described as the "overall" track probability), and for 24-hour forecast periods out to 7 days (in order to evaluate the change in performance over different lead times). Track probabilities are defined as the probability of the named tropical cyclone passing within 120 km of each grid point in the given time period using the commonly used "strike probability" definition first defined in Van der Grijn *et al.* (2004). In this article the term track probability is preferred, because a tropical cyclone is a large weather system that can also "strike" and lead to impacts further away than 120 km from the centre of the storm track, for example from storm surges, flooding, and extreme winds including tornadoes. Thus the term "strike probability" could be misleading to the public as it could falsely imply that those areas outside of 120 km from the centre of the cyclone will not be impacted. Tropical cyclone genesis (storms that are forecast to form during the forecast) is not included in the verification in this article. Observed track data are collected routinely at the Met Office from Global Telecommunications System (GTS) bulletins from the RSMCs, tropical cyclone warning centres, and the Joint Typhoon Warning Center (used if no track is available from the RSMC), to accumulate observed tropical cyclone positions for all named storms to verify the forecasts. The mean track or "consensus" of the three deterministic model tracks is calculated and used as the reference forecast to fully assess the forecast skill.

Since 2015 the verification has been produced biannually, at the end of the Northern and Southern Hemisphere seasons, for tropical cyclones in all global basins over the previous 12 months, in order to evaluate the most recent configurations of each ensemble. This study verifies a 2-year period (January 2017 to December 2018), in order to increase the sample of cases, whilst ensuring the model configurations are still relatively recent.

For each of the named tropical cyclones in the study period, the following steps are carried out in the overall track probability verification process:

1. All observed positions for this storm are read in and the observations are included from the first time period that it has an observed intensity of greater than 34 knots to the last time that it exceeds this value. This ensures that only tropical cyclones that reach tropical storm strength or greater are verified, and that for the included storms, all observations in between these two times are included even if the intensity dips below this threshold in between.
2. A storm-specific verification grid is created for the area covered by the storm track at a resolution of 0.5°.
3. For each of the 0000/1200 UTC forecast run times at which this tropical cyclone was named, a track probability gridded data file is calculated for the observations, the reference consensus forecast, and each forecast model. For the observation data this is a file containing 1's (for grid points which fall within 120 km of the observed track of the storm over the next 7 days), and 0's (those that do not). To create the reference consensus forecast, the mean of the three deterministic forecast tracks is calculated, and a file created containing 1's (for grid points which fall within 120 km of the consensus track of the storm over the next 7 days), and 0's (those that do not). For the ensemble forecast models, each grid point contains a probability value between 0 and 1, calculated by counting how many members have forecast tracks for this storm that lie within 120 km of that point at some point between T+0 and T+168 h, and dividing by the number of ensemble members.
4. For the forecast model data (consensus and ensemble forecasts), the forecasts are only included up to the last time that a matching observation is available. Therefore this verification is focussed on the ability of the models to predict the future track position of the tropical cyclone while it is at tropical storm strength or above, rather than on the ability of the models to weaken and dissipate the storm at the same time as shown in the observations, as in practice the point at which the forecast tracks end is highly dependent on the choice of thresholds for dissipation in the storm tracker.
5. As many forecast runs do not have a full 7 days of matching observed tracks to verify against, there are a considerably smaller overall sample size of tracks verifying at day 7 than at day 1. To prevent the overall track probability verification from being dominated by the verification of shorter lead-time forecasts, which are not so relevant for society given that important decisions such as evacuations generally need to be made on forecasts of two or more days, only those forecast runs

1477870x, 2020, 726, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3712 by Test, Wiley Online Library on [29/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where there are matching observed tracks for T+48 or longer are included.

A similar process is then followed to verify for each 24-hour period from T+0–T+24 through to T+144–T+168. In this case however, the forecasts are verified at the basin-scale, again to replicate the forecast product, which is a basin-based animation of the track probability forecasts for every 24-hour period.

Gridded verification is then carried out comparing the track probability forecast data with the corresponding observed data. A range of probabilistic verification statistics are calculated to assess the skill, reliability and value of the ensemble forecasts. These include the Relative Operating Characteristic (ROC), reliability diagrams, relative economic value and Brier skill score (BSS), as described in Jolliffe and Stephenson (2012), and summarised below:

- The ROC plot assesses the skill of the forecast at discriminating between events and non-events. The points along the curve are the hit rates and false alarm rates for each probability bin. Perfect skill would produce a curve from bottom left to top left to top right, and no skill is indicated by the diagonal line from (0,0) to (1,1).
- Reliability diagrams display how well the predicted probabilities correspond to their observed frequencies. Perfect reliability would be a diagonal line from (0,0) to (1,1), a line above the diagonal indicates under-forecasting and below the diagonal shows over-forecasting, while a line which falls below the diagonal for high probabilities and above it for low probabilities exhibits over-confidence (Wilks, 2011).
- For a given user, their cost-loss ratio is the term given to the ratio of the cost of a preventative measure to the loss averted, and can be used to guide the probability threshold above which to take action. At each probability threshold there will be a 2x2 contingency table containing the number of hits, false alarms, misses and correct rejections. Assuming the user takes action whenever an event is forecast, then a cost (C) can be associated with the hits and false alarms and a loss (L) associated with the misses. The relative economic value can then be calculated for a range of cost-loss ratios and visualised using a relative economic value plot which displays the relative improvement in economic value between the sample climatology and a perfect forecast for all cost-loss ratios (Richardson, 2000). The relative economic value is a useful additional user-focussed measure for comparing forecasts. In practice the user's cost-loss ratio may be difficult to determine and may have to be estimated; however, if one forecast consistently has higher value than another forecast across all cost-loss ratios then it clearly has greater value for any user.
- The BSS assesses the relative skill of the probabilistic forecast over that of a reference forecast in terms of predicting whether an event occurred. A score of 0 indicates no skill when compared to the reference forecast and a score of 1 would be a perfect score. In this article the reference forecasts are the mean or consensus track from the three deterministic forecasts. This is a deliberately challenging reference forecast compared to climatology-based reference forecasts, and the implications of this are discussed in Section 4.

Verification statistics are produced for all tropical cyclones, and then are also split into six global basins to allow for an inter-region comparison, as shown in Figure 2.

3 | RESULTS

3.1 | Overall track probability forecasts for named tropical cyclones

During the verification period of January 2017 to December 2018, 130 named tropical cyclones across the six basins had at least one forecast run where the verification criteria were met and all forecast data were available (24 in NAT, 25 in NEP, 41 in NWP, 7 in NI, 15 in SWI and 18 in AUS, as shown in Figure 2). Overall this two-year period had close to the average number of named tropical cyclones per year, with above-average activity in the North Atlantic basin, and below-average activity in the Southern Hemisphere.

The ROC plot in Figure 3 shows excellent skill for all models; the multi-model ensemble has the largest ROC area (0.985), followed by ECMWF ENS (0.967), MOGREPS-G (0.956) and NCEP GEFS (0.925). All of the models have very low false alarm rates for the majority of forecast probabilities, while the corresponding hit rates vary more significantly across the forecast probabilities. Figure 4 indicates that all models show good reliability, particularly ECMWF ENS and the combined multi-model ensemble. MOGREPS-G and NCEP GEFS are slightly over-confident, with the line falling above the diagonal for low probabilities and below it for higher probabilities. Over-confident reliability diagrams are a common characteristic of ensemble forecasts and are characteristic of an under-spread in the ensemble. So when a high track probability is predicted the ensemble member forecast tracks are sometimes too closely clustered, and the tropical cyclone track probability is less likely than forecast, but when a low probability is forecast the ensemble members are again too closely clustered and so the probability should sometimes be a bit higher.

1477870x, 2020, 726, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3712 by Test, Wiley Online Library on [29/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

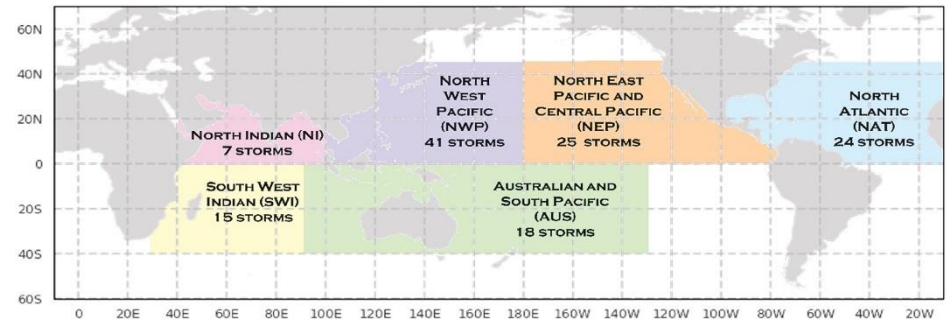


FIGURE 2 The six tropical cyclone basins verified in this study, and the number of named tropical cyclones included in the verification in each basin in the study period (2017 and 2018)

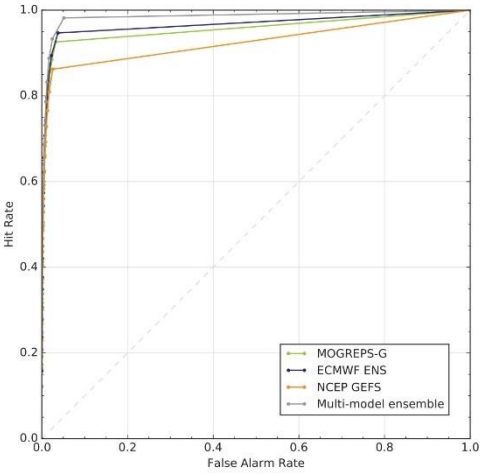


FIGURE 3 ROC plot showing the hit rate and false alarm rate at each probability threshold, for the overall track probability forecasts from the three individual ensembles (MOGREPS-G, ECMWF ENS and NCEP GEFS) and the combined multi-model ensemble

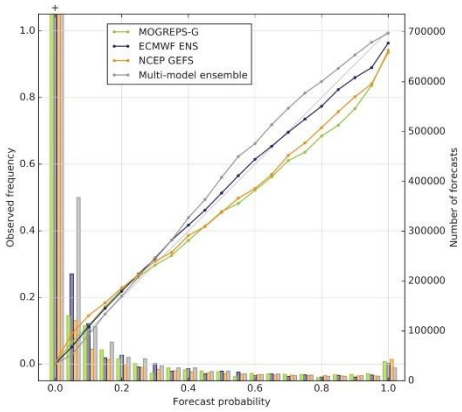


FIGURE 4 Reliability diagram for the overall track probability forecasts for the three individual ensembles (MOGREPS-G, ECMWF ENS and NCEP GEFS) and the combined multi-model ensemble. The number of forecasts (grid points) in each bin is also indicated by bars along the *x*-axis

Figure 5 demonstrates that the combined multi-model ensemble has the greatest relative economic value for all cost-loss ratios, with the multi-model ensemble curve fully encompassing the curves of the three individual models. All models show the greatest relative economic value for low cost-loss ratios (0–0.2) where the loss associated with an event is significantly greater than the cost of acting to mitigate against it when forecast. This is a promising result as although the cost-loss ratios vary for different forecast users, tropical cyclones tend to be associated with low cost-loss ratios due to the severity of their associated hazards and potential impacts.

The BSS, calculated using the consensus of the deterministic models as the reference forecast, is displayed in Figure 6, both for all named tropical cyclones, and split into their relevant basins. Although the absolute and relative performance of the individual ensembles varies from basin to basin, the combined multi-model ensemble has the largest BSS across all basins, showing that the multi-model ensemble adds forecast skill in every basin. The strongest performing individual ensemble is ECMWF ENS in all basins except the Southwest Indian where NCEP GEFS has the largest BSS of the individual ensembles. The relative performance of MOGREPS-G and NCEP GEFS varies between basins.

1477870x, 2020, 726, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3712 by Test, Wiley Online Library on [29/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

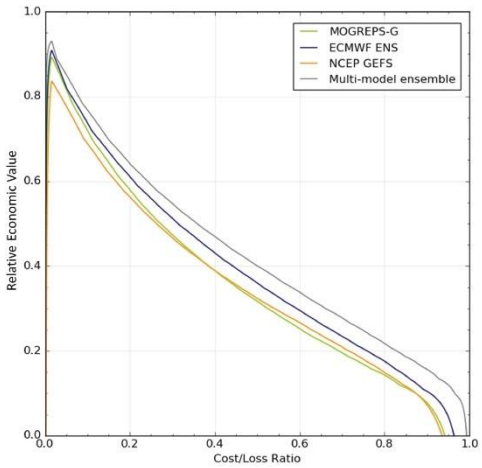


FIGURE 5 Relative economic value plot displaying the relative economic value at each cost-loss ratio for the overall track probability forecasts from the three individual ensembles (MOGREPS-G, ECMWF ENS and NCEP GEFS) and the combined multi-model ensemble

3.2 | Track probability forecasts for named tropical cyclones over each 24-hour forecast period

To illustrate the changes in forecast characteristics and performance against lead time, the reliability diagrams and relative economic value plots shown in Figures 7 and 8 display the 24-hour track probability centred on six lead times from T+24 to T+144. The reliability diagrams (Figure 7) all exhibit some degree of over-confidence in the forecasts, which varies with model and lead time. The lowest probability bin is slightly above the diagonal (as there are some forecasts where the forecasts are confident there will be no tropical cyclone and in fact the tropical cyclone does track that way), but this is not visible by eye due to the very large numbers of correct near-zero probabilities. The other probability bins are below the diagonal, where the forecasts are over-forecasting the tropical cyclone track probability. At T+24 (Figure 7a) there is good reliability for all models, but at the higher probability bins (>0.5), the advantage of the multi-model ensemble is evident compared to the three individual ensembles, which all show a similar degree of over-confidence. As lead time increases, reliability begins to decrease in the NCEP GEFS and MOGREPS-G ensembles, while ECMWF ENS, and especially the multi-model ensemble, maintain very good reliability particularly at probabilities greater than 0.4 at T+72 and T+96 (Figure 7c,d). MOGREPS-G and NCEP GEFS are showing significant over-confidence by T+120 (Figure 7e), indicating that these models are significantly under-dispersive in their five-day track forecasts. At T+120, the multi-model ensemble remains the most reliable at probabilities less than 0.8. This is an important result, as it shows that MOGREPS-G and NCEP GEFS, despite being significantly under-dispersive by this point, are still adding benefit to the ECWMF ENS in the multi-model ensemble at most forecast probabilities. However, at this lead time and in particular at T+144 (Figure 7f)

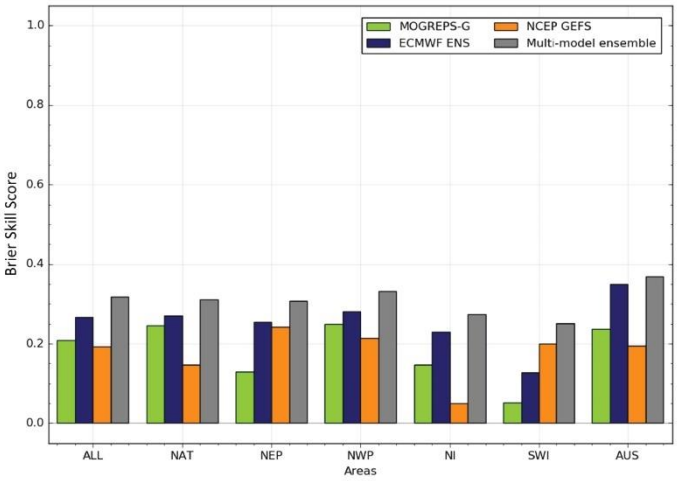


FIGURE 6 BSS for the overall track probability forecasts for all named tropical cyclones (left) and for named tropical cyclones in each of the individual six basins, for each individual ensemble and the multi-model ensemble. The reference forecast used in the skill score calculation is the consensus (mean) track of the three deterministic forecasts

1477870x, 2020, 726, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3712 by Test, Wiley Online Library on [29/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1477870x, 2020, 726, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3712 by Test, Wiley Online Library on [29/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

there are far fewer cases in the higher probability bins, leading to more noisy results.

In Figure 8, nearly all cost-loss ratios (0–0.9) have relative economic value at T+24 (Figure 8a). The peak value is at low cost-loss ratios, with approximately 0.9 relative economic value for a cost-loss ratio of 0–0.1. As lead time increases there is a gradual drop in value, particularly in the higher cost-loss ratios, but even by T+144 significant value still remains at 0–0.1 cost-loss ratios (Figure 8f). At all lead times the multi-model ensemble value curve fully encompasses those for the individual ensembles. Out to T+72 (Figure 8c) the value of the multi-model ensemble is significantly greater than any of the individual ensembles, but by T+120 (Figure 8e) the value for the ECMWF ENS is noticeably greater than the other two ensembles, and closer to the value shown by the multi-model ensemble.

The area under the ROC curve (AUC) values are shown against lead time in Figure 9. The AUC decreases with lead time for all models. The multi-model ensemble maintains an AUC of over 0.93 at all lead times, whereas the AUC score for the individual ensemble decreases more significantly with lead time (in particular for NCEP GEFS). The Brier skill score (BSS) results in Figure 10 show that the skill of the 24-hour track probability forecasts varies with lead time in each basin, with the skill compared to the deterministic consensus forecast rising with increasing lead time. This is as expected, and shows that the benefit of probabilistic forecasts over a consensus of deterministic forecasts becomes greater as the forecast lead time increases. When all tropical cyclones are included, irrespective of the basin (the light green line in Figure 10), the multi-model ensemble probabilistic forecasts show positive skill over the multi-model consensus from T+60 onwards. However, this hides big differences between the basins. In the North Atlantic and North East Pacific basins the BSS is positive throughout and rise to a value of 0.3 by T+156. In the Northwest Pacific basin the BSS is positive from T+48 onwards. However, in the North Indian Ocean, the Southwest Indian Ocean and the Australian basins, the BSS stays negative throughout, showing that in these basins the ensembles are not adding skill to the consensus reference forecast for the 24-hour track probability forecasts in this study period. The negative BSS for these basins in their 24-hour track probability forecasts is in contrast to the positive BSS in their overall 7-day track probability forecasts that was shown in Figure 6, and highlights that the 7-day verification is more forgiving of along-track errors. It is also worth noting that these are also the three basins where the sample size of storms is the smallest. A case-by-case assessment of the tropical cyclones in the worst-performing SWI basin reveals that the deterministic errors were relatively low compared to previous seasons, making the consensus reference forecast hard to beat for this sample of storms.

3.3 | Verification for two high-profile tropical cyclones

Figure 11 compares the BSS for the storm-based verification of two high-profile tropical cyclones in the North Atlantic basin: Hurricane Matthew in 2016 and Hurricane Irma in 2017. For each storm a different individual ensemble displays the highest skill (MOGREPS-G for *Matthew* and ECMWF ENS for *Irma*), illustrating that even within the same basin, the strongest performing individual ensemble varies from storm to storm, rather than one ensemble always being the most skilful in a given basin. In both cases, the multi-model ensemble shows comparable forecast skill to the strongest performing model. Figure 12 shows one of the forecasts included in the storm-based verification in each case, illustrating the strong performance of MOGREPS-G for *Matthew* and ECMWF ENS for *Irma*. For Hurricane Matthew, the MOGREPS-G ensemble was the first model to give a strong signal for the storm to track just off the Florida coast with an eventual landfall in South Carolina. For Hurricane Irma, the ECMWF ensemble was the only ensemble to contain the observed track in the ensemble track spread in the early forecast runs. At the time of the forecast it is not known which of the individual ensemble forecast models will have the greatest forecast skill for that particular storm, and so the key result here is that in both cases the multi-model ensemble probability forecasts were able to provide equivalent forecast skill to the best performing individual ensemble forecasts.

4 | DISCUSSION

In Section 1, five key questions were laid out to be addressed by the probabilistic evaluation of tropical cyclone track probability forecasts in this study. This discussion section is organised to answer these questions, and draw out the key implications from these:

- *How do forecast performance and characteristics vary between the global ensembles, and with lead time?*

All ensembles exhibit good reliability and value in the named tropical cyclone track probability forecasts, particularly at low cost-ratios. This shows how the probability forecasts have huge potential to be useful to decision makers and downstream users of tropical cyclone forecasts, who will often have low cost-loss ratios due to high potential losses to property and personal safety, compared

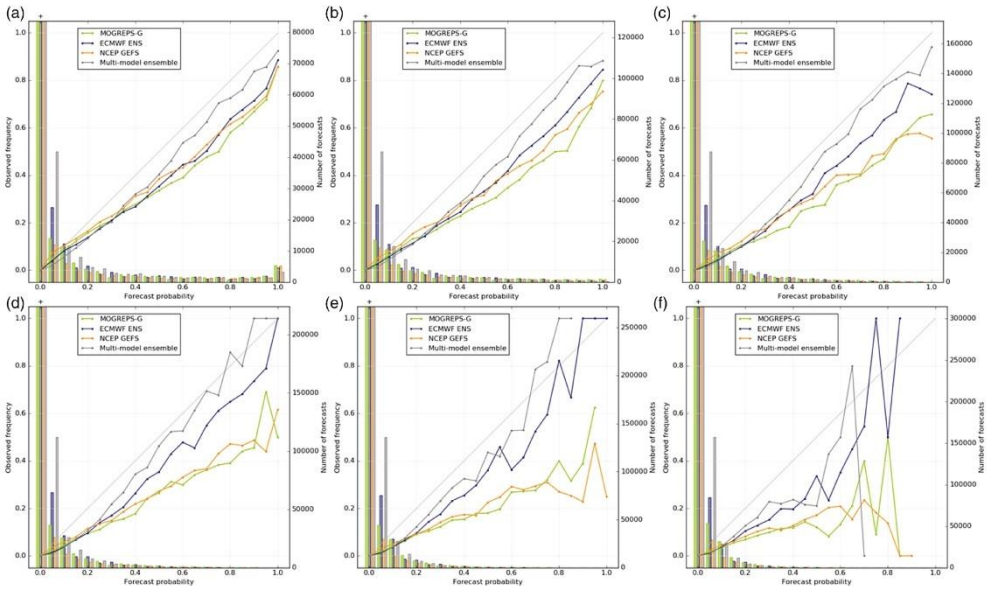


FIGURE 7 Reliability diagrams for 24-hour track probability forecasts for all named storms for each individual ensemble and the multi-model ensemble, for 24-hour periods centred on (a) T+24, (b) T+48, (c) T+72, (d) T+96, (e) T+120 and (f) T+144

to the cost of mitigative efforts such as putting up shutters or evacuating. It highlights the importance of initiatives to increase the pull-through of probabilistic situation-based uncertainty information into operational warnings, such as the collaboration through the WMO HIWeather project that is discussed in Titley *et al.* (2019).

The best performing of the three individual ensembles included in the study, in terms of the verification statistics presented for tropical cyclone track probability forecasts, is the ECMWF ENS, followed by MOGREPS-G and NCEP GEFS, which are both over-confident in their track probability forecasts, indicating the known tendency of these models to be under-dispersive. There are many differences between the three ensemble forecast systems, including differences in data assimilation influencing both the initial conditions of the storm itself and the wider environmental steering flow, the model formulation, and the ensemble perturbation strategies. The relative contribution of perturbations from an ensemble of data assimilations, singular vectors, and stochastic model perturbations to ECMWF ENS track spread was presented in Lang *et al.* (2012). Benefits of the ECMWF ENS perturbation strategy includes the ability to target singular vectors on tropical cyclones, and an enhanced ability to be able to tune perturbations to give improved spread and reliability, compared with the perturbation schemes in the other ensembles.

In the verification of the track probability forecasts for each 24-hour period, all models were shown to be over-confident in their track probability forecasts, but this became much more pronounced at longer lead times for MOGREPS-G and NCEP GEFS, where the higher probabilities were forecast far too often compared to the observed frequency. These results are important to model developers as they show the importance of increasing the ensemble spread in the MOGREPS-G and NCEP GEFS ensembles. In MOGREPS-G, a major upgrade was scheduled to go live in autumn 2019 to try to address this issue. The ensemble perturbation system will be changed from Ensemble Transform Kalman Filter (ETKF) to an ensemble of data assimilations (En-4D-En-Var: Bowler *et al.*, 2017). In the new system, data assimilation is performed for each member, creating increments relative to its own background trajectory. A partial re-centring around the deterministic analysis gives an additional increase in skill and reduces jumpiness. Comparative trials of the new En-4D-En-Var ensemble have shown faster spread growth across many variables including 850 hPa wind speed in the tropics, with a much better match to observed errors. The ensemble trials have also been processed through

1477870x, 2020, 726, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3712 by Test, Wiley Online Library on [29/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1477870x, 2020, 726, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3712 by Test, Wiley Online Library on [29/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

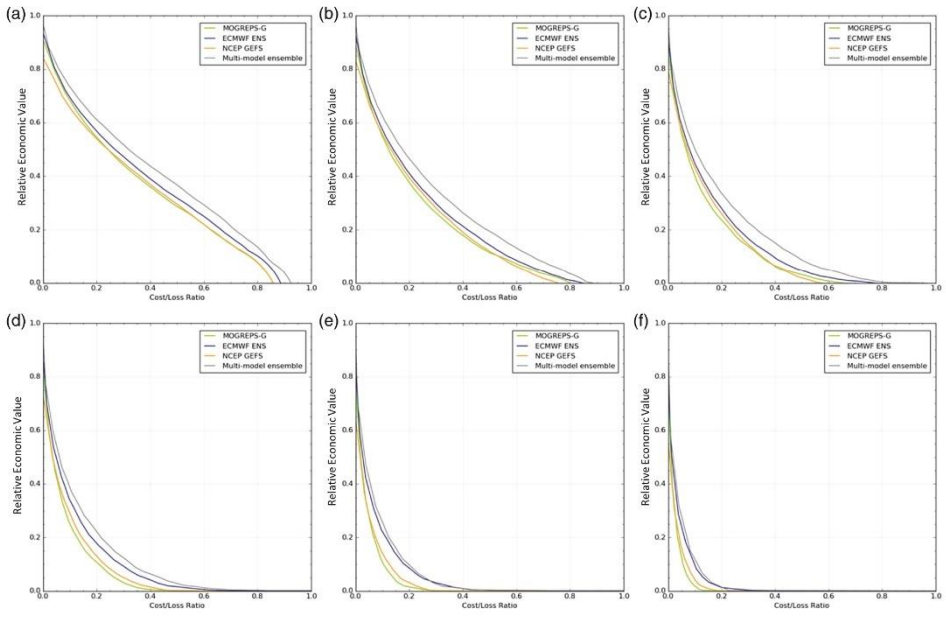


FIGURE 8 Relative economic value curves for 24-hour track probability forecasts for all named storms for each individual ensemble and the multi-model ensemble, for 24-hour periods centred on (a) T+24, (b) T+48, (c) T+72, (d) T+96, (e) T+120 and (f) T+144

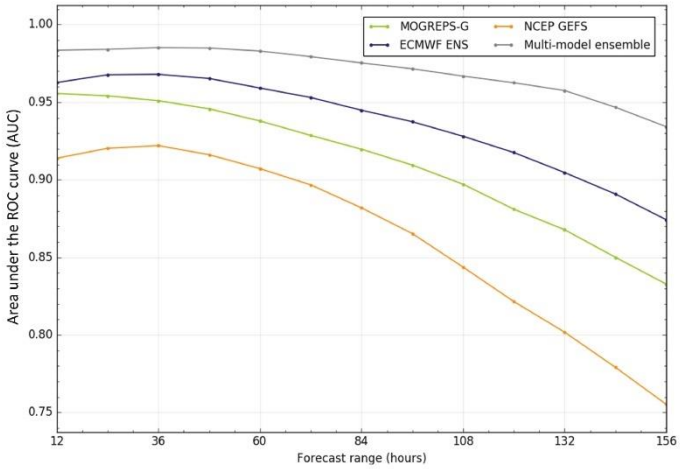


FIGURE 9 The AUC for each individual ensemble and the multi-model ensemble, for the 24-hour track probability forecasts centred on each forecast range

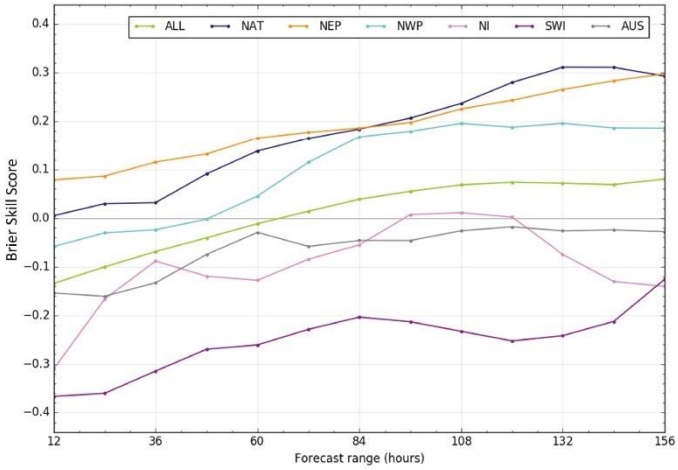


FIGURE 10 BSS for the multi-model ensemble, for all storms (in green) and split into the six tropical cyclone basins, for the 24-hour track probability forecasts centred around each forecast range

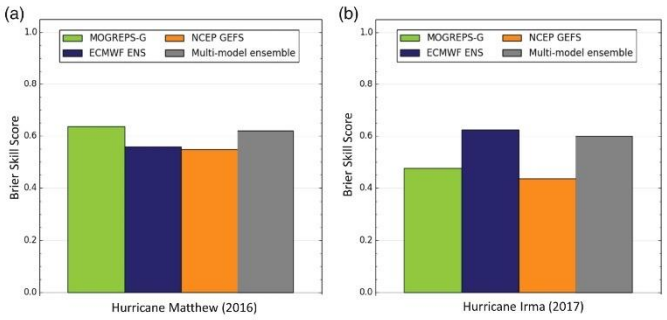


FIGURE 11 Example storm-specific verification for all forecast runs of (a) Hurricane Matthew (2016) and (b) Hurricane Irma (2017): Brier skill score for MOGREPS-G, ECMWF ENS, NCEP GEFS and multi-model ensemble forecasts (reference forecast in this verification is the sample climatology)

the tropical cyclone tracking and post-processing system, and show a significant improvement in track spread at all lead times. Meanwhile at NCEP a new version of the deterministic Global Forecast System (GFS) with a new Finite-Volume Cubed-Sphere Dynamical Core (FV3) went operational in June 2019. The potential of this new model to improve hurricane forecast performance was described by Chen *et al.* (2019), who showed the improved performance in re-runs of the active 2017 Atlantic hurricane season. The FV3 dynamical core will be implemented into the GEFS ensemble in 2020, in an upgrade that will also see improved stochastic physics and an increase in ensemble members. The ECMWF ENS also continues to be upgraded, with IFS Cycle 46r1 implemented in June 2019, including a data assimilation upgrade which improves the initial conditions of the ensemble forecasts. Additional work to improve the tropical cyclone intensity forecasts in the ECMWF ENS is described in Magnusson *et al.* (2019). The track probability verification described in this article will continue to run every 6 months to verify the three ensemble models and their combined multi-model ensemble to investigate the impact of these model upgrades on the skill of tropical cyclone track probability forecasts.

- *Is there benefit in terms of probabilistic forecast skill, reliability and value, from combining the three global ensembles into a multi-model ensemble?*

As established in the introduction, the rationale of multi-model ensemble forecasting lies in combining independent and skilful models, each with its own strengths and weaknesses. The evaluation in this article clearly shows that additional forecast skill and value can be gained from combining the members from the three individual ensembles included in this study into a multi-model ensemble. The three ensembles have different data assimilation strategies, model formulations and ensemble perturbation schemes, that when combined together are

1477870x, 2020, 726, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3712 by Test, Wiley Online Library on [29/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

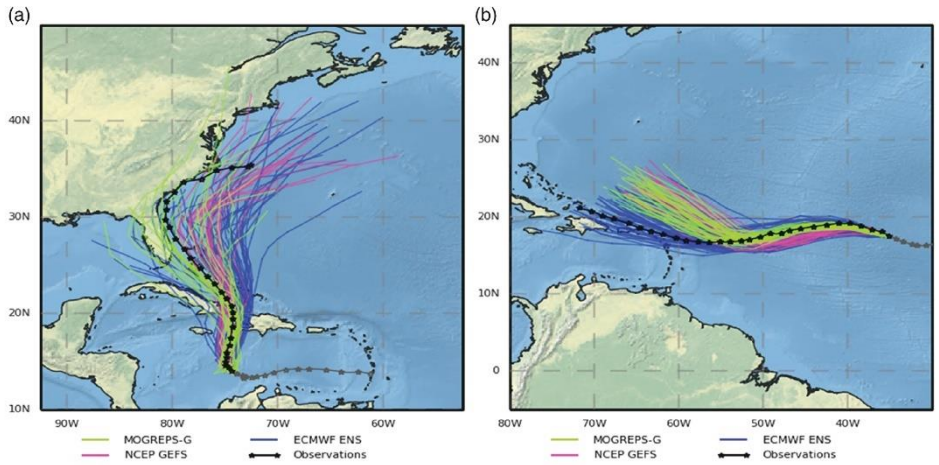


FIGURE 12 Ensemble member tracks from one of the forecast runs included in the verification of each case: (a) 1200 UTC 2 October 2016 forecast for Hurricane Matthew; (b) 0000 UTC 1 September 2017 forecast for Hurricane Irma. Tracks are coloured according to model (green = MOGREPS-G, blue = ECMWF ENS and pink = NCEP GEFS)

shown to collectively provide more realistic estimates of tropical cyclone track probabilities. In the overall track probability verification, the relative economic value curve of the multi-model ensemble fully encompasses that of the individual ensembles, and in the verification for each 24-hour period it is particularly noteworthy that the AUC decreases more significantly with lead time in the individual ensembles compared to the multi-model ensemble. This illustrates that the amount of potential forecast skill to be gained from using a multi-model ensemble to derive your track probabilities increases significantly with forecast lead time. This finding is important for operational tropical cyclone forecasters, forecast centre managers, and numerical weather prediction centres, showing the importance of improving access to multiple ensemble forecast model forecasts in order to allow multi-model ensemble information to be used in the operational tropical cyclone forecasting process.

- *Does using the full probability forecast information via the multi-model ensemble add skill compared to a "consensus" forecast of the parent deterministic models?*

The choice of reference forecast is crucial in the Brier Skill Score (BSS) calculation, as the BSS measures the improvement of the probabilistic forecast relative to a reference forecast. This forecast reference is often calculated from the sample climatology, or, in the case of tropical cyclone forecasts, from a combined climatology and persistence forecast (CLIPER: Knaff *et al.*, 2003). However, in order to be more relevant for the current common forecasting practice of establishing a consensus forecast, and in order to provide a more skilful and challenging reference for the ensemble probabilities to be compared to, a consensus forecast based on the mean track of the three parent deterministic models was created and used as the reference in the BSS calculation. The BSS for the overall 7-day track probability skill scores shows positive skill for all of the ensemble forecast models in all basins when compared to the reference consensus track, with the BSS being highest for the combined multi-model ensemble. This shows the added benefit to be gained in operational tropical cyclone track forecasting if the full probability information provided by the ensembles can be pulled through into operational warnings, as has begun to happen with the incorporation of dynamic uncertainty information in warning products from several centres including RSMC La Réunion and RSMC Tokyo (Titley *et al.*, 2019). When the track probability verification is split into 24-hour periods it is harder to beat the consensus reference, particularly at short lead times, but the BSS increases with forecast range, showing the increasing value of ensemble prediction over consensus forecasting with lead time. Overall there is positive forecast skill from T+60 onwards. The BSS was also calculated for the 24hour track probability verification using the sample climatology (not shown) and showed the reverse pattern, with the highest BSS (0.6) at shorter lead times, slowly decreasing with forecast range. This is to be expected since the sample climatology reference has no forecast capability, whereas the consensus has a very

1477870x, 2020, 726, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3712 by Test, Wiley Online Library on [29/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

strong forecast capability at short range which reduces with increasing lead time. This illustrates how important it is to be clear about the implications and rationale of selecting a particular reference forecast.

- *When comparing basin to basin, does forecast performance vary, both overall compared to a reference consensus forecast, and relatively between the different ensembles?*

In the overall track probability forecasts BSS results (Figure 6), the combined multi-model ensemble has the largest BSS of all the models, and varies between 0.25 and 0.4 with each basin, showing that the multi-model ensemble is adding skill in every basin. The strongest performing individual ensemble is ECMWF ENS in all basins except the SWI basin where NCEP GEFS has the highest skill. The relative performance of MOGREPS-G and NCEP GEFS varies between basins, with MOGREPS-G performing well in the NAT, NWP and AUS basins, and NCEP GEFS performing well in the SWI and NEP basins. This result would provide useful guidance to operational tropical cyclone forecasters in these regions, as it provides a breakdown of the current levels of performance of each model in their area of interest, and confirms that all areas would see additional value in computing track probability forecasts from multi-model ensemble forecast data. It is also of interest to model developers, who could investigate why their model performs better relatively for tropical cyclones in one basin over another, leading to potential ideas on how to improve forecast skill. As discussed earlier, each model has different perturbation strategies and data assimilation schemes, with differences in which observations are included (including differences in quality-control systems and in the assimilation of RSMC tropical cyclone observations, known as bogussing, which is not currently carried out at ECMWF). These results show that some aspects of each model system may be better suited to one region over another, and further investigations including the implications of a tropical cyclone being in a data-rich vs. data-poor area may lead to further insights and improvements in the model forecasts.

The inter-basin comparison of BSS using the 24-hour track probability forecasts centred on each lead time (Figure 10) reveals large differences from basin to basin in the skill compared to the consensus reference forecast, with positive skill in the NAT, NWP and NEP, but not in the NI, SWI or AUS basins. This emphasises that the traditional deterministic consensus is hard to beat in regions where the tracks of the named storms in the study period are well forecast by the deterministic models. The lower BSS for the 24-hour track probability forecasts highlights the need to focus on improving the ensemble forecasts for the translation speed of the storm, as the results when splitting the track probability forecasts in to 24-hour periods will be impacted by along-track errors in addition to the cross-track errors, whereas the cross-track errors are the most influential factor in the overall 7-day track probability results. Although forecast users predominantly need to know whether or not they will be impacted by a hurricane rather than when, the timings are also important in forecast preparedness activities, and future verification could investigate the relative along- and cross-track errors in the ensemble vs. the deterministic consensus.

- *When comparing storm to storm, how does forecast performance vary, and what does it tell us about the benefit of multi-model ensembles?*

Although when averaged across all tropical cyclones the ECMWF ENS is clearly currently the most skilful of the individual ensembles, the case-study analysis in Figure 11 showed that in a particular case this will not always hold true. Sometimes a different model is more skilful in a particular storm (as in the case for MOGREPS-G for Hurricane Matthew). At the time of the live forecasts, operational forecasters will not know which individual model is destined to be the most skilful for that storm, and therefore the result showing that the multi-model ensemble skill matches that of the strongest performing individual ensemble is an important result, and clearly illustrates the benefit of combining the ensemble members into a multi-model ensemble when computing track probability forecasts.

5 | CONCLUSIONS AND FUTURE PLANS

This study has shown that combined multi-model ensemble tropical cyclone track probability forecasts, calculated from all members of ECMWF ENS, MOGREPS-G and NCEP GEFS, have increased skill and value over the best-performing individual ensemble. This result is consistent when verifying all global named tropical cyclones together, and when the verification is carried out for each individual basin. The improved skill and value of the multi-model ensemble is found for both the full (up to 7-day) track probability forecasts, and for track probabilities split into 24-hour forecast periods. The verification results from the three individual ensembles show that the track probability forecasts from ECMWF ENS display the best reliability, skill and value. MOGREPS-G and NCEP GEFS become increasingly over-confident and under-dispersive with increasing forecast range, emphasising the importance of ongoing work at both centres to improve the perturbation strategy and increase the spread in the ensemble forecasts. However, even when there is an individual ensemble model that on average performs

1477870x, 2020, 726, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3712 by Test, Wiley Online Library on [29/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1477870x, 2020, 726, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3712 by Test, Wiley Online Library on [29/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

better in a particular storm, basin, or overall, there is still skill, reliability and value to be gained from adding additional ensemble models into a combined multi-model ensemble, indicating that they have independent systematic strengths and errors and collectively provide more realistic estimates of tropical cyclone track probabilities than individually.

Storm-based verification illustrated that the best-performing individual ensemble can change from case to case, but that the multi-model ensemble matches the best-performing individual ensemble, which would not be known in advance, in terms of overall probabilistic forecast skill. The mean, or consensus, of the three higher-resolution deterministic forecasts is hard to beat in some basins, but overall the additional probabilistic skill of the ensembles is shown, particularly at longer lead times or when computing the full 7-day track probability forecasts. This study helps to illustrate the potential value and skill to be gained if operational tropical cyclone forecasting can continue to migrate away from a deterministic-focussed forecasting environment to one where the probabilistic situation-based uncertainty information provided by the ensembles can be pulled through into operational forecasts and warnings.

There are many ideas for where this work could be further applied and extended in the future. In this study no weighting is applied to the ensemble members in the multi-model ensemble, with each member from each model given an equal weight. It would be interesting to look at different options for combining together the ensemble members, and also investigate how much of the additional value is from additional members as opposed to the inclusion of members from other models. For example, would the multi-model ensemble still add value to the ECMWF ENS if it were restricted to the same number of members? Additional global ensemble forecast models, including those available in the TIGGE cyclone CXML archive from the Japan Meteorological Agency, the Canadian Meteorological Centre and Météo-France, could be added to the multi-model ensemble, to investigate the optimal combination of ensembles in track probability forecasts. At the Met Office the ensemble tropical cyclone products and verification will continue to be used to evaluate important model upgrades and trials, such as the impact of the forthcoming move from ETKF to Ens-4D-En-Var perturbations. The products and verification are also run on high-resolution convective-permitting ensembles in Southeast Asia. Work is also underway to extend the ensemble tropical cyclone verification capability at the Met Office to incorporate a verification of forming storms (tropical cyclone genesis) and intensity trends. It is also important to look beyond the traditional tropical cyclone track and intensity forecasts and move towards verifying the associated hazards, for example to assess how the uncertainty and predictability of the track translates through to uncertainty and predictability of the precipitation, and the downstream flood hazard. Ongoing collaborations between global numerical weather prediction centres, researchers and operational forecasting centres continue to be essential to ensure that future research and ensemble model developments are of maximum benefit to operational tropical cyclone forecasting.

ACKNOWLEDGEMENTS

The authors would like to thank three reviewers for their helpful comments and suggestions that have led to improvements to the article, Julian Heming for his support with the use of MOTCTracker and for providing the observations for use in the verification, and Ken Mylne, Philip Gill and Rutger Dankers for their support during this work. This work was partially supported by the Met Office Weather and Climate Science for Service Partnership (WCSSP) India Programme as part of the Newton–Bhabha Fund.

ORCID

Helen A. Titley https://orcid.org/0000-0003-1654-9826
Rebecca L. Bowyer https://orcid.org/0000-0001-7584-4425
Hannah L. Cloke https://orcid.org/0000-0002-1472-868X

REFERENCES

Bowler, N.E., Arribas, A., Mylne, K.R., Robertson, K.B. and Beare, S.E. (2008) The MOGREPS short-range ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 134, 703–722.

Bowler, N.E., Clayton, A.M., Jardak, M., Jermey, P.M., Lorenc, A.C., Wlasak, M.A., Barker, D.M., Inverarity, G.W. and Swinbank, R. (2017) The effect of improved ensemble covariances on hybrid variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143(703), 785–797.

Chen, J.H., Lin, S.J., Magnusson, L., Bender, M., Chen, X., Zhou, L., Xiang, B., Rees, S., Morin, M. and Harris, L. (2019) Advancements in hurricane prediction with NOAA's next-generation forecast system. *Geophysical Research Letters*, 46, 4495–4501.

Du, Y.G., Qi, L.B. and Cao, X.G. (2016) Selective ensemble-mean technique for tropical cyclone track forecast by using time-lagged ensemble and multi-centre ensemble in the western North Pacific. *Quarterly Journal of the Royal Meteorological Society*, 142, 2452–2462.

Dupont, T., Plu, M., Caroff, P. and Faure, G. (2011) Verification of ensemble-based uncertainty circles around tropical cyclone track forecasts. *Weather and Forecasting*, 26, 664–676.

Goerss, J.S. (2000) Tropical cyclone track forecasts using an ensemble of dynamical models. *Monthly Weather Review*, 128, 1187–1193.

Goerss, J.S. (2007) Prediction of consensus tropical cyclone track forecast error. *Monthly Weather Review*, 135, 1985–1993.

Goerss, J.S. and Sampson, C.R. (2014) Prediction of consensus tropical cyclone intensity forecast error. *Weather and Forecasting*, 29, 750–762.

Hagedorn, R., Buizza, R., Hamill, T.N., Leutbecher, M. and Palmer, T.N. (2012) Comparing TIGGE multi-model forecasts with reforecast-calibrated ECMWF ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 138, 1814–1827.

Hagedorn, R., Doblas-Reyes, F.J. and Palmer, T.N. (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting. Part I: Basic concept. *Tellus*, 57A, 219–233.

Hamill, T.M. (2012) Verification of TIGGE multimodel and ECMWF reforecast-calibrated probabilistic precipitation forecasts over the contiguous United States. *Monthly Weather Review*, 140, 2232–2252.

Heming, J.T. (2017) Tropical cyclone tracking and verification techniques for Met Office numerical weather prediction models. *Meteorological Applications*, 24, 1–8.

Hodges, K.I. (1995) Feature tracking on the unit sphere. *Monthly Weather Review*, 123, 3458–3465.

Johnson, C. and Swinbank, R. (2009) Medium-range multimodel ensemble combination and calibration. *Quarterly Journal of the Royal Meteorological Society*, 135, 777–794.

Jolliffe, I.T. and Stephenson, D.B. (2012) *Forecast Verification: A practitioner's guide in atmospheric science*, 2nd edition. Chichester: Wiley and Sons Ltd.

Knaff, J.A., DeMaria, M., Sampson, C.R. and Gross, J.M. (2003) Statistical, 5-day tropical cyclone intensity forecasts derived from climatology and persistence. *Weather and Forecasting*, 18, 80–92.

Lang, S.T.K., Leutbecher, M. and Jones, S.C. (2012) Impact of perturbation methods in the ECMWF ensemble prediction system on tropical cyclone forecasts. *Quarterly Journal of the Royal Meteorological Society*, 138, 2030–2046.

Leonardo, N.M. and Colle, B.A. (2017) Verification of multimodel ensemble forecasts of North Atlantic tropical cyclones. *Weather and Forecasting*, 32, 2083–2101.

Magnusson, L., Bidlot, J.R., Bonavita, M., Brown, A.R., Browne, P.A., De Chiara, G., Dahoui, M., Lang, S.T.K., McNally, T., Mogensen, K.S., Pappenberger, F., Prates, F., Rabier, F., Richardson, D.S., Vitart, F. and Malardel, S. (2019) ECMWF activities for improved hurricane forecasts. *Bulletin of the American Meteorological Society*, 100, 445–458.

Majumdar, S.J. and Finocchio, P.M. (2010) On the ability of global ensemble prediction systems to predict tropical cyclone track probabilities. *Weather and Forecasting*, 25, 659–680.

Matsueda, M. and Nakazawa, T. (2015) Early warning products for severe weather events derived from operational medium-range ensemble forecasts. *Meteorological Applications*, 22, 213–222.

Palmer, T.N. (2019) The ECMWF ensemble prediction system: looking back (more than) 25 years and projecting forward 25 years. *Quarterly Journal of the Royal Meteorological Society*, 145(S1), 12–24. https://doi.org/10/1002/qj.3383.

Park, Y.Y., Buizza, R. and Leutbecher, M. (2008) TIGGE: preliminary results on comparing and combining ensembles. *Quarterly Journal of the Royal Meteorological Society*, 134, 2029–2050.

Richardson, D.S. (2000) Skill and relative economic value of the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 126, 649–667.

Robbins, J.C. and Titley, H.A. (2018) Evaluating high-impact precipitation forecasts from the Met Office Global Hazard Map (GHM) using a global impact database. *Meteorological Applications*, 25, 548–560.

Sampson, C.R., Franklin, J.L., Knaff, J.A. and DeMaria, M. (2008) Experiments with a simple tropical cyclone intensity consensus. *Weather and Forecasting*, 23, 304–312.

Swinbank, R., Kyouda, M., Buchanan, P., Froude, L., Hamill, T.M., Hewson, T.D., Keller, J.H., Matsueda, M., Methven, J., Pappenberger, F., Scheuerer, M., Titley, H.A., Wilson, L. and Yamaguchi, M. (2016) The TIGGE project and its achievements. *Bulletin of the American Meteorological Society*, 97, 49–67.

Tallapragada, V., Bernardet, L., Gopalakrishnan, S., Kwon, Y., Liu, Q., Marchok, T., Sheinin, D., Tong, M., Trahan, S., Tuleya, R., Yablonsky, R. and Zhang, X. (2013) *Hurricane Weather Research and Forecasting (HWRF) model: 2013 scientific documentation*. HWRF Development Testbed Center, Technical Report 99. Available at: http://www.dtcenter.org/HurrWRF/users/docs/scientific_documents/HWRFv3.5a_ScientificDoc.pdf [Accessed 4th December 2019].

Titley, H.A., Yamaguchi, M. and Magnusson, L. (2019) Current and potential use of ensemble forecasts in operational TC forecasting: results from a global forecaster survey. *Tropical Cyclone Research and Review*, 8, 166–180.

Toth, Z. and Kalnay, E. (1997) Ensemble forecasting at NCEP and the breeding method. *Monthly Weather Review*, 125, 3297–3318.

Van der Grijn, G. (2002) *Tropical Cyclone Forecasting at ECMWF: New products and validation*. ECMWF Technical Memorandum, Vol. 386. Reading, UK: ECMWF, 1–13.

Van der Grijn, G., Paulsen, J.E., Lalaurette, F. and Leutbecher, M. (2004) *Early medium-range forecasts of tropical cyclones*. ECMWF Newsletter, No. 102, Reading, UK, 7–14. Available at: https://www.ecmwf.int/file/28462/download?token=i0qbDD6S.

Vitart, F. and Stockdale, T.N. (2001) Seasonal forecasting of tropical storms using coupled GCM integrations. *Monthly Weather Review*, 129, 2521–2537.

Wilks, D.S. (2011) *Statistical Methods in the Atmospheric Sciences. International Geophysics Series, 100*, 3rd edition. San Diego, CA: Elsevier, Academic Press.

Yamaguchi, M., Ishida, J., Sato, H. and Nakagawa, M. (2017) WGNE intercomparison of tropical cyclone forecasts by operational NWP models: a quarter century and beyond. *Bulletin of the American Meteorological Society*, 98, 2337–2349.

Yamaguchi, M. and Koide, N. (2017) Tropical cyclone genesis guidance using the early stage Dvorak analysis and global ensembles. *Weather and Forecasting*, 32, 2133–2141.

Yamaguchi, M., Nakazawa, T. and Hoshino, S. (2012) On the relative benefits of a multi-centre grand ensemble for tropical cyclone track prediction in the western North Pacific. *Quarterly Journal of the Royal Meteorological Society*, 138, 2019–2029.

Yamaguchi, M., Vitart, F., Lang, S.T., Magnusson, L., Elsberry, R.L., Elliott, G., Kyouda, M. and Nakazawa, T. (2015) Global distribution of the skill of tropical cyclone activity forecasts on short- to medium-range time scales. *Weather and Forecasting*, 30, 1695–1709.

Zhang, X.P. and Yu, H. (2017) A probabilistic tropical cyclone track forecast scheme based on the selective consensus of ensemble prediction systems. *Weather and Forecasting*, 32, 2143–2157.

How to cite this article: Titley HA, Bowyer RL, Cloke HL. A global evaluation of multi-model ensemble tropical cyclone track probability forecasts. *Q J R Meteorol Soc*. 2020;146:531–545. https://doi.org/10.1002/qj.3712

1477870x, 2020, 726, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3712 by Test, Wiley Online Library on [29/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

A3: Published Article: Key Factors Influencing the Severity of Fluvial Flood Hazard from Tropical Cyclones

This appendix contains the published version of chapter 5 of this thesis, with the following reference:

Titley, H.A., Cloke, H.L., Harrigan, S., Pappenberger, F., Prudhomme, C., Robbins, J.C., Stephens, E.M. and Zsótér, E. (2021) Key factors influencing the severity of fluvial flood hazard from tropical cyclones. *Journal of Hydrometeorology*, 22 (7): 1801–1817. doi:10.1175/JHM-D-20-0250.1.

Key Factors Influencing the Severity of Fluvial Flood Hazard from Tropical Cyclones

H. A. TITLEY,[a,b] H. L. CLOKE,[b,c,d] S. HARRIGAN,[e] F. PAPPENBERGER,[e] C. PRUDHOMME,[e,f,g] J. C. ROBBINS,[a] E. M. STEPHENS,[b,h] AND E. ZSÓTÉR[b,e]

[a] *Met Office, Exeter, United Kingdom*
[b] *University of Reading, Reading, Berkshire, United Kingdom*
[c] *Uppsala University, Uppsala, Sweden*
[d] *Centre for Natural Hazards and Disaster Science, Uppsala, Sweden*
[e] *ECMWF, Reading, Berkshire, United Kingdom*
[f] *University of Loughborough, Loughborough, United Kingdom*
[g] *U.K. Centre for Ecology and Hydrology, Wallingford, United Kingdom*
[h] *Red Cross Red Crescent Climate Centre, The Hague, Netherlands*

(Manuscript received 20 October 2020, in final form 12 March 2021)

ABSTRACT: Knowledge of the key drivers of the severity of river flooding from tropical cyclones (TCs) is vital for emergency preparedness and disaster risk reduction activities. This global study examines landfalling TCs in the decade from 2010 to 2019 to identify those characteristics that influence whether a storm has an increased flood hazard. The highest positive correlations are found between flood severity and the total precipitation associated with the TC. Significant negative correlations are found between flood severity and the translation speed of the TC, indicating that slower-moving storms that rain over an area for longer tend to have higher flood severity. Larger and more intense TCs increase the likelihood of having a larger area affected by severe flooding but not its duration or magnitude, and it is found that the fluvial flood hazard can be severe in all intensity categories of TC, including those of tropical storm strength. Catchment characteristics such as antecedent soil moisture and slope also play a role in modulating flood severity, and severe flooding is more likely in cases in which multiple drivers are present. The improved knowledge of the key drivers of fluvial flooding in TCs can help to inform research priorities to help with flood early warning, such as increasing the focus on translation speed in model evaluation and impact-based forecasting.

SIGNIFICANCE STATEMENT: Knowing ahead of landfall which TCs are likely to lead to significant river flooding will help those responsible for emergency planning make appropriate decisions to minimize loss of life and property. We compare 280 TCs and find that the cases with slow-moving, large, and intense cyclones, affecting areas with wet antecedent conditions, have the highest likelihood of experiencing widespread flooding. Slower-moving storms also have an increased risk of longer and more extreme floods. Our results show the importance of considering aspects such as the speed of forward movement along the whole flood early warning chain, from model evaluation and development, through to warning design and communication, to better inform forecast-based action prior to TC landfall.

KEYWORDS: Flood events; Hurricanes/typhoons; Precipitation; Tropical cyclones; Communications/decision making; Emergency preparedness; Emergency response

1. Introduction

Considering fluvial flood hazards in tropical cyclone (TC) forecasting and warning is important because this is a leading cause of mortality and damages (Rezapour and Baldock 2014). In the United States, drowning from excessive rainfall occurs in more TCs than deaths from any other cause (Rappaport 2014). Many of these fatalities occur outside of landfall counties (Czajkowski and Kennedy 2010) and in inland counties (Rappaport 2000). The U.S. residential losses from TC freshwater flooding are 2 times the losses from TC storm surge, with nearly half of these being in inland areas (Czajkowski et al. 2017). A multihazard approach considering both wind speed and rainfall has been shown to be more appropriate for risk-informed decision-making (Song et al. 2020), but studies investigating evacuation decision-making during hurricanes in the United States have shown that the key determining factor is the intensity of the storm on the Saffir–Simpson scale based on wind speed (Whitehead et al. 2000), with no significant relationship between the perceived risk of flooding and evacuation (Stein et al. 2010). Therefore, it is important to increase public awareness of the dangers of inland flooding, and provide a better understanding of those factors that influence the severity of flood hazard to those involved in emergency preparedness and disaster risk reduction activities.

Heavy rainfall can present a risk to communities regardless of the storm's intensity category. The U.S. National Hurricane Center (NHC) and the U.S. National Weather Service (NWS) both indicate the importance of other information beyond the storm's intensity-based category. The NHC states that "Rainfall amounts are not directly related to the strength of

Denotes content that is immediately available upon publication as open access.

Corresponding author: Helen Titley, helen.titley@metoffice.gov.uk

DOI: 10.1175/JHM-D-20-0250.1

© 2021 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Unauthenticated | Downloaded 06/29/23 02:14 PM UTC

TCs but rather to the speed and size of the storm, as well as the geography of the area'' (https://www.nhc.noaa.gov/prepare/hazards.php#rain). The NWS states that "It is common to think the stronger the storm the greater the potential for flooding. However, this is not always the case. A weak, slow-moving tropical storm can cause more damage due to flooding than a more powerful fast-moving hurricane'' (https://www.weather.gov/jetstream/tc_hazards). The total rainfall from a TC in a given location is largely determined by the length of time a TC spends over that location, which is dependent on the size of the rainfall area and translational speed of the storm (Rogers et al. 2009). A slow along-track motion, or stalling near or after landfall, can lead to higher amounts of rainfall and a greater flood hazard, as seen during Hurricanes Harvey (2017) and Florence (2018). Larger TCs can also increase the rainfall hazard as they precipitate upon one spot for a longer time than smaller TCs moving at the same speed. They can also lead to more widespread flooding, leading to increased challenges for responders. Forecasting rainfall induced by TC landfall is determined by many factors, including the TC track, intensity, size and structure, as well as interaction with topography (Qiu et al. 2019), vertical wind shear (Chen et al. 2006), and other meteorological systems in the wider atmospheric environment. This can lead to seemingly similar landfall locations having different rainfall distributions (Cheung et al. 2018; Shi et al. 2017). Hydrological factors, such as soil conditions and orography, are also thought to have been important in determining which TC cases had elevated impacts from flooding (Rappaport 2000; Sturdevant-Rees et al. 2001).

While there is broad understanding, largely through case study or regional analysis (Saha et al. 2015; Hernández Ayala and Matyas 2016; Touma et al. 2019; Yu et al. 2017), of those factors influencing the rainfall related to TCs, systematic global analysis to objectively confirm these drivers and their relationships to downstream flood hazard is lacking. This study aims to address this gap and help provide more specific information in support of efforts to raise awareness of fluvial flood risk from TCs, by undertaking a systematic global analysis of the key meteorological and hydrological factors that lead to an increased fluvial flood hazard from TCs.

An improved understanding of the key factors that influence the severity of flood hazard from TCs is important to guide research work aiming to understand the predictability of fluvial flooding from TCs, to inform research priorities to improve the forecasts of flooding, and to guide planning and preparedness activities in the event of an advancing TC. Information on which TCs are likely to have storm surge and winds as their main hazard and which will also have substantial flood hazard is of vital importance for disaster risk reduction. While the strongest winds and largest storm surge usually occur near the center of intense TCs, rainfall and flooding often occur far from the center, spread far inland, and last beyond when the cyclone has weakened or dissipated (Villarini et al. 2014; Khouakhi et al. 2017). This has important consequences for evacuation, emergency management and recovery programs. If planners and responders better understand the likely severity of fluvial flooding for a given storm, and the locations likely to be at greatest risk from flooding, this can help with evacuation and emergency planning, response and recovery efforts.

This study develops three indices to represent the severity of fluvial flooding for TC cases in terms of the flood area, duration, and magnitude. These three flood severity indices are compared with TC and catchment characteristics in 280 landfalling TCs in the 10-yr period from 2010 to 2019 to identify the key factors that influence the severity of river flooding in TCs.

The remainder of this paper is set out as follows: section 2 details the data sources used in the study, and section 3 describes the methods used to select the TC cases, and calculate the TC characteristics, catchment characteristics and flood severity metrics for each case. Section 4 then compares the flood severity metrics for each storm with the TC and catchment characteristics to analyze the key factors that influence the severity of flood hazard from TCs. Section 5 discusses some important aspects of the results and where future work would be beneficial, and section 6 contains the main conclusions of the study.

2. Data sources

a. *IBTrACS*

The main observed TC track dataset used in this study is the International Best Track Archive for Climate Stewardship (IBTrACS) (Knapp et al. 2010, 2018). Where IBTrACS data are still listed as provisional (the 2018/19 and 2019/20 Southern Hemisphere seasons, and the 2019 season in the North Indian and North West Pacific basins), the track points are supplemented by the initial positions in the real-time advisories from each Regional Specialized Meteorological Centre (RSMC), which are collated at the Met Office for use in verifying TC track forecasts (Heming 2017; Titley et al. 2020).

The IBTrACS data are used to calculate the land footprint for each TC case (section 3a), over which to calculate the flood severity indices. The IBTrACS data also provide the TC intensity data [section 3b(1)], and are used to calculate the translation speed [section 3b(2)].

b. *Global Precipitation Measurement (GPM) IMERG*

Precipitation data are taken from the latest Integrated Multisatellite Retrievals for GPM (GPM IMERG Final Run V06) (Huffman et al. 2019). The "Final Run'' data are used because they include gauge data where available to calibrate the rainfall satellite-derived observations and have been shown to capture TC precipitation patterns well, with closer agreement with gauge-based measurements than their predecessor for extreme events (Yuan et al. 2021). The data have a horizontal resolution of 0.1° and are extracted at a 30-min temporal resolution and then accumulated to give 24-h precipitation accumulation data. The precipitation summary metrics are described in section 3b(4).

c. *ERA5*

ERA5 is the European Centre for Medium-Range Weather Forecasts (ECMWF)'s latest comprehensive atmospheric reanalysis with a global horizontal grid resolution of approximately 31 km, or 0.281 25° (Hersbach et al. 2020). ERA5 mean

Unauthenticated | Downloaded 06/29/23 02:14 PM UTC

sea level pressure (MSLP) data were extracted from the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) (Hersbach et al. 2018a) at 6-hourly intervals and used to calculate the size of the TC [section 3b(3)]. Daily ERA5 soil moisture content data were also extracted, along with soil type, for use in calculating the antecedent soil moisture saturation [section 3c(1)]. Precipitation accumulations were extracted hourly and summed to give 24-h precipitation totals to compare with GPM precipitation accumulations. Several studies have found large improvements in performance for precipitation in ERA5 relative to ERA-Interim (e.g., Beck et al. 2019; Tarek et al. 2020; Nogueira 2020), including for TC cases (Hersbach et al. 2018b). A recent study found that ERA5 agrees well with gridded gauge data in terms of the spatial distribution of typhoon precipitation, although it can underestimate the most extreme precipitation (Jiang et al. 2021).

d. GloFAS-ERA5 reanalysis

The severity of fluvial flooding from TCs is difficult to calculate in a consistent way from observations given the relative sparsity of river discharge observations, particularly in many areas impacted by TCs (Lavers et al. 2019). Therefore, this study uses the GloFAS-ERA5 global river discharge reanalysis (Harrigan et al. 2020b), as a proxy for river discharge observations. The Global Flood Awareness System (GloFAS) is designed to provide a global overview of upcoming flood events to decision-makers such as humanitarian organizations (e.g., Coughlan de Perez et al. 2016) as part of the Copernicus Emergency Management Service for floods. In GloFAS, ensemble meteorological forecasts from ECMWF are processed by the revised land surface hydrology scheme (HTESSEL) to create land surface runoff fields, with additional hydrological processes such as flow routing provided by the LISFLOOD hydrological model to forecast river discharge (Harrigan et al. 2020a; Alfieri et al. 2013; Balsamo et al. 2009, 2011; Van Der Knijff et al. 2010; Hirpa et al. 2018). Daily GloFAS-ERA5 river discharge reanalysis data, with a horizontal resolution of 0.1°, were extracted from the C3S CDS (Harrigan et al. 2019), for each day of 2010–19 and were used to define the flood severity in each landfalling TC case (section 3d).

Although GloFAS-ERA5 is partially dependent on model-derived precipitation and soil moisture, the groundwater and river routing parameters were calibrated against daily river discharge from 1287 observation stations worldwide (Hirpa et al. 2018) and a recent comprehensive global evaluation found it was skillful against a mean flow benchmark in 86% of catchments (Harrigan et al. 2020b). Other studies have compared GloFAS-ERA5 and observed discharge across regions (Ficchì and Stephens 2019; Towner et al. 2020), or in case studies e.g., Emerton et al. (2020), for Cyclones Idai and Kenneth in Mozambique. The use of GloFAS-ERA5 reanalysis data allows an analysis of the flooding across the whole area affected by each TC, rather than just at isolated points on the river network. It also allows a direct comparison of flood severity across all TC cases, using openly available data, in a way that would not be possible using discharge observations.

3. Methods

a. Selecting TC cases and calculating their land footprint

For the period 2010–19, the IBTrACS TC track data are filtered to only include those TCs with a landfall at tropical storm strength or above [≥34 kt (1 kt ≈ 0.51 m s^{-1})]. The 6-hourly latitude and longitude track position data are extracted where the storm is at or above tropical storm strength, and is over land or within 500 km of land. To decide the area, henceforth termed the "footprint," in which to calculate the flood severity for each TC, the following steps are taken (illustrated for Cyclone Eliakim, which impacted Madagascar in 2018, in Fig. 1):

1) The TC positions at the start and the end of each day are extracted and used to create a daily mask representing the area of precipitation defined as being associated with the TC, by calculating all grid points that lie within 500 km of that 24-h storm track segment. Most studies that have assigned a set radius for defining TC-related rainfall, have used a radius of either 500 km (Prat and Nelson 2013, 2016; Jiang et al. 2011; Luitel et al. 2018), 550 km (Zhou and Matyas 2017) or 5° (Guo et al. 2017).
2) The daily masks are combined to create an overall mask for the track of each TC.
3) The overall mask is combined with a land sea mask to create the final land footprint used to analyze the flood severity.
4) The number of valid 0.1° by 0.1° points (the GloFAS-ERA5 reanalysis resolution) in the land footprint is calculated, and the small number of TCs with fewer than 1000 points in their footprint (equating to an area of ~100 000 km^2) are excluded to ensure the footprints are large enough to allow a sensible comparison of the area affected by severe flooding.

This process results in 280 landfalling TC cases from 2010 to 2019 being included in the study (broken down by basin in Fig. 2). For each case, a set of TC characteristics, catchment characteristics, and flood severity measures are calculated to allow an analysis across the cases of the key factors influencing the severity of fluvial flooding.

b. TC characteristics

1) INTENSITY

The minimum MSLP and maximum sustained wind speed are extracted from IBTrACS at each track point, and the prelandfall value is used to identify the storm intensity for consideration in each case. To ensure that wind speed data are comparable between basins, 1-min maximum sustained winds are used where available.

2) TRANSLATION SPEED

Translation speed (speed in a forward motion) is calculated for each 6-h track segment by dividing the distance between the track points by 6 h. Three speed measures are then taken forward to consider in the analysis: (i) translation speed during the track segment where it makes landfall, (ii) average speed while the storm track is over land (the mean of the translation speeds

Unauthenticated | Downloaded 06/29/23 02:14 PM UTC

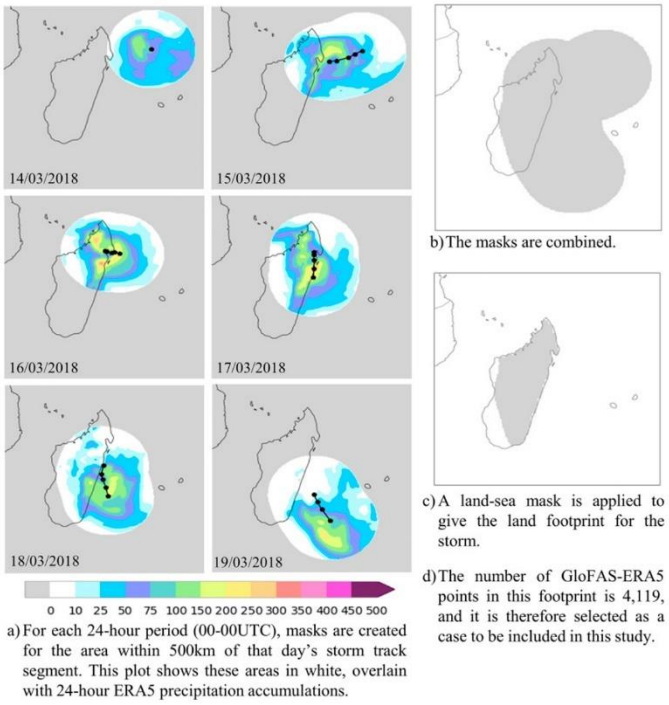


FIG. 1. Method for calculating the land footprint of each TC case, illustrated for Cyclone Eliakim, which impacted Madagascar in March 2018.

from all 6-hourly track segments over land), and (iii) minimum translation speed while the storm is over land (the minimum of the translation speeds from all 6-hourly track segments over land), to take account of storms that stagnate after landfall.

3) SIZE AND SHAPE

TCs span a huge range of sizes, with gale-force winds extending anywhere from ~20 km from the center to over 1000 km. A number of methods are commonly used to calculate storm size and shape, including the radial extent of winds reaching certain threshold values in knots (R43, R50, R64), the radius of the maximum winds (RMW), and the radius of the outermost closed isobar (ROCI) (Merrill 1984; Weber et al. 2014). The ROCI delimits the outermost extent of a TC's wind circulation and is the most relevant of these metrics for this study. Each RSMC will have its own, often subjective, procedures to calculate the ROCI recorded in IBTrACS, so to allow global comparisons this study recalculates the ROCI using ERA5 MSLP fields along the TC tracks. For each track point the MSLP field is centered on the storm location and then decomposed into tangential and radial coordinates, essentially unwrapping the field around the storm center. A search is carried out for the outermost closed MSLP isobar on each radial angle. The distance to this isobar is then averaged for all radial angles to give the overall ROCI, and is averaged for each quadrant (left front, right front, right rear, and left rear) to give shape information. Storm asymmetry is calculated by the ratio of the quadrant with the highest ROCI to that with the lowest ROCI. The average size and asymmetry in the day prior to landfall in each case are used in the subsequent analysis.

4) PRECIPITATION

The daily masks containing all points within 500 km of the storm track in each 24-h period (Fig. 1a) are applied to the 24-h precipitation accumulation data from GPM and ERA5, to obtain the precipitation considered to be associated with the TC. Three subsequent days are also included using the final mask to ensure that precipitation from any slow-moving remnants of the storm that may still contribute to flooding are included in the total storm precipitation. For the day prior to landfall, the proportion of the masked area with precipitation accumulations greater than 100 mm is calculated and included as a characteristic, to see how this information, which could be better estimated ahead of landfall, compares to flood severity.

To calculate the total storm precipitation, the masked daily precipitation fields are combined to give the overall storm-total

Unauthenticated | Downloaded 06/29/23 02:14 PM UTC

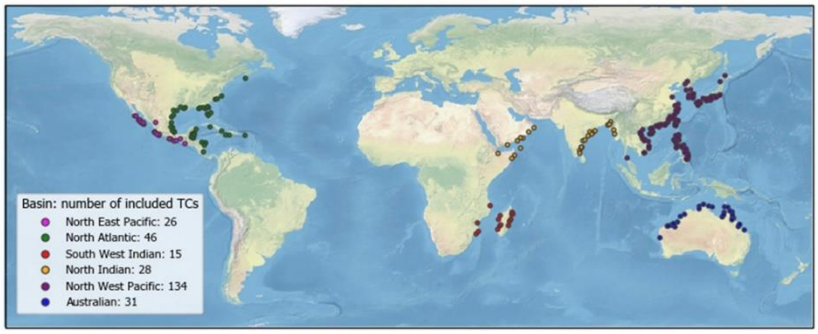


FIG. 2. The locations of the first land point in the track of the 280 TCs included in this study (2010–19), including the breakdown of cases by the TC basin. Note that no TCs are included in the South Pacific basin, where the only landfalls are in small island nations and therefore the land footprints were too small to allow for a fair comparison with the other cases.

GPM and ERA5 precipitation in the land footprint (illustrated for Cyclone Eliakim in Fig. 3). To investigate a range of precipitation characteristics in terms of their influence on the flood severity metrics, the following values are calculated from both GPM and ERA5:

1) the highest precipitation accumulation at any point in the land footprint area,
2) the average precipitation accumulation across the land footprint area,
3) the average precipitation accumulation for points in the land footprint area where the total is greater than 25 mm (giving an estimate of the average precipitation accumulated in those areas where the storm led to rainfall), and
4) the proportion of the land footprint area with precipitation accumulations greater than 25, 100, and 200 mm.

c. *Catchment characteristics*

1) ANTECEDENT SOIL MOISTURE SATURATION

ERA5-derived soil moisture saturation fields were calculated as a percentage of the progress along a scale from the permanent wilting point soil water content to the saturation soil water content [as specified by soil type in Table 8.9 of ECMWF (2020)]. The data were calculated on two levels: layer 1 (0–7 cm), and layers 1–3 combined (0–100 cm). For each TC case, the antecedent soil moisture saturation data were taken from the date prior to the storm coming within 500 km of land, to ensure the prestorm saturation levels were being recorded. The antecedent soil moisture saturation for the Cyclone Eliakim case is shown in Fig. 3c. For each layer, the average value in the storm land footprint was calculated, along with the proportion of the footprint with saturation greater than 90%.

2) OROGRAPHY AND GRADIENT

Two of the GloFAS input variables, detailing the orography and the gradient of each grid point, were extracted for each footprint area. The average orography and gradient within the storm footprint area were calculated for consideration in the data analysis.

d. *Calculating flood severity*

The period used to calculate the flood severity for each TC case was defined by setting the first date as the landfall date and the last date as a week after the final point in the cyclone track, to allow time for the affected watersheds to respond to the rainfall. For each day within this case period the GloFAS-ERA5 discharge data in the land footprint area were compared with return period flood thresholds to calculate the return period exceeded at each point. The return period thresholds were taken directly from GloFAS, where they are determined from the GloFAS-ERA5 river discharge reanalysis, by fitting a Gumbel extreme value distribution on the annual maxima time series over the 1979–2018 period (Alfieri et al. 2019; Harrigan et al. 2020a; Zsótér et al. 2020).

Flood severity can be defined in several ways, each measuring a contrasting property of the flood hazard that may be of particular relevance to different stakeholders. In this study the flood severity is calculated in the following three ways for each storm:

1) *Flood severity (area)*: This is defined as the percentage of GloFAS grid points within the storm's footprint where the discharge exceeds the 5-yr return period value for that point at any date during the case period. This has been termed "area" to simplify the decomposition of flood severity into three component parts, but it is more akin to "floodiness" (Stephens et al. 2015) defined over a storm's footprint than it is to measures of inundation extent. For Cyclone Eliakim, the flood severity (area) was 22% (Fig. 3d).
2) *Flood severity (duration)*: This is defined by selecting those points where the 5-yr return period was exceeded on at least one day, and then calculating the average number of days for which the 5-yr return period was exceeded at those points. This value is set to zero for

Unauthenticated | Downloaded 06/29/23 02:14 PM UTC

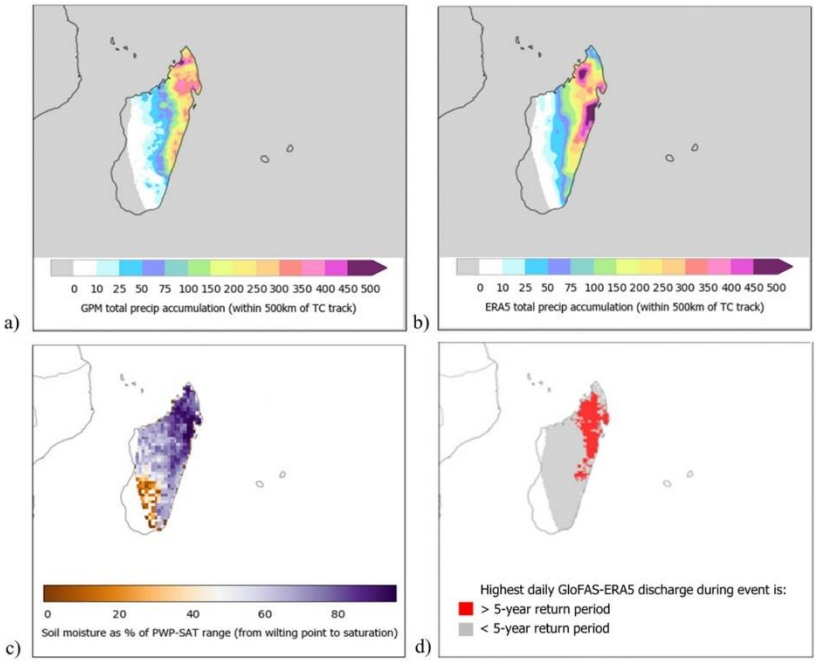


FIG. 3. Cyclone Eliakim (2018): (a) total event precipitation calculated from GPM, (b) total event precipitation calculated from ERA5, (c) antecedent soil moisture saturation in the top layer of soil, and (d) GloFAS-ERA5 points that exceeded the 5-yr return period during the event (22% of land footprint).

cases in which no points exceed the threshold at all during the case period.

3) *Flood severity (magnitude)*: For this magnitude-based definition the data are filtered to only include points that have an upstream area of greater than 1000 km^2, to focus on the magnitude of the flooding on larger rivers in the footprint. The magnitude of the flood is defined by first calculating the maximum return period that is exceeded at each of these river points during this TC event, and then ranking the points from highest to lowest and calculating the average return period exceeded over the top 100 points. The return periods are capped at 200 years as a sensible upper limit, given that the return period thresholds are generated on a 40-yr sample (Harrigan et al. 2020a). In this way the magnitude-based definition of flood severity compares the most severely affected river points in each TC case, in terms of the extremeness of the discharge levels experienced.

4. Results

a. Comparison of flood severity measures

The flood severity of all the TC cases, split geographically by the TC basin, is summarized in Fig. 4. For all basins other than the South Indian Ocean, most of the cases have relatively low flood severity (area) (median of 1%–5%), but each basin has several cases with higher flood severity. However, in the South Indian Ocean basin, more of the cases have a higher flood severity (area), with the median of 13% (Fig. 4a) found by a Kruskal–Wall H test to be significantly different from the other basins (significance level $p < 0.01$). For flood severity (duration) the median duration for those points exceeding the 5-yr return period threshold is between one and three days in all regions (Fig. 4b). The median duration was higher for the Australian basin and the lower for the North West Pacific (both significant at $p < 0.01$). For flood severity (magnitude) the South Indian Ocean is again shown to have a higher median value than other regions have (significant at $p < 0.01$). All basins have some cases in which very high return periods are exceeded (Fig. 4c).

There are significant positive Spearman correlations between the three flood severity measures (Fig. 5). Although the correlations are strong (>0.5 in each case), there are some cases in which storms have a very high flood severity in one measure and not in another. Good examples of this are Typhoon Hagibis (2019) and Typhoon Lan (2017), which can be seen in the bottom right of the plot in Fig. 5a, with highest flood

Unauthenticated | Downloaded 06/29/23 02:14 PM UTC

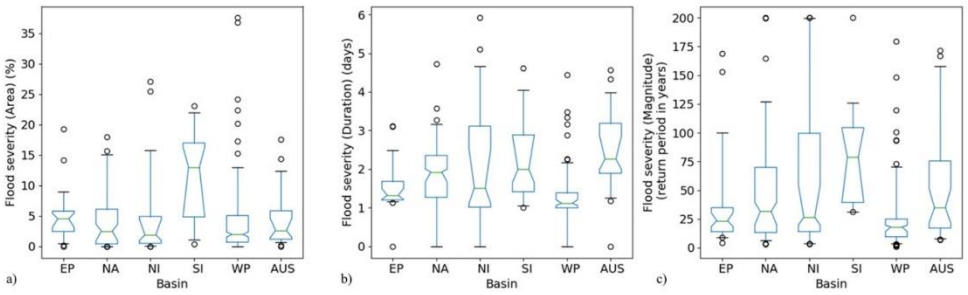


FIG. 4. Box-and-whisker plot showing (a) flood severity (area), (b) flood severity (duration), and flood severity (magnitude) grouped by TC basin [Australian (AUS), eastern Pacific (EP), North Atlantic (NA), "North Indian" (NI), "South Indian" (SI), and "North West Pacific" (WP)]. In all of the box-and-whisker plots in this study, the box extends from the Q1 to Q3 quartile values of the data and the whiskers extend from the 5th to the 95th percentiles. Outliers beyond this point are plotted as circles. The median is shown by a green line, with the 95% confidence bounds in that median shown by the notched area of the box, calculated by bootstrapping.

severity (area), but where most of that flooding was very short lived, with flood severity (duration) of only around 1 day. This emphasizes the importance of considering multiple indices, to take into consideration all characteristics of the flood severity.

The top 10 most severe cases according to each of the flood severity indices are shown in Table 1. All basins have storms represented in the lists. Many storms appear in the top 10 for two of the indices, and two storms, Cyclone Luban (2018) and Cyclone Idai (2019) appear in all three lists. Table 2 contains a summary of the impacts from the TC cases that appear in at least two of the three top-10 lists. While it is harder to ascertain whether there were cases with severe impacts that were not highlighted, the fact that all of the storms in Table 2 had significant impacts from flooding indicates that the methods used in this paper have correctly highlighted high-impact cases of fluvial flooding from TCs.

b. Relationships between flood severity and case characteristics

1) CORRELATIONS

The Spearman's rank correlation coefficients between the flood severity metrics and all the calculated storm and catchment characteristics are shown in Fig. 6. A comparison between GPM and ERA5 rainfall (not shown) revealed very strong correlations, so only the ERA5 precipitation variables are included. All the flood severity variables are significantly correlated with each other, as was seen in Fig. 5. All precipitation variables (both for the precipitation field prior to landfall and for the accumulated precipitation variables within the footprint) are significantly positively correlated with the flood severity (area) index, with the highest correlations for the average accumulation in the footprint where it rained (0.58) and for the percentage of the footprint over 100-/200-mm

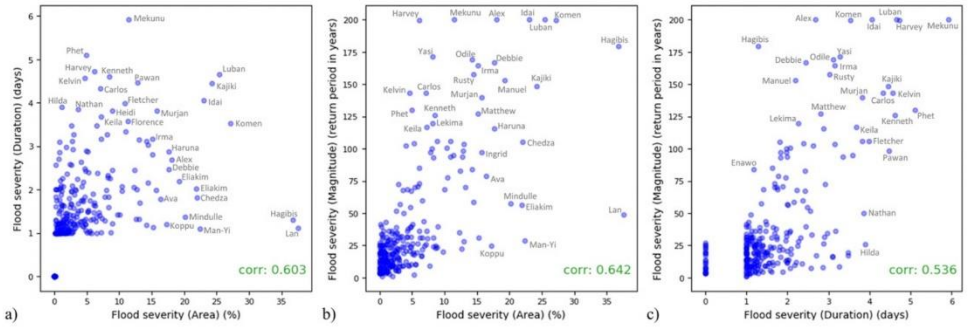


FIG. 5. Scatterplots of the different flood severity indices: (a) flood severity (area) vs flood severity (duration); (b) flood severity (area) vs flood severity (magnitude); and (c) flood severity (duration) vs flood severity (magnitude). The Spearman's rank correlation coefficient is displayed on each plot. Points are annotated with the storm name if they have high values on either axis or are outliers to the general pattern.

Unauthenticated | Downloaded 06/29/23 02:14 PM UTC

TABLE 1. The 10 highest-ranked TCs for each flood severity measure (in boldface type if that TC case is in the top 10 for two of the three indices, and in boldface italic if it is in the top 10 of all three indices). The regions are defined in the caption of Fig. 4.

Rank	Flood severity (area)	Flood severity (duration)	Flood severity (magnitude)
1	Lan (WP; 2017)	**Mekunu** (NI; 2018)	Alex (NA; 2010)
2	**Hagibis** (WP; 2019)	Phet (NI; 2010)	***Luban*** (NI; 2018)
3	**Komen** (NI; 2015)	**Harvey** (NA; 2017)	**Mekunu** (NI; 2018)
4	***Luban*** (NI; 2018)	***Luban*** (NI; 2018)	***Idai*** (SI; 2019)
5	**Kajiki** (WP; 2019)	Kenneth (SI; 2019)	**Harvey** (NA; 2017)
6	***Idai*** (SI; 2019)	Kelvin (AUS; 2018)	**Komen** (NI; 2015)
7	Man-Yi (WP; 2013)	Pawan (NI; 2019)	**Hagibis** (WP; 2019)
8	Chedza (SI; 2015)	**Kajiki** (WP; 2019)	Yasi (AUS; 2011)
9	Eliakim (SI; 2018)	Carlos (AUS; 2011)	Odile (EP; 2014)
10	Mindulle (WP; 2016)	***Idai*** (SI; 2019)	Debbie (AUS; 2017)

accumulations (0.57/0.54). For flood severity (duration), there is no significant correlation with the precipitation prior to landfall, but there are significant correlations with most total storm accumulation measures. For flood severity (magnitude), there are positive correlations with all precipitation variables apart from the percentage of the footprint over the lowest value of 25 mm.

The storm *intensity* variables, MSLP and wind speed at landfall, are significantly negatively (−0.18) and positively (0.13) correlated with flood severity (area), respectively. This shows that generally the more intense the storm (lower MSLP minima and stronger winds), the greater the flood severity. However, neither intensity variable is significantly correlated for flood severity (duration) or flood severity (magnitude), so it is not always the case that a strong storm will lead to severe flooding in terms of duration or magnitude, or that a weaker storm cannot cause significant flood events. A comparison of the distribution of flood severity values for TC cases split into those that are tropical storms at landfall with those that have equivalent hurricane or major hurricane strength (Fig. 7) confirms that all storm categories have some cases with both low and high flood severities. There are no significant differences between the severity of flooding in tropical storms and hurricanes, but for major hurricanes there is a notable increase in the upper quartile range and the median is significantly higher for the area and magnitude definitions of flood severity (Kruskal–Wall H test with $p < 0.01$).

The three *translation speed* variables are all significantly negatively correlated with all the flood severity indices i.e., slower-moving storms tending to have greater flood severity when considered in terms of area, duration and magnitude. The strongest correlations are with the duration of flooding (−0.31, −0.32, and −0.33, respectively, for the translation speed at landfall, the average translation speed over land, and the minimum translation speed over land).

The *storm size* as measured by the ROCI is significantly positively correlated with flood severity (area) (0.12), indicating that larger storms tend to have a greater percentage of the land footprint exceeding the 5-yr return period. There is no significant correlation with flood severity (duration) or flood severity (magnitude), and no significant correlation of *storm asymmetry* with any of flood severity measures.

The relationship between flood severity and *soil saturation* is complex and varies with each of the flood severity indices. For

TABLE 2. A summary of the impacts from those TCs that appear in the top 10 for at least two of the flood severity measures (in boldface type if that TC case is in the top 10 for two of the three indices, and in boldface italic if it is in the top 10 of all three indices). All are TCs that led to significant impacts.

Case	Summary of impacts
Idai (SI; 2019)	Severe flooding in Mozambique, Malawi, and Zimbabwe, leading to a major humanitarian crisis; damages of at least $2.2 billion (2019 USD), 1.85 million people in need of humanitarian assistance, and over 600 fatalities in Mozambique alone (Emerton et al. 2020); deadliest TC recorded in the South West Indian Ocean basin
Luban (NI; 2018)	Flooding in Somalia, Oman, and Yemen, with 14 deaths in Yemen, with damages in the country estimated at U.S. $1 billion (Associated Press 2018)
Hagibis (WP; 2019)	98 confirmed deaths in Japan and more than 270 000 households left without power across the country; severe damage to transport infrastructures
Komen (NI; 2015)	Komen led to 45 deaths in Bangladesh, 83 in India, and 39 deaths in Myanmar
Kajiki (WP; 2019)	At least 6 fatalities in Vietnam, with agricultural losses estimated at U.S. $12.9 million
Mekunu (NI; 2018)	31 fatalities (Socotra island, Yemen, and Oman); severe flooding led to power outages, and landslides
Harvey (NA; 2017)	Unprecedented flooding in Texas inundating hundreds of thousands of homes, displacing more than 30 000 people, and prompting more than 17 000 rescues; over 100 fatalities

Unauthenticated | Downloaded 06/29/23 02:14 PM UTC



FIG. 6. Spearman's rank correlation coefficients of the flood severity variables with the TC characteristics and catchment characteristics for each TC case. Those with a Spearman's rank correlation coefficient with a significance value <0.05 are marked in boldface type with an asterisk.

flood severity (area), all the soil saturation metrics have positive correlations, showing that cases with more saturated antecedent soil conditions tend to have greater flood severity. However, a different pattern is shown for the other two flood severity indices (duration and magnitude), where there are significant negative correlations with the average soil moisture saturation in the top 100-cm layer (−0.27 and −0.15, respectively).

There is a significant positive relationship of average *orography* and flood severity (area) (0.24), and there are significant

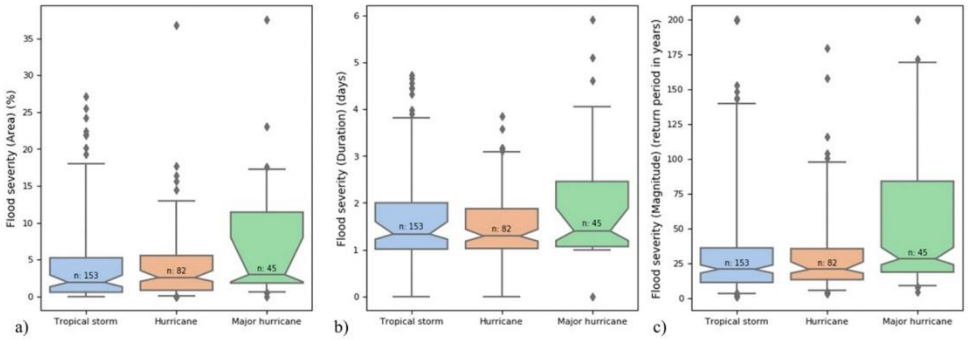


FIG. 7. Flood severity [(a) area, (b) duration, and (c) magnitude], for three categories of TC calculated based on the maximum wind speed at landfall taken from IBTrACS: (i) tropical storm (34–63 kt), (ii) equivalent hurricane strength (64–95 kt), (iii) equivalent major hurricane strength (96+ kt). In actuality, different storm categories and thresholds are used by different RSMCs when defining intensity, but to compare across basins the categories from the Saffir–Simpson scale are used here. The 95% confidence bounds in the median is shown by the notched area of the box, calculated by bootstrapping. The number of cases in each category is overlain.

Unauthenticated | Downloaded 06/29/23 02:14 PM UTC

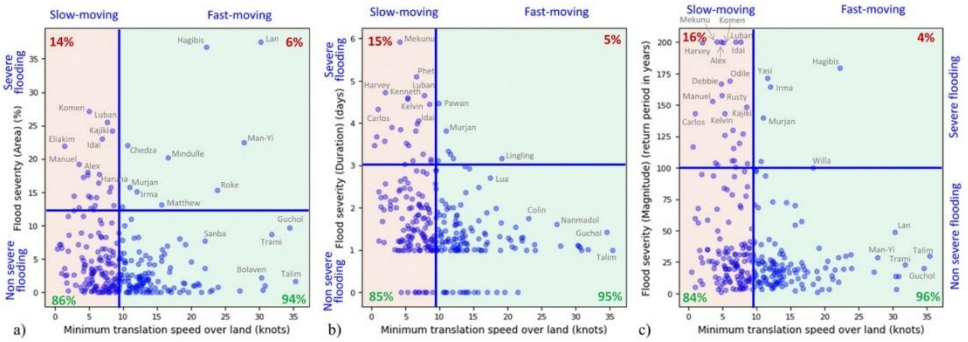


FIG. 8. Illustration of the conditional probability technique for examining the relationship between minimum translation speed over land and flood severity. The blue lines divide the plot into quadrants based on the median translation speed (9.19 kt) and the 90th percentile of the flood indices [12.86% for flood severity (area), 3.10 days for flood severity (duration), and 102.59 for flood severity (magnitude)]. There are, therefore, equal numbers of cases in the orange and green halves of the plots, and the upper section of the plot displays the 10% of the TC cases with the most severe flooding. The numbers in the top-right and top-left quadrants can be compared to see if there is an elevated risk of severe flooding for slow-moving storms. Points are annotated with the storm name if they have high values on either axis or are outliers to the general pattern. The Spearman's rank correlation coefficients for each plot are (a) −0.22, (b) −0.33, and (c) −0.26.

negative correlations of the average *gradient* with flood severity (duration) (−0.44) and flood severity (magnitude) (−0.31), so lower gradients would tend to have greater flood severity in terms of the flood duration and magnitude.

2) CONDITIONAL PROBABILITIES

Conventional wisdom is that translation speed is a significant risk factor leading to increased flood severity. Significant relationships were shown between the flood severity indices and the translation speed of the TC in Fig. 6, but the correlations are perhaps lower than anticipated, at −0.22, −0.33, and −0.26 for the minimum translation speed over land when compared with flood severity as measured by area, duration, and magnitude, respectively. However, if we look in more detail at scatterplots for these relationships, and apply a conditional probability technique to split the plot into four quadrants, further useful information can be found (Fig. 8). The *x* axis is split by the median value of the minimum translation speed over land, to designate slow-moving and fast-moving storms. A sensitivity analysis (not shown) was used to select which percentile of the *y* axes (the flood severity scores) to use to designate nonsevere and severe flooding cases. The 90th percentile was selected as the threshold that gave the best combination of having sufficient cases included in the upper section while also focusing in on the more severe cases (the top 10%). The conditional probability of having severe flooding if the storm is slow moving is then calculated. For flood severity (area), the conditional probability of having severe flooding if the storm is slow moving is 14%, more than double the conditional probability if the storm is fast moving (6%) (Fig. 8a). So although there are a few cases (e.g., Hagibis, Man-Yi, and Lan) in the upper-right quadrant of Fig. 8a in which flood severity is high in fast-moving storms and there are many cases in which slow-moving storms do not lead to severe flooding, overall there is a greater risk of severe flooding in slow-moving storms, with a significant cluster of storms with high flood severities and low translation speeds (e.g., Komen, Idai, Luban, Kajiki, and Eliakim). The results when we apply the same conditional probability technique to the other two flood severity metrics are even clearer, with 3 times the probability of having a severe flooding event for slow-moving storms for flood severity (duration), and 4 times the probability of a severe flooding event for slow-moving storms for flood severity (magnitude) (Figs. 8b,c).

Fast movement of TCs appears to often mitigate severe flooding, and slow movement increases the likelihood of severe flooding, but these patterns are not seen in all cases. The role of other factors in moderating this relationship can be examined by comparing conditional probability plots of flood severity with several different characteristics, selected based on the correlation results shown in Fig. 6 (Fig. 9). Flood severity (area) is used to examine in more detail the interplay between characteristics, as the extent of the severe flooding across the whole area affected by the TC is perhaps the most relevant when considering preparedness and recovery activities such as evacuation planning. The strong relationship between flood severity and the percentage of the footprint with high precipitation totals can be seen in Fig. 9a, with only 2% of the cases lower than the median precipitation having severe flooding. When the conditional probability technique is applied to the average soil saturation in the top layer of the soil, there is found to be more than double the risk (14% as compared with 6%) of having severe flooding where there is high soil saturation (Fig. 9b). Figures 9c–e show three of the TC characteristics that may play a role in increasing the overall precipitation accumulation: intensity, translation speed, and size. For the more

Unauthenticated | Downloaded 06/29/23 02:14 PM UTC

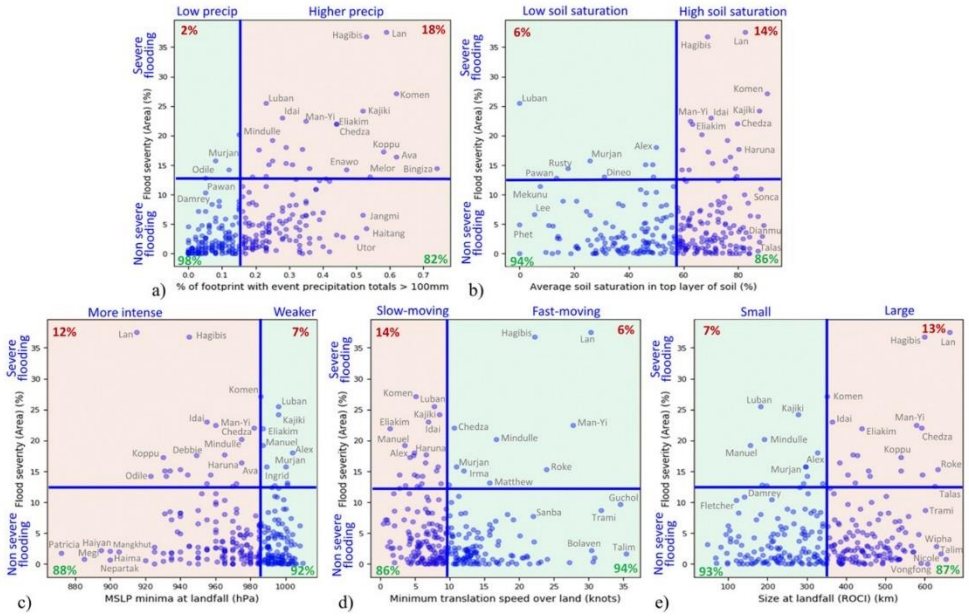


FIG. 9. Scatterplots for the flood severity (area) indices against (a) the percentage of the footprint with event precipitation totals greater than 100 mm, (b) average soil saturation in the top layer of soil, (c) mean sea level pressure minima at landfall, (d) minimum translation speed over land, and (e) size at landfall (ROCI). Each plot has the conditional probability technique applied using the median of the variable and the 90th percentile of the flood severity to create the quadrants. The Spearman's rank correlation coefficients for each plot are 0.57 for (a), 0.15 for (b), −0.18 for (c), −0.22 for (d), and 0.12 for (e).

intense storms, there is a small increased probability of severe flooding (12%) when compared with the less intense storms (7%) (Fig. 9c). The doubling of the likelihood of severe flooding for slow translations speeds, as shown in Fig. 8a, are repeated in Fig. 9d, while Fig. 9e shows the conditional probability results for TC size, with nearly double the risk of severe flooding for large storms (13%) relative to small storms (7%).

Typhoons Man-Yi (2013), Hagibis (2019) and Lan (2017) appear as relative outliers in the upper right quadrant of Fig. 9d (fast-moving storms where flood severities were unexpectedly high or very high). On Fig. 9e it can be seen that these are all very large storms, indicating that the size of these storms may have been a significant factor contributing to their high flood severity, acting to override the mitigative nature of their fast translation speed. They are also categorized in the more intense category for intensity (Fig. 9c), and the high soil saturation category (Fig. 9b). Thus, in these cases the high translation speed of these typhoons did not prevent a severe flood event due to the presence of several other risk factors.

A similar conditional probability analysis for the other two flood severity measures (duration and magnitude) (not shown) found that the main driving factor from the TC characteristics was the translation speed of the storm. The other main risk factors were from the catchment characteristics, with drier soils and lower gradients being associated with a large increase in the risk of having a longer and higher magnitude flood event. These results show that the influence of the catchment characteristics on the duration and magnitude of flooding is greater than for the flooding measured in terms of the area.

3) INTERPLAY OF CHARACTERISTICS

To investigate further how the combination of several factors helps to determine the overall flood severity, two characteristics, TC translation speed and TC size, have been used to split the cases into four categories based on whether they are larger or smaller than the median size and faster or slower than the median translation speed (Fig. 10). For flood severity (area), cases in which the TC is both slow and large have the highest median and upper quartile range, followed by slow and small, then fast and large, and finally fast and small cases. Meanwhile for flood severity (duration) and flood severity (magnitude) there is a clearer split between the fast types and the slow types, showing that for flood duration and magnitude the translation speed of the storm is a more important factor than the size of the storm.

This process is taken further by grouping each TC case depending on how many of the four risk factors found in Fig. 9

Unauthenticated | Downloaded 06/29/23 02:14 PM UTC

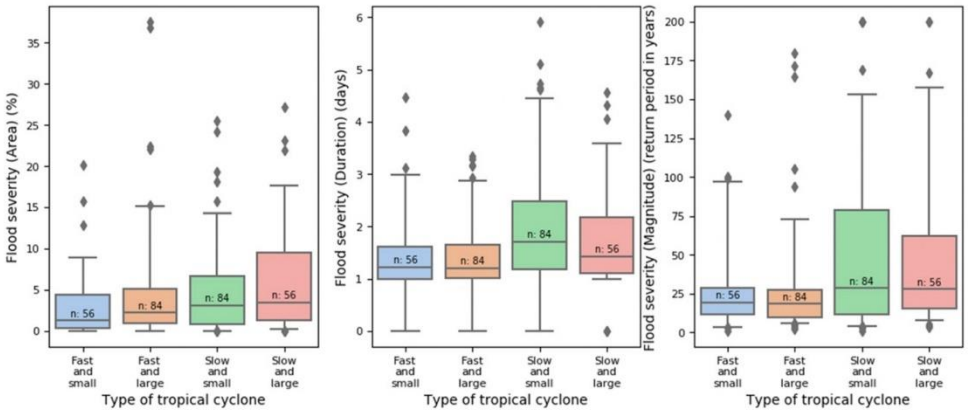


FIG. 10. Flood severity [(left) area, (center) duration, and (right) magnitude] for four categories of TC: (i) fast and small, (ii) fast and large, (iii) slow and small, and (iv) slow and large. The number of cases in each category is overlain.

(slow-moving storm, large storm, intense storm and wet antecedent conditions) are present (Fig. 11). Total precipitation is not included here as we are focusing on the factors that influence that precipitation, that are easier to estimate a priori when assessing the risk associated with a storm in forecast mode. Figure 11a shows that the likelihood of a higher flood severity increases the more risk factors are present. Figure 11b splits the categories by which of the risk factors are present. The storms that are classed as having all four risk factors (slow, large, strong, and wet) have the highest median flood severity and the highest upper quartile range. Within those cases that have three risk factors, cyclones that are slow and large with wet antecedent conditions stand out as having a high upper quartile range. For those cases with two risk factors, the highest median value is for cases with slow TCs with wet antecedent conditions, while where only one risk factor is present, there are higher flood severities for slow storms.

5. Discussion of the key outcomes

a. Applying flood severity metrics to TC footprints

This study has implemented three flood severity indices designed to measure the severity of flooding in TCs, based on (i) the area affected; (ii) the duration of the flooding; and (iii) the magnitude of the flooding in the most severely affected areas. Although the different flood severity measures are all significantly positively correlated with each other, there are cases in which a storm has a very high flood severity in one measure and not in another, for example, where a large area is affected but flood duration is low. Some of the drivers of an enhanced flood severity found in this study are consistent whether considering the area affected, duration, or magnitude e.g., high total precipitation and low translation speed, whereas some drivers are only significant for one or two of the indices e.g., large storms and the area-based index, and gentler slopes with the duration- and magnitude-based indices. This emphasizes the importance of considering multiple indices to take into consideration all characteristics of the flood hazard, especially given that stakeholders may have differing priorities. An organization with responsibility for emergency planning across a large geographical area would be particularly concerned about the extent of the area being affected by flooding, whereas an individual in an area prone to flooding may be most concerned about the magnitude of flooding likely to be experienced.

b. Importance of precipitation and translation speed

The strongest relationships are found with the metrics summarizing the event total precipitation, showing the importance of increasing the availability and profile of precipitation forecasts in TC forecast communication and verifying the ability of models that drive flood forecasting systems to forecast the precipitation associated with TCs.

Of the other TC characteristics, the speed of forward movement of the TC is found to be a key factor influencing the severity of the fluvial flood hazard, in terms of its area, duration and magnitude. This confirms that slower-moving storms, which rain over a given area for longer, tend to have higher flood severity, and that greater focus should be given to translation speed in forecast warnings and communication. Climate change is thought to have caused a general weakening of summertime tropical circulation, decreasing the translation speed of TCs (Kossin 2018). This, combined with an expected increase in the magnitude of rainfall rates in TCs (Knutson et al. 2010), is likely to lead to an increased fluvial flood hazard from these events in the future.

c. Presence of flood hazard at all levels of storm intensity

TC intensity is positively correlated with the flood severity as defined by area, but not significantly correlated with the duration or magnitude of the flooding. A further analysis splitting cases into those of equivalent tropical storm, hurricane, and

Unauthenticated | Downloaded 06/29/23 02:14 PM UTC

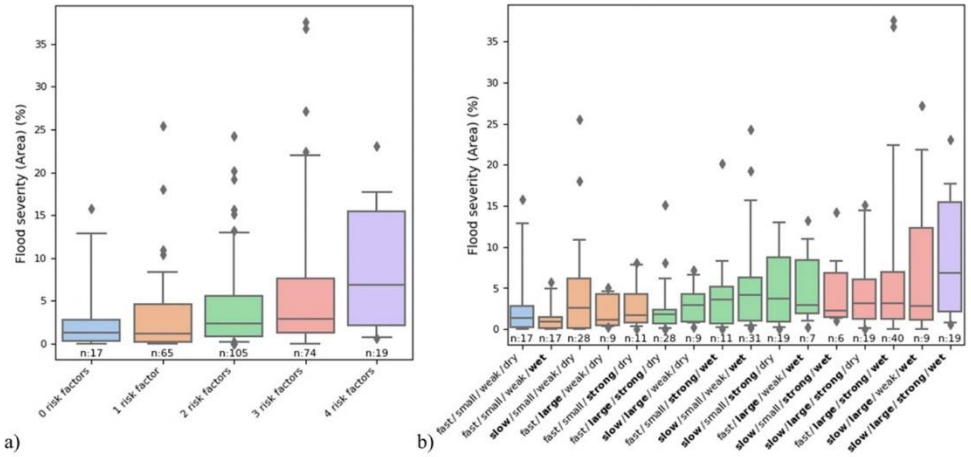


FIG. 11. Flood severity (area), for TCs grouped by risk factors (slow-moving storm, large storm, intense storm, and wet antecedent conditions) (a) grouped according to how many of the risk factors they possess and (b) further split by which risk factors (highlighted in boldface type) are present in each group. The number of cases in each category is displayed below the box plot.

major hurricane strength, showed that all three have cases spanning the full range of flood severity values. This shows the importance of forecasting agencies continuing to emphasize to both the public and to organizations involved with disaster risk reduction preparedness and response activities that the fluvial flood hazard can be severe in all categories of TC.

d. Complex relationship between flood severity and antecedent soil moisture

Conventional wisdom, established by looking at small to medium sized flood events, is that higher levels of soil moisture saturation would lead to an increased risk of flooding (Berghuijs et al. 2019). However, when dealing with the extreme precipitation totals that can be found in TC events, the relationship appears more complex. For flood severity (area), there are indeed more cases of severe flooding in those cases with high soil saturation. However, a different pattern is shown for the other two flood severity indices (duration and magnitude), where there is a higher conditional probability of severe flooding for lower (drier) antecedent soil moisture. Areas that are drier climatologically (e.g., the Arabian Peninsula) show higher return periods being exceeded, which is likely due to a combination of flash flooding, and the difficulty of accurately estimating return period thresholds in arid areas. Using soil moisture anomaly data rather than soil moisture saturation in future studies may help to further unpack this complex relationship.

e. Interplay of characteristics in modulating flood severity

The interplay between different characteristics acts to modulate the overall flood severity. For example, although fast-moving storms can mitigate fluvial flooding, there can still be a significant flood event if there are other risk factors present. An example of this is Typhoon Hagibis, which led to widespread and severe flooding in Japan in 2019. Here, the relative mitigation of Hagibis being fast moving was outweighed by the storm's size, intensity, and wider atmospheric environment, leading to very heavy precipitation and severe flooding, albeit over a relatively short period (Takemi and Unuma 2020). Contrastingly, although slow-moving storms significantly increase the conditional probability of a severe flood event, many slow-moving storms do not lead to significant flooding e.g., Tropical Storm Beryl (2012), which moved very slowly over northern Florida, but was relatively small and weak, and passed over an area where the antecedent conditions were dry. As this study included over 250 cases it was not possible to explore in detail the varied complexities that contributed to the detailed flood response in each individual case, such as the wider atmospheric environment, in a way that would be possible if only examining one or two cases. Nevertheless it provides a systematic analysis of potential meteorological and hydrological drivers, and a key finding is that the more driving factors a case possesses, the higher the likelihood of severe flooding, with in particular TCs that were slow, large, and strong, with wet antecedent conditions, having a higher flood severity in terms of the area that was affected.

f. Implications for research into the predictability of fluvial flooding from TCs

The severity of flooding in TC cases is shown to be dependent on a range of both meteorological and hydrological factors, showing the importance of a multidisciplinary Earth system approach to global flood prediction (Harrigan et al. 2020c). The flood forecasts from hydrometeorological forecasting systems will be highly dependent on how reliably the

Unauthenticated | Downloaded 06/29/23 02:14 PM UTC

input meteorological forecasts are able to predict not only the track and the intensity of TCs, but also their precipitation extent and intensity. This in turn will be dependent on the size and in particular the translation speed of the storm. There has generally been less focus on verifying these characteristics of TCs in ensemble forecasts, with most studies focusing on the predictability of track, intensity, and genesis. To understand and improve the forecasts of fluvial flooding from TCs, studies evaluating these aspects of forecast performance are recommended. It would also be beneficial to examine how uncertainty in each aspect of the forecast, and in each part of the forecasting chain, impacts on the overall predictability of fluvial flooding for TCs.

g. Implications for disaster risk reduction, preparedness, and response

The finding that the forward speed of the storm is of more importance than the intensity in modulating the severity of fluvial flooding is an important result. The expected speed of movement of the storm tends to have far less coverage in both traditional and social media than the intensity of the storm. This means that the public, local emergency responders, and those involved in planning disaster risk reduction activities may become too focused on storm intensity categories when making evacuation decisions, undertaking storm preparedness activities, or planning deployment of resources. Significant impacts from fluvial flooding can occur regardless of whether it is a strong storm, where there will likely also be significant wind and storm surge hazards, or a weak storm, where the fluvial flooding may be the main hazard. Both circumstances have important implications for forecast-based early action. An example of a TC that caused significant flooding while being relatively weak is Cyclone Komen, which impacted Myanmar, Bangladesh and India in 2015. Komen only reached equivalent strength of a tropical storm, but was very slow moving, with a broad circulation leading to prolonged heavy rainfall over areas of Myanmar already saturated due to weeks of monsoonal flooding, leading to severe impacts. Raising the profile of cases such as these where the wind hazards may be relatively low, but there is potential for a slow-moving storm impacting vulnerable areas, meaning the fluvial flood hazard and impacts are likely to be more severe, will help to ensure that preparations are made to ameliorate impact. For intense TCs, where evacuations and deployment of aid may be being organized in advance of landfall due to the anticipated wind impacts, it is also vital to know whether that storm also has an elevated likelihood of fluvial flooding due to other risk factors such as a slow translation speed or saturated antecedent soil conditions. If it does it would be important to select evacuation routes, shelters and aid deployment sites that would not be impacted in the event of extensive fluvial flooding, which may extend far from the landfall location. For example, Cyclone Idai, which led to catastrophic flooding in Mozambique in March 2019, exhibited most of the factors shown in this study to be key drivers of fluvial flood hazard: it was a strong storm, but significantly it was also a slow-moving storm at and after landfall, a relatively large storm, and also had wet antecedent soil moisture conditions. Highlighting this co-occurrence of flood drivers to decision-makers in future cases in which a storm is approaching landfall will help to ensure early action can be taken to plan for and mitigate the impacts of severe flooding.

6. Conclusions

Landfalling TCs over the last decade have been classified in terms of the severity of fluvial flooding that was associated with each storm using GloFAS-ERA5 reanalyses. Flood severity has been calculated using three different indices, each focusing on a different characteristic of the associated fluvial flooding (area, duration and magnitude). The key factors that impact the severity of fluvial flooding were investigated by comparing these flood severities with a range of TC characteristics (storm speed, size, shape, intensity, and precipitation), and catchment characteristics (antecedent soil moisture, orography, and gradient). The key findings are the following.

- There are cases that have very high flood severity in one measure and not in another, highlighting the importance of considering the area, duration and magnitude characteristics of TC flood events.
- Flood severity is strongly positively correlated with metrics summarizing the event total precipitation, showing the importance of increasing the availability and profile of precipitation forecasts in TC forecast communication and verifying the ability of models that drive flood forecasting systems to forecast the precipitation associated with TCs.
- Slow-moving storms were found to have a much higher conditional probability of a severe flood event relative to fast-moving storms, confirming that slower-moving storms, where storms are raining over an area for longer, tend to have higher flood severity.
- All intensities of TC have cases spanning the full range of flood severity values, including several cases of tropical storm strength that were associated with severe flooding. This confirms the need for forecasting agencies to continue emphasizing that the fluvial flood hazard can be severe in all categories of TC.
- Large storms have a higher conditional probability of severe flooding than small storms in terms of the area affected. The duration and magnitude of the flooding are found to be more dependent on the catchment characteristics within the land footprint of the storm, with gentler slopes having a much greater conditional probability of severe flooding. The relationship between flood severity and antecedent soil moisture is found to be complex and not consistent across the different flood severity indices.
- Several TC and catchment characteristics often combine to influence the overall flood severity and negate or enhance other factors e.g., fast-moving storms still having a severe flood hazard if they are large, intense and passing over saturated ground. The more driving factors that a case possesses the higher the flood risk is found to be, with the highest likelihood of having a large area affected by flooding found for TCs that were slow, large, intense, and affecting areas with wet antecedent conditions.

The reanalysis data used in this study provide a proxy for the observed severity of flooding, and, although this allows a direct

Unauthenticated | Downloaded 06/29/23 02:14 PM UTC

comparison of flood severity across all TC cases, studies of fluvial flooding in TCs would benefit hugely from an increase in the availability and consistency of hydrological observations (Lavers et al. 2019).

This study has highlighted the characteristics of TCs that have the most control on flood hazard. For those involved in communicating early warnings and taking forecast-based action before a storm our results show the importance of considering aspects such as storm speed when assessing the risks. For developers of hydrometeorological ensemble forecasts of river flows our work underlines that the input meteorological ensemble forecasts for TCs need to be able to reliably predict not just their track and intensity, but also their precipitation, size and translation speed. Future research needs to focus on verifying and improving these characteristics for agencies to be able to provide more accurate forecasts of flood hazard.

Acknowledgments. The authors thank three anonymous reviewers for their helpful comments and suggestions, which led to improvements to the paper. Authors Helen Titley and Joanne Robbins acknowledge the support of the Weather and Climate Science for Service Partnership (WCSSP) India and South East Asia projects, supported by the U.K. government's Newton Fund. Authors Liz Stephens and Hannah Cloke acknowledge funding from the U.K.'s Natural Environment Research Council (NERC) and Foreign Commonwealth & Development Office (FCDO) (formerly DFID) Science for Humanitarian Emergencies and Resilience (SHEAR) research program Forecasts for Anticipatory Humanitarian Action (FATHUM) project Grant NE/P000525/1. Author Hannah Cloke also acknowledges funding from the U.K.'s Natural Environment Research Council (NERC) The Evolution of Global Flood Risk (EVOFLOOD) project Grant NE/S015590/1.

Data availability statement. The data used in this study were accessed from the following freely available sources: (i) GPM IMERG final data (https://doi.org/10.5067/GPM/IMERG/3B-HH/06), (ii) ERA5 (precipitation, MSLP, and soil moisture) (https://doi.org/10.24381/cds.adbb2d47), (iii) GloFAS-ERA5 (https://doi.org/10.24381/cds.a4fdd6b9), and (iv) IBTrACS (https://doi.org/10.25921/82ty-9e16).

REFERENCES

Alfieri, L., P. Burek, E. Dutra, B. Krzeminski, D. Muraro, J. Thielen, and F. Pappenberger, 2013: GloFAS—Global ensemble streamflow forecasting and flood early warning. *Hydrol. Earth Syst. Sci.*, **17**, 1161–1175, https://doi.org/10.5194/hess-17-1161-2013.

——, E. Zsoter, S. Harrigan, F. Aga Hirpa, C. Lavaysse, C. Prudhomme, and P. Salamon, 2019: Range-dependent thresholds for global flood early warning. *J. Hydrol. X*, **4**, 100034, https://doi.org/10.1016/j.hydroa.2019.100034.

Associated Press, 2018: Yemen says 14 killed in Cyclone Luban. Accessed 15 September 2020, https://apnews.com/705065c5517545adae9e4fb313bb8381.

Balsamo, G., A. Beljaars, K. Scipal, P. Viterbo, B. van den Hurk, M. Hirschi, and A. K. Betts, 2009: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the Integrated Forecast System. *J. Hydrometeor.*, **10**, 623–643, https://doi.org/10.1175/2008JHM1068.1.

——, F. Pappenberger, E. Dutra, P. Viterbo, and B. van den Hurk, 2011: A revised land hydrology in the ECMWF model: A step towards daily water flux prediction in a fully-closed water cycle. *Hydrol. Processes*, **25**, 1046–1054, https://doi.org/10.1002/hyp.7808.

Beck, H. E., and Coauthors, 2019: Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrol. Earth Syst. Sci.*, **23**, 207–224, https://doi.org/10.5194/hess-23-207-2019.

Berghuijs, W. R., S. Harrigan, P. Molnar, L. J. Slater, and J. W. Kirchner, 2019: The relative importance of different flood-generating mechanisms across Europe. *Water Resour. Res.*, **55**, 4582–4593, https://doi.org/10.1029/2019WR024841.

Chen, S. S., J. A. Knaff, and F. D. Marks, 2006: Effects of vertical wind shear and storm motion on tropical cyclone rainfall asymmetries deduced from TRMM. *Mon. Wea. Rev.*, **134**, 3190–3208, https://doi.org/10.1175/MWR3245.1.

Cheung, K., and Coauthors, 2018: Recent advances in research and forecasting of tropical cyclone rainfall. *Trop. Cyclone Res. Rev.*, **7**, 106–127, https://doi.org/10.6057/2018TCRR02.03.

Coughlan de Perez, E., and Coauthors, 2016: Action-based flood forecasting for triggering humanitarian action. *Hydrol. Earth Syst. Sci.*, **20**, 3549–3560, https://doi.org/10.5194/hess-20-3549-2016.

Czajkowski, J., and E. Kennedy, 2010: Fatal tradeoff? Toward a better understanding of the costs of not evacuating from a hurricane in landfall counties. *Popul. Environ.*, **31**, 121–149, https://doi.org/10.1007/s11111-009-0097-x.

——, G. Villarini, M. Montgomery, E. Michel-Kerjan, and R. Goska, 2017: Assessing current and future freshwater flood risk from North Atlantic tropical cyclones via insurance claims. *Sci. Rep.*, **7**, 41609, https://doi.org/10.1038/srep41609.

ECMWF, 2020: IFS documentation CY47R1—Part IV: Physical processes. ECMWF Rep., 228 pp., https://doi.org/10.21957/cpmkqvhja.

Emerton, R., and Coauthors, 2020: Emergency flood bulletins for Cyclones Idai and Kenneth: A critical evaluation of the use of global flood forecasts for international humanitarian preparedness and response. *Int. J. Disaster Risk Reduct.*, **50**, 101811, https://doi.org/10.1016/j.ijdrr.2020.101811.

Ficchì, A., and L. Stephens, 2019: Climate variability alters flood timing across Africa. *Geophys. Res. Lett.*, **46**, 8809–8819, https://doi.org/10.1029/2019GL081988.

Guo, L., N. P. Klingaman, P. L. Vidale, A. G. Turner, M. Demory, and A. Cobb, 2017: Contribution of tropical cyclones to atmospheric moisture transport and rainfall over East Asia. *J. Climate*, **30**, 3853–3865, https://doi.org/10.1175/JCLI-D-16-0308.1.

Harrigan, S., E. Zsoter, C. Barnard, F. Wetterhall, P. Salamon, and C. Prudhomme, 2019: River discharge and related historical data from the Global Flood Awareness System, v2.1. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), accessed 8 October 2020, https://doi.org/10.24381/cds.a4fdd6b9.

——, ——, H. Cloke, P. Salamon, and C. Prudhomme, 2020a: Daily ensemble river discharge reforecasts and real-time forecasts from the operational Global Flood Awareness System. *Hydrol. Earth Syst. Sci. Discuss.*, https://doi.org/10.5194/hess-2020-532.

——, and Coauthors, 2020b: GloFAS-ERA5 operational global river discharge reanalysis 1979–present. *Earth Syst. Sci. Data*, **12**, 2043–2060, https://doi.org/10.5194/essd-12-2043-2020.

Unauthenticated | Downloaded 06/29/23 02:14 PM UTC

——, H. Cloke, and F. Pappenberger, 2020c: Innovating global hydrological prediction through an Earth system approach. *WMO Bull.*, **69**, https://public.wmo.int/en/resources/bulletin/innovating-global-hydrological-prediction-through-earth-system-approach.

Heming, J. T., 2017: Tropical cyclone tracking and verification techniques for Met Office numerical weather prediction models. *Meteor. Appl.*, **24**, 1–8, https://doi.org/10.1002/met.1599.

Hernández Ayala, J. J., and C. J. Matyas, 2016: Tropical cyclone rainfall over Puerto Rico and its relations to environmental and storm-specific factors. *Int. J. Climatol.*, **36**, 2223–2237, https://doi.org/10.1002/joc.4490.

Hersbach, H., and Coauthors, 2018a: ERA5 hourly data on single levels from 1979 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), accessed 8 October 2020, https://doi.org/10.24381/cds.adbb2d47.

——, and Coauthors, 2018b: Operational global reanalysis: Progress, future directions and synergies with NWP. ECMWF ERA Rep. Series 27, 63 pp., https://doi.org/10.21957/tkic6g3wm.

——, and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hirpa, F. A., P. Salamon, H. E. Beck, V. Lorini, L. Alfieri, E. Zsoter, and S. J. Dadson, 2018: Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data. *J. Hydrol.*, **566**, 595–606, https://doi.org/10.1016/j.jhydrol.2018.09.052.

Huffman, G. J., E. F. Stocker, D. T. Bolvin, E. J. Nelkin, and T. Jackson, 2019: GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree × 0.1 degree V06. Goddard Earth Sciences Data and Information Services Center (GES DISC), accessed 6 July 2020, https://doi.org/10.5067/GPM/IMERG/3B-HH/06.

Jiang, H., C. Liu, and E. J. Zipser, 2011: A TRMM-based tropical cyclone cloud and precipitation feature database. *J. Appl. Meteor. Climatol.*, **50**, 1255–1274, https://doi.org/10.1175/2011JAMC2662.1.

Jiang, Q., and Coauthors, 2021: Evaluation of the ERA5 reanalysis precipitation dataset over Chinese Mainland. *J. Hydrol.*, **595**, 125660, https://doi.org/10.1016/j.jhydrol.2020.125660.

Khouakhi, A., G. Villarini, and G. A. Vecchi, 2017: Contribution of tropical cyclones to rainfall at the global scale. *J. Climate*, **30**, 359–372, https://doi.org/10.1175/JCLI-D-16-0298.1.

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone best track data. *Bull. Amer. Meteor. Soc.*, **91**, 363–376, https://doi.org/10.1175/2009BAMS2755.1.

——, H. J. Diamond, J. P. Kossin, M. C. Kruk, and C. J. Schreck III, 2018: International Best Track Archive for Climate Stewardship (IBTrACS) Project, version 4 (subset: since 1980). NOAA/National Centers for Environmental Information, accessed 28 September 2020, https://doi.org/10.25921/82ty-9e16.

Knutson, T. R., and Coauthors, 2010: Tropical cyclones and climate change. *Nat. Geosci.*, **3**, 157–163, https://doi.org/10.1038/ngeo779.

Kossin, J. P., 2018: A global slowdown of tropical-cyclone translation speed. *Nature*, **558**, 104–107, https://doi.org/10.1038/s41586-018-0158-3.

Lavers, D., S. Harrigan, E. Andersson, D. S. Richardson, C. Prudhomme, and F. Pappenberger, 2019: A vision for improving global flood forecasting. *Environ. Res. Lett.*, **14**, 121002, https://doi.org/10.1088/1748-9326/ab52b2.

Luitel, B., G. Villarini, and G. A. Vecchi, 2018: Verification of the skill of numerical weather prediction models in forecasting rainfall from U.S. landfalling tropical cyclones. *J. Hydrol.*, **556**, 1026–1037, https://doi.org/10.1016/j.jhydrol.2016.09.019.

Merrill, R. T., 1984: A comparison of large and small tropical cyclones. *Mon. Wea. Rev.*, **112**, 1408–1418, https://doi.org/10.1175/1520-0493(1984)112<1408:ACOLAS>2.0.CO;2.

Nogueira, M., 2020: Inter-comparison of ERA-5, ERA-Interim and GPCP rainfall over the last 40 years: Process-based analysis of systematic and random differences. *J. Hydrol.*, **583**, 124632, https://doi.org/10.1016/j.jhydrol.2020.124632.

Prat, O. P., and B. R. Nelson, 2013: Mapping the world's tropical cyclone rainfall contribution over land using the TRMM Multi-satellite Precipitation Analysis. *Water Resour. Res.*, **49**, 7236–7254, https://doi.org/10.1002/wrcr.20527.

——, and ——, 2016: On the link between tropical cyclones and daily rainfall extremes derived from global satellite observations. *J. Climate*, **29**, 6127–6135, https://doi.org/10.1175/JCLI-D-16-0289.1.

Qiu, W., F. Ren, L. Wu, L. Chen, and C. Ding, 2019: Characteristics of tropical cyclone extreme precipitation and its preliminary causes in Southeast China. *Meteor. Atmos. Phys.*, **131**, 613–626, https://doi.org/10.1007/s00703-018-0594-5.

Rappaport, E. N., 2000: Loss of life in the United States associated with recent Atlantic tropical cyclones. *Bull. Amer. Meteor. Soc.*, **81**, 2065–2074, https://doi.org/10.1175/1520-0477(2000)081<2065:LOLITU>2.3.CO;2.

——, 2014: Fatalities in the United States from Atlantic tropical cyclones: New data and interpretation. *Bull. Amer. Meteor. Soc.*, **95**, 341–346, https://doi.org/10.1175/BAMS-D-12-00074.1.

Rezapour, M., and T. E. Baldock, 2014: Classification of hurricane hazards: The importance of rainfall. *Wea. Forecasting*, **29**, 1319–1331, https://doi.org/10.1175/WAF-D-14-00014.1.

Rogers, R., F. Marks, and T. Marchok, 2009: Tropical cyclone rainfall. *Encyclopedia of Hydrological Sciences*, M. G. Anderson and J. J. McDonnell, Eds., John Wiley and Sons, https://doi.org/10.1002/0470848944.hsa030.

Saha, K. K., M. M. Hasan, and A. Quazi, 2015: Forecasting tropical cyclone-induced rainfall in coastal Australia: Implications for effective flood management. *Austral. J. Environ. Manage.*, **22**, 446–457, https://doi.org/10.1080/14486563.2015.1028109.

Shi, C., J. Wu, and L. Qi, 2017: Analysis on precipitation difference of two typhoons with similar tracks. *J. Mar. Meteor.*, **37**, 36–45.

Song, J. Y., A. Alipour, H. Moftakhari, and H. Moradkhani, 2020: Toward a more effective hurricane hazard communication. *Environ. Res. Lett.*, **15**, 064012, https://doi.org/10.1088/1748-9326/ab875f.

Stein, R. M., L. Dueñas-Osorio, and D. Subramanian, 2010: Who evacuates when hurricanes approach? The role of risk, information, and location. *Soc. Sci. Quart.*, **91**, 816–834, https://doi.org/10.1111/j.1540-6237.2010.00721.x.

Stephens, E., J. J. Day, F. Pappenberger, and H. Cloke, 2015: Precipitation and floodiness. *Geophys. Res. Lett.*, **42**, 10 316–10 323, https://doi.org/10.1002/2015GL066779.

Sturdevant-Rees, P., J. A. Smith, J. Morrison, and M. L. Baeck, 2001: Tropical storms and the flood hydrology of the central Appalachians. *Water Resour. Res.*, **37**, 2143–2168, https://doi.org/10.1029/2000WR900310.

Takemi, T., and T. Unuma, 2020: Environmental factors for the development of heavy rainfall in the eastern part of Japan during Typhoon Hagibis (2019). *SOLA*, **16**, 30–36, https://doi.org/10.2151/sola.2020-006.

Tarek, M., F. P. Brissette, and R. Arsenault, 2020: Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrol. Earth Syst. Sci.*, **24**, 2527–2544, https://doi.org/10.5194/hess-24-2527-2020.

Unauthenticated | Downloaded 06/29/23 02:14 PM UTC

Titley, H. A., R. L. Bowyer, and H. L. Cloke, 2020: A global evaluation of multi-model ensemble tropical cyclone track probability forecasts. *Quart. J. Roy. Meteor. Soc.*, **146**, 531–545, https://doi.org/10.1002/qj.3712.

Touma, D., S. Stevenson, S. J. Camargo, D. E. Horton, and N. S. Diffenbaugh, 2019: Variations in the intensity and spatial extent of tropical cyclone precipitation. *Geophys. Res. Lett.*, **46**, 13 992–14 002, https://doi.org/10.1029/2019GL083452.

Towner, J., A. Ficchí, H. L. Cloke, J. Bazo, E. Coughlan de Perez, and E. M. Stephens, 2020: Influence of ENSO and tropical Atlantic climate variability on flood characteristics in the Amazon basin. *Hydrol. Earth Syst. Sci. Discuss.*, https://doi.org/10.5194/hess-2020-580.

Van Der Knijff, J. M., J. Younis, and A. P. J. De Roo, 2010: LISFLOOD: A GIS-based distributed model for river basin scale water balance and flood simulation. *Int. J. Geogr. Info. Sci.*, **24**, 189–212, https://doi.org/10.1080/13658810802549154.

Villarini, G., R. Goska, J. A. Smith, and G. A. Vecchi, 2014: North Atlantic tropical cyclones and U.S. flooding. *Bull. Amer. Meteor. Soc.*, **95**, 1381–1388, https://doi.org/10.1175/BAMS-D-13-00060.1.

Weber, H. C., C. F. Lok, N. E. Davidson, and Y. Xiao, 2014: Objective estimation of the radius of the outermost closed isobar in tropical cyclones. *Trop. Cyclone Res. Rev.*, **3**, 1–21, https://doi.org/10.6057/2014TCRR01.01.

Whitehead, J. C., R. Edwards, M. Van Willigen, J. R. Maiolo, K. Wilson, and K. T. Smith, 2000: Heading for higher ground: Factors affecting real and hypothetical hurricane evacuation behavior. *Global Environ. Change*, **2B**, 133–142, https://doi.org/10.1016/S1464-2867(01)00013-4.

Yu, Z., Y. Wang, H. Xu, N. Davidson, Y. Chen, Y. Chen, and H. Yu, 2017: On the relationship between intensity and rainfall distribution in tropical cyclones making landfall over China. *J. Appl. Meteor. Climatol.*, **56**, 2883–2901, https://doi.org/10.1175/JAMC-D-16-0334.1.

Yuan, S., L. Zhu, and S. M. Quiring, 2021: Comparison of two multisatellite algorithms for estimation of tropical cyclone precipitation in the United States and Mexico: TMPA and IMERG. *J. Hydrometeor.*, **22**, 923–939, https://doi.org/10.1175/JHM-D-19-0296.1.

Zhou, Y., and C. J. Matyas, 2017: Spatial characteristics of storm-total rainfall swaths associated with tropical cyclones over the eastern United States. *Int. J. Climatol.*, **37**, 557–569, https://doi.org/10.1002/joc.5021.

Zsótér, E., H. L. Cloke, C. Prudhomme, S. Harrigan, P. de Rosnay, J. Muñoz-Sabater, and E. Stephens, 2020: Trends in the GloFAS-ERA5 river discharge reanalysis. ECMWF Tech. Memo. 871, 73 pp., https://www.ecmwf.int/sites/default/files/elibrary/2020/19762-trends-glofas-era5-river-discharge-reanalysis.pdf.

Unauthenticated | Downloaded 06/29/23 02:14 PM UTC

A4: Submitted Article: Using ensembles to analyse predictability links in the tropical cyclone flood forecast chain

This appendix contains a formatted version of chapter 6 of this thesis that was submitted to the Journal of Hydrometeorology following reviews, with the following reference:

Titley, H.A., Cloke, H.L., Stephens, E.M., Pappenberger, F. and Zsótér, E. (2023) Using ensembles to analyse predictability links in the tropical cyclone flood forecast chain. Journal of Hydrometeorology (re-submitted June 2023 following review).

Using ensembles to analyse predictability links in the tropical cyclone flood forecast chain

H. A. Titley [a,b*], H. L. Cloke [b], E. M. Stephens [b,c], F. Pappenberger [d], and E. Zsoter [d]

[a] *Met Office, Exeter, United Kingdom*

[b] *University of Reading, Reading, Berkshire, United Kingdom*

[c] *Red Cross Red Crescent Climate Centre, The Hague, Netherlands*

[d] *ECMWF, Reading, Berkshire, United Kingdom*

Corresponding author: Helen Titley, helen.titley@metoffice.gov.uk

File generated with AMS Word template 2.0

ABSTRACT

Fluvial flooding is a major cause of death and damages from tropical cyclones (TCs), so it is important to understand the predictability of river flooding in TC cases, and the potential of global ensemble flood forecast systems to inform warning and preparedness activities. This paper demonstrates a methodology using ensemble forecasts to follow predictability and uncertainty through the forecast chain in the Global Flood Awareness System (GloFAS), to explore the connections between the skill of the TC track, intensity, precipitation and river discharge forecasts. Using the case of Hurricane Iota, which brought severe flooding to Central America in November 2020, we assess the performance of each ensemble member at each stage of the forecast, along with the overall spread and change between forecast runs, and analyse the connections between each forecast component. Strong relationships are found between track, precipitation and river discharge skill. Changes in TC intensity skill only result in significant improvements in discharge skill in river catchments close to the landfall location that are impacted by the heavy rains around the eye wall. The rainfall from the wider storm circulation is crucial to flood impacts in most of the affected river basins, with a stronger relationship with the post-landfall track error rather than the precise landfall location. We recommend the wider application of this technique in TC cases, to investigate how this cascade of predictability varies with different forecast and geographical contexts, to help inform flood early warning in TCs.

SIGNIFICANCE STATEMENT

This study demonstrates a methodology to analyse the cascade of predictability and uncertainty through the various stages of the tropical cyclone (TC) flood forecasting chain, illustrating how it can provide useful information to modellers interested in optimising flood forecast skill, and to those who prepare and communicate flood forecasts with stakeholders and end-users in TC cases. The results highlight the importance of improving verification of ensemble TC precipitation forecasts, and of focusing on more than just the category of the storm and landfall location when forecasting and communicating flood impacts in TC cases.

File generated with AMS Word template 2.0

1. Introduction

Despite the major role of rain-induced flooding in deaths and damages related to tropical cyclones (TCs) (Rappaport, 2014; Czajkowski and Kennedy, 2010; Czajkowski et al., 2013) and predictions that TC rainfall rates and associated flooding impacts will increase due to climate change (Knutson et al., 2010; 2013; 2015; 2020; Wang et al., 2015; Wright et al., 2015), the main determining factor in evacuation decision-making remains the intensity of the TC (Stein et al., 2010; Whitehead et al., 2000; Senkbeil et al., 2019). The potential for heavy rainfall is present regardless of the storm's category (Titley et al., 2021) so it is vital to increase public awareness of the dangers of river flooding in TCs. The importance of increasing the focus on the probabilistic forecasting of downstream hazards has been recognised by key international research programmes (Golding, 2022; Titley et al., 2019). Improving the understanding of the factors that influence the predictability of flood-related hazards during TCs can help optimise forecast guidance and therefore improve preparedness in advance of an approaching storm. However, investigating this predictability is hampered by a limited sample size of forecast TCs affecting a single region, and the difficulties of aggregating across regions due to geographical differences (e.g. river length and orientation, topography). Ensemble forecasts have the potential to help overcome this issue by providing a larger sample of potential outcomes for what are rare events in any given location.

This work therefore aims to develop a novel methodology to use ensemble forecasts to explore how each part of the TC flood forecast chain (track, intensity, precipitation, and hydrology) influences the overall predictability of river flooding in TC cases. The ensemble forecasts are examined across several forecast runs to create hundreds of sample forecasts for a particular high-impact case. Taking inspiration from previous ensemble-based sensitivity analysis (Hakim & Torn, 2008; Lynch & Schumacher, 2014; Zhang & Meng, 2018; Shen et al., 2020), in this study we develop methods to use the ensemble forecasts to investigate the interconnectivity of the various forecast components of the TC flood forecast chain and give valuable insights as to the controls on TC flood predictability. The following questions will be addressed: i) How do the errors and the spread in the TC track and intensity forecasts relate to the errors and spread in the precipitation forecasts and the river discharge forecasts?; and ii) Do the members with the best/worst track forecasts necessarily lead to good/poor precipitation and river discharge forecasts?

72 The forecast system selected for use in this paper is the Global Flood Awareness System
73 (GloFAS, www.globalfloods.eu), an operational system of the European Commission's
74 Copernicus Emergency Management Service, which is used for producing flood forecast
75 bulletins for humanitarian users for TCs (e.g. Emerton et al., 2020; Speight et al., 2022).
76 GloFAS produces probabilistic river discharge forecasts driven by ensemble weather
77 forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF),
78 providing early indication of upcoming flood events. The 51 ensemble members within the
79 GloFAS forecasting system can be traced through the complete forecast chain from TC track
80 and intensity, through to their rainfall and river discharge forecasts, and over several forecast
81 runs, providing a larger sample size of forecasts that can be used to assess the relationships
82 between each forecast component. Using the GloFAS forecasting system in this way can also
83 allow an initial assessment of the lead times at which it can provide useful guidance for
84 potential flooding in TC cases, and what forecast and geographical contexts can cause
85 variations in this. The methods are demonstrated using the case of Hurricane Iota which
86 affected over 5 million people in Central America in November 2020 (UN-OCHA, 2020a;
87 2020b).

88 **2. Case background**

89 Hurricane Iota was a strong category 4 hurricane on the Saffir-Simpson scale, that caused
90 severe and widespread river flooding in Nicaragua, Honduras and Guatemala, resulting in at
91 least 67 deaths and total damages estimated at $1.4 billion (Stewart, 2021). Iota originated as
92 a tropical wave that moved into the Eastern Caribbean on 10th November 2020 before
93 strengthening and being named Tropical Storm Iota on 13th November. Iota became
94 hurricane-strength on 15th November, and rapidly intensified to a Category 4 hurricane, with
95 peak maximum wind speeds of 135 kt, before making landfall with an intensity of 125 kt on
96 17 November at 03:40 UTC, south of Puerto Cabezas in northeast Nicaragua. Several sites
97 across Nicaragua and Honduras reported event rainfall totals over 200mm, with a maximum
98 total of 510mm in Guatemala (Stewart, 2021). The extreme rainfall led to estimated floods
99 exceeding the 1 in 20-year thresholds across major rivers in these countries (Figure 1), as
100 well as widespread flash flooding and mudslides. Flooding was exacerbated by pre-existing
101 flood conditions caused by Hurricane Eta, which had made landfall in a similar location two
102 weeks earlier. In total 5.2 million people were affected with widespread damage to properties
103 and infrastructure (IFRC, 2020; UN OCHA, 2020a; UN OCHA, 2020b).

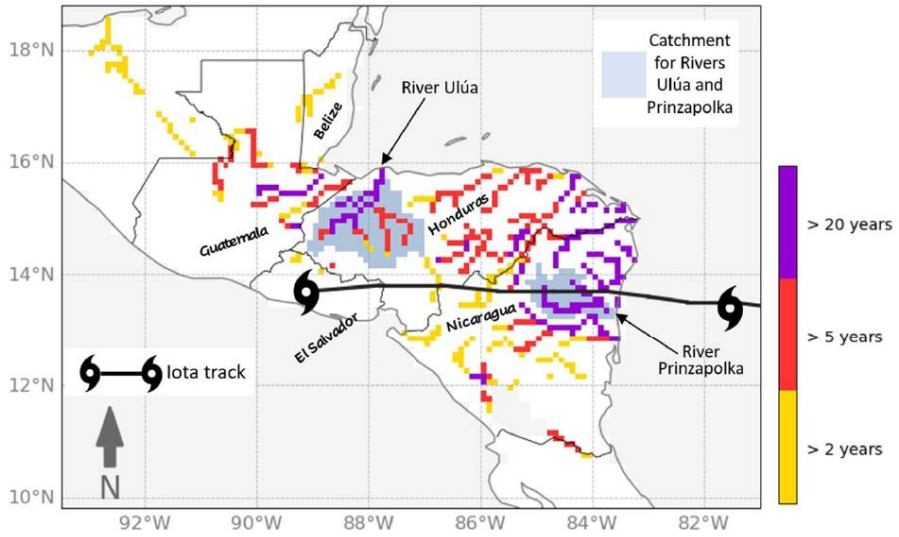


Figure 1: The return period estimated to have been exceeded from 16th-20th November 2020 in the GloFAS-ERA5 reanalysis. Two of the most severely affected rivers are labelled (the Ulúa and the Prinzapolka, for which the forecasts are examined in Section 4), with their catchment areas shown in light blue.

3. Data and Methods

The GloFAS forecast system is driven by ECMWF ensemble (ENS) forecasts. The ECMWF ENS consists of an ensemble of 51 forecasts with a current spatial resolution of 18km; one unperturbed control member and 50 perturbed members whose initial states and model physics have been perturbed to explore the currently understood range of uncertainty in the observations and the model. The initial condition uncertainty is sampled using perturbations from an ensemble of data assimilations (EDA) and perturbations constructed from the leading singular vectors, while the model uncertainties are represented with the Stochastically Perturbed Parameterization Tendencies scheme (SPPT) that simulates the effect on forecast uncertainty of random model errors due to the parametrized physical processes (ECMWF, 2020), The ensemble members provide a range of possible future weather states, with the methods designed to ensure that each ENS member is truly independent of all the others (Owens and Hewson, 2018).

From the set-up of the GloFAS forecast system it is possible to assess the performance of each component of the flood forecast chain (track, intensity, precipitation and discharge)

within each ensemble member from each successive forecast in the lead up to the flood event. Comparing the performance of the ensemble forecasts between components and from run to run allows for an assessment of the links between the predictability of each of the different forecast components (Figure 2), with a particular focus on how the discharge forecast skill relates to the errors in the TC track, intensity, and precipitation forecasts. Table 1 summarizes the forecast and verification data that was sourced for each of the components in the forecast chain. It also describes the verification metric that was selected and calculated for each ensemble member from each forecast run. Further detail on each component and the evaluation of the predictability linkages is given in the following sections.

a. Tropical cyclone track (position and intensity)

1) FORECAST DATA

The forecasted TC track position and intensity information from each of the 51 ECMWF ENS ensemble members were extracted from the TIGGE CXML archive (NCEP et al., 2008), for each 00 UTC forecast run from 8th to 17th November 2020.

2) VERIFICATION DATA AND METHOD

The verification data for TC track position and intensity for Hurricane Iota were obtained from the International Best Track Archive for Climate Stewardship (IBTrACS) (Knapp et al., 2010; 2018). Forecast tropical cyclone tracks are deemed to be matched with Hurricane Iota if any point on the track is within 500km of a matching observed position from IBTrACS. All points from that track are then included in the analysis providing their forecast position is within 1,000km of the matching observed position. The TC track error in each of the ensemble members with a matching tropical cyclone was then calculated as follows: i) the landfall track error, defined as the distance between the observed and forecast position, in km, at the time of the last pre-landfall observed track point (00 UTC 17th November 2020); ii) the average track error across land; in the Iota case this is defined as the average of the distances (in km) from each of the 6-hourly forecast and observed track positions in the first 24 hours that the storm was over land, to ensure that storm tracks are available for most members. The intensity error calculated is the error in mean sea level pressure (mslp) minima at the last pre-landfall point, prior to any weakening on landfall.

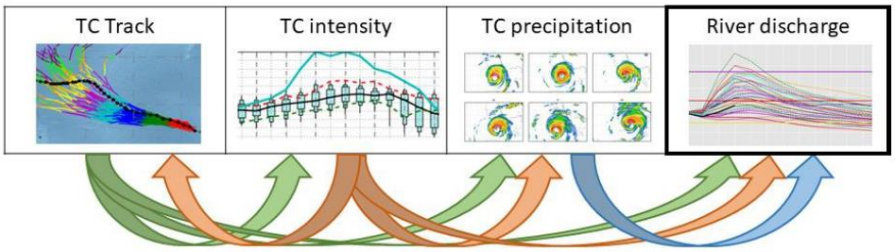


Figure 2: Conceptual diagram illustrating four of the main components of the TC river flood forecast process, along with the potential links in predictability between them. The boxes contain the four aspects of the forecast that are verified in each ensemble member in each forecast run in the lead up to the flood event, while the arrows represent the potential of each component to influence the skill of the others (green arrows = from TC track; orange arrows = from TC intensity; blue arrows = from TC precipitation). It is these links and dependencies that are evaluated in this paper (e.g. each arrow has a matching scatter plot in Figure 8).

Table 1: A summary of the forecast data, verification data and metrics computed for each component of the TC flood forecast chain.

	TC Track	TC intensity	TC precipitation	River discharge
Forecast data	ECMWF ENS tracks from TIGGE CXML archive (NCEP et al., 2008)	ECMWF ENS tracks from TIGGE CXML archive (NCEP et al., 2008)	ECMWF ENS precipitation forecast data (0.2°x0.2° resolution) (ECMWF, 2022)	Global Flood Awareness System (GloFAS) (v2.1, 0.1°x0.1° resolution), (Zsoter et al., 2019)
Verification data	International Best Track Archive for Climate Stewardship (IBTrACS) (Knapp et al., 2010; 2018)	International Best Track Archive for Climate Stewardship (IBTrACS) (Knapp et al., 2010; 2018)	GPM IMERG (final version, 0.1°x0.1° resolution) precipitation data (Huffman et al., 2019)	GloFAS-ERA5 operational global river discharge reanalysis (0.1°x0.1° resolution) (Harrigan et al., 2019; 2020)
Verification metric	• Track error at landfall • Track error in the first 24 hours over land	Error in mean sea level pressure minima at landfall	Mean absolute error of storm total precipitation	Modified Kling-Gupta efficiency (KGE') (Kling et al., 2012) computed for key river points and across all severely impacted river points

File generated with AMS Word template 2.0

168 *b. Tropical cyclone precipitation*

169 1) FORECAST DATA

170 ECMWF ENS precipitation forecast data for each 00UTC forecast run from 8th to 17th
171 November 2020 was extracted from the ECMWF archive (ECMWF, 2022). The storm total
172 precipitation was calculated as being the 96-hour accumulation from 00 UTC 16th to 00 UTC
173 20th November within 500km of the Hurricane Iota track. This represents the area of
174 precipitation defined as being associated with the cyclone, as used in multiple studies that
175 assign a set radius when defining TC-related rainfall (Prat and Nelson, 2013, 2016; Jiang et
176 al., 2011; Luitel et al., 2018; Titley et al., 2021). A land/sea mask was applied to focus on
177 precipitation over land, as relevant for river flooding.

178 2) VERIFICATION DATA AND METHOD

179 The observed precipitation for the matching storm total 96-hour period was calculated
180 from the Integrated Multi-satellitE Retrievals for GPM (IMERG) (Huffman et al., 2019),
181 selecting the IMERG Final Run data. The IMERG Final Run product, recommended for
182 research uses, is available around 3 months after valid time and incorporates additional
183 microwave passes and a monthly gauge correction. The same masks (land/sea and within 500
184 km of the observed track) were applied as in the forecast fields. Prior to verification, the
185 IMERG data is re-gridded to the model grid using linear interpolation. The Mean Absolute
186 Error (MAE) was then calculated for each ensemble member by comparing the storm total
187 precipitation in the forecasted and observed fields.

188 *c. River discharge*

189 1) FORECAST DATA

190 The set-up of the operational GloFAS forecast at the time of Hurricane Iota (v2.1)
191 GloFAS (Alfieri et al., 2013; Harrigan et al., 2020a), utilised ECMWF ENS meteorological
192 forecasts coupled with the land surface module HTESSEL (Balsamo et al., 2011) and the
193 Lisflood hydrological routing model (Van Der Knijff et al., 2010), to produce 51 possible
194 forecast evolutions of the streamflow through the global river network. GloFAS version 2.1
195 forecasts (Zsoter et al., 2019) were downloaded from the Copernicus Climate Change Service
196 (C3S) Climate Data Store (CDS) for each 00 UTC forecast run from 8th to 17th November
197 2020.

File generated with AMS Word template 2.0

198 2) VERIFICATION DATA AND METHOD

199 Ongoing difficulties in the availability of reliable, spatially distributed, river discharge
200 observations from river gauges (Lavers et al., 2019) mean that reanalyses provide the only
201 practical source of river discharge verification data for the purposes of this study. Therefore,
202 the verification dataset utilised in this study is the GloFAS-ERA5 operational global river
203 discharge reanalysis which accompanies GloFAS v2.1 (Harrigan et al., 2019; 2020b). An
204 objective analysis of the GloFAS forecasts is carried out by calculating the Modified Kling-
205 Gupta efficiency (KGE') (Gupta et al., 2009; Kling et al., 2012) to compare the forecast and
206 verification discharge data. The KGE' is an expression of the distance away from ideal model
207 performance as described by three components (correlation, variability bias and mean bias).
208 KGE' = 1 indicates perfect agreement between simulations and observations. The KGE' score
209 for a mean flow benchmark is KGE'≈-0.41 (Knoben et al., 2019). The KGE' is calculated for
210 the 1 to 15 day forecasts from all GloFAS ensemble members for key river points and for all
211 severely impacted river points across the region (defined as those points with an upstream
212 area greater than 1,000km^2 that experienced greater than a 5-year return period during the
213 event). The return period thresholds used in the study were as adopted within the GloFAS
214 web interface, calculated by fitting a Gumbel extreme value distribution to the GloFAS-
215 ERA5 river discharge reanalysis over the 1979-2018 period (Alfieri et al., 2019; Harrigan et
216 al., 2020a; Zsoter et al., 2020).

217 *d. Proposed methodology to evaluate the predictability linkages along the forecast chain*

218 Once all of the forecast components have been verified for each ensemble member at
219 each forecast run, the key part of the proposed methodology is to use this information to
220 analyse the links between forecast track, intensity, precipitation and discharge skill across
221 hundreds of sample forecasts for a particular high-impact case; in this case Hurricane Iota.
222 Several steps are carried out as follows:

223 i. Box and whisker plots are used to represent how the track, intensity and precipitation
224 errors and the discharge KGE' scores (across all ensemble members) compare in terms
225 of their mean and spread for each forecast run in the lead up to the event.

226 ii. A “divergence” score (Richardson et al., 2020) is also calculated to allow a comparison
227 of the run-to-run differences between each component of the forecast. In this score the
228 difference between two ensembles is calculated using the divergence function

File generated with AMS Word template 2.0

associated with the Continuous Ranked Probability Score (CRPS) (Gneiting and Raftery, 2007). In our case the two ensembles we are comparing are the ensemble member scores from each component on two successive runs, with the divergence score indicating how much the forecast performance is changing from run to run.

iii. Comparing the divergence scores from each part of the forecast (the track, intensity, precipitation and discharge components) allows us to assess how closely matched the run-to-run jumps in forecast performance are between each component.

iv. The correlation between the individual ensemble member scores for each component from each forecast lead time is also calculated using Spearman Rank Correlation, and displayed alongside scatter plots to show how closely each aspect of forecast performance is tied to each other.

v. 'Good' and 'bad' members are identified from their performance in each part of the forecast chain and their performance in other components examined, in order to delve deeper into the controls on predictability and how these vary from place to place and from run to run.

4. Results

a. Evolution of forecasts for Iota

A summary of the TC position and intensity, precipitation and discharge forecasts for Hurricane Iota is shown in Figure 3. At seven days ahead of landfall around two thirds of members are forecasting a TC to develop, but with a very large spread in forecast position. By five days ahead of landfall over 90% of members are forecasting a TC, and by three days ahead the forecast storm location is much more confident. The intensity of Hurricane Iota is underpredicted at all lead times, as the forecast models struggled to resolve the rapid intensification process. The observed mean sea level pressure (mslp) minima at the pre-landfall point was 918mb, but even at a one-day lead time the minimum mslp forecasted by any of the ensemble members was 951mb (Figure 3 top row). Despite this, the heavy rainfall along the north coast of Honduras is predicted with high probabilities (> 60%) around 5 days ahead of landfall (Figure 3 middle row), but heavy rainfall close to the landfall location on the east coast of Nicaragua was not predicted with high probabilities until later forecast runs closer to the time of landfall. This suggests that the heavy rainfall close to the landfall location was more reliant on having a more accurate forecast for the landfall location and

intensity, which was only possible at shorter lead times. The northward bias of some earlier runs (see top row Figure 3) still produced forecasts of significant rainfall for Northern Honduras.

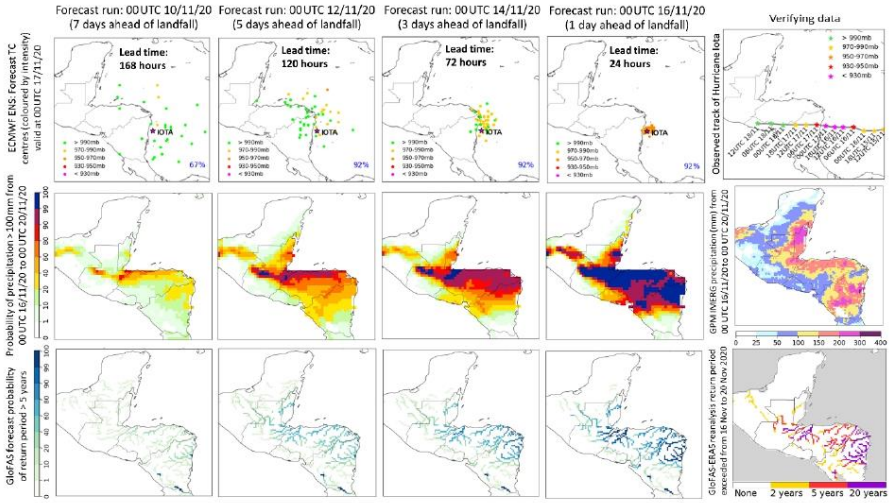


Figure 3: Hurricane Iota cyclone position (top row), rainfall (middle row) and flood (bottom) forecasts from the forecast runs of ECMWF ENS and GloFAS in the lead up to the event. The first four columns show the forecast runs from 7, 5, 3 and 1 days prior to landfall, while the right column shows the corresponding verification data for track, rainfall, and flood return period exceedance. In the top row the star shows the observed centre position of TC Iota at 00UTC 17/11/2020 (just prior to landfall) and the dots show the forecast positions valid at this time from each ensemble member, coloured by the mean sea level pressure minima. The percentages indicate the percentage of members that had a matched cyclone centre forecast at this lead time. The second row shows the precipitation probability forecasts from EC ENS for 4-day event precipitation totals > 100mm. The bottom row shows the GloFAS forecast probability for river discharge exceeding the 5-year return period on each river point.

A signal for significant flooding appears 5 days ahead of landfall (Figure 3 bottom row), with probabilities of 5-year flood return periods being exceeded being greater than 40% across a large part of the region. This coincides with the increase in confidence in the TC position and rainfall accumulations. However, the forecast confidence varies from river to river, and even one day prior to landfall not all the areas simulated to have exceeded the 20-year return period (the highest threshold displayed in GloFAS) have high probabilities. For example the Río Grande de Matagalpa in Central Nicaragua (the most southerly river that experienced return periods greater than 20 years as shown in the bottom right plot in Figure 3) only has a probability of 40-60% of exceeding a 5 year return period, which is much lower than the other impacted rivers. Although the rainfall forecast one day prior to landfall is

correctly identifying two main areas of high precipitation, on the north coast of Honduras and near the landfall location in Nicaragua, the area near the forecast track is displaced too far to the north, and the forecast has missed the extension of the rainfall area to the SW of the TC track. These results indicate that the factors controlling the forecast skill for the flood forecasts for Iota vary with lead time and scale. Once the forecast confidence increased for the TC genesis and track of Iota (around 5 days ahead of landfall) there is a clear signal for a significant flood event in the overall region that was affected. At the catchment scale, the detailed structure and precipitation forecasts for the storm influence how well the flooding is forecast on each individual river, even at shorter lead times.

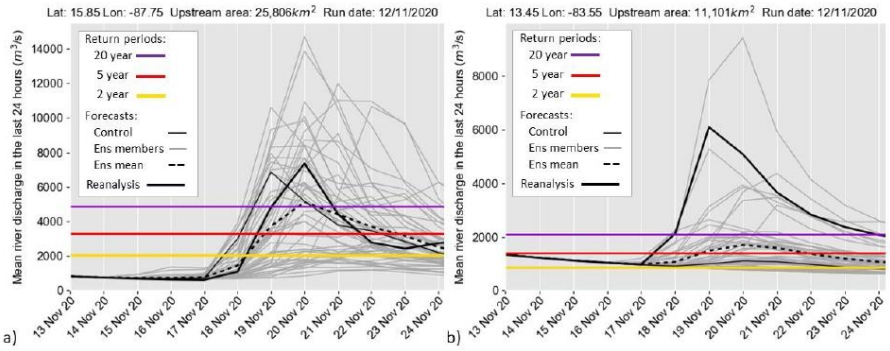


Figure 4: GloFAS forecast hydrographs from 00UTC run on 12th November 2020 for a) River Ulúa and b) River Prinzapolka (ensemble members: grey; control member: thin black; ensemble mean discharge forecast: black dashed). The yellow, red and purple horizontal lines show the 2-, 5- and 20-year return period discharge levels for that river. The reanalysis discharge is overlain in a thick black line.

Two of the rivers which experienced the most severe flooding from Hurricane Iota were the River Ulúa in northern Honduras and the River Prinzapolka in eastern Nicaragua. To examine the river discharge forecasts for these rivers, the GloFAS outlet river point was selected for each river: 15.85N, 87.75W for the Ulúa and 13.45N, 83.55W for Prinzapolka, with upstream areas of 25,806km^2 and 11,101km^2 respectively (see annotated map in Figure 1 for the locations and catchment areas). GloFAS forecast hydrographs for these points from 5 days prior to landfall are shown in Figure 4 and illustrate that the majority of members were predicting severe flooding on the Ulúa, whereas only a small number of members predicted the extreme discharge levels experienced on the Prinzapolka. This shows the sensitivity of rivers closer to the landfall location to the precise forecast of the landfall, as these were areas affected by the heavy rainfall close to the centre of the storm. The upstream area of the

311 Prinzapolka site is also smaller, resulting in a faster response to heavy rainfall compared to
312 the Ulúa site, and increased sensitivity to accurate timing of the landfall.

313 *b. Flow of uncertainty and predictability through the forecast chain*

314 The previous section presented the evolution of forecasts for Hurricane Iota, identifying
315 potential linkages between the different forecast components. To investigate the flow of
316 uncertainty and predictability more thoroughly, a detailed analysis comparing the ensemble
317 performance at each part of the forecast chain is proposed and illustrated. Forecast
318 performance across the ensemble members is presented side-by-side for each component of
319 the forecast: track error (Figure 5a), intensity error (Figure 5b), precipitation error (Figure 5c)
320 and KGE' score for the discharge forecasts (5d) across all severely impacted river points as
321 well as the two rivers that were examined in Figure 4 (the Ulúa and the Prinzapolka).
322 Flooding on the River Ulúa was well predicted further in advance than the flooding on the
323 River Prinzapolka, which was closer to the landfall location, confirming the findings in
324 Section 4a. Using the Towner et al. (2019) thresholds to evaluate the KGE' scores ("Good"
325 (KGE' $\geqslant$ 0.75), "Intermediate" (0.75 > KGE' $\geqslant$ 0.5), "Poor" (0.5 > KGE' > 0) and "Very poor"
326 (KGE' $\leqslant$ 0), the majority of ensemble members have "intermediate" performance 5 days
327 ahead of landfall for the River Ulúa, but only 2 days ahead of landfall for the River
328 Prinzapolka. An examination of IMERG satellite data (Figure 6) reveals that the Ulúa
329 catchment rainfall was mainly associated with onshore flow of the outer rainbands with
330 orographic enhancement, which fell over a longer period and is less sensitive to the exact
331 landfall location. For the River Prinzapolka, the centre of the storm passed directly over the
332 catchment leading to very heavy rainfall in a shorter period during and after landfall.
333 Successful flood forecasts therefore relied on the accurate prediction of landfall location and
334 of the heavy rainfall close to the TC centre, which was less predictable at longer lead times.
335 The average KGE' score across all impacted points (all river points where the 5-year return
336 period was exceeded in the reanalysis) falls between these two extremes.

File generated with AMS Word template 2.0

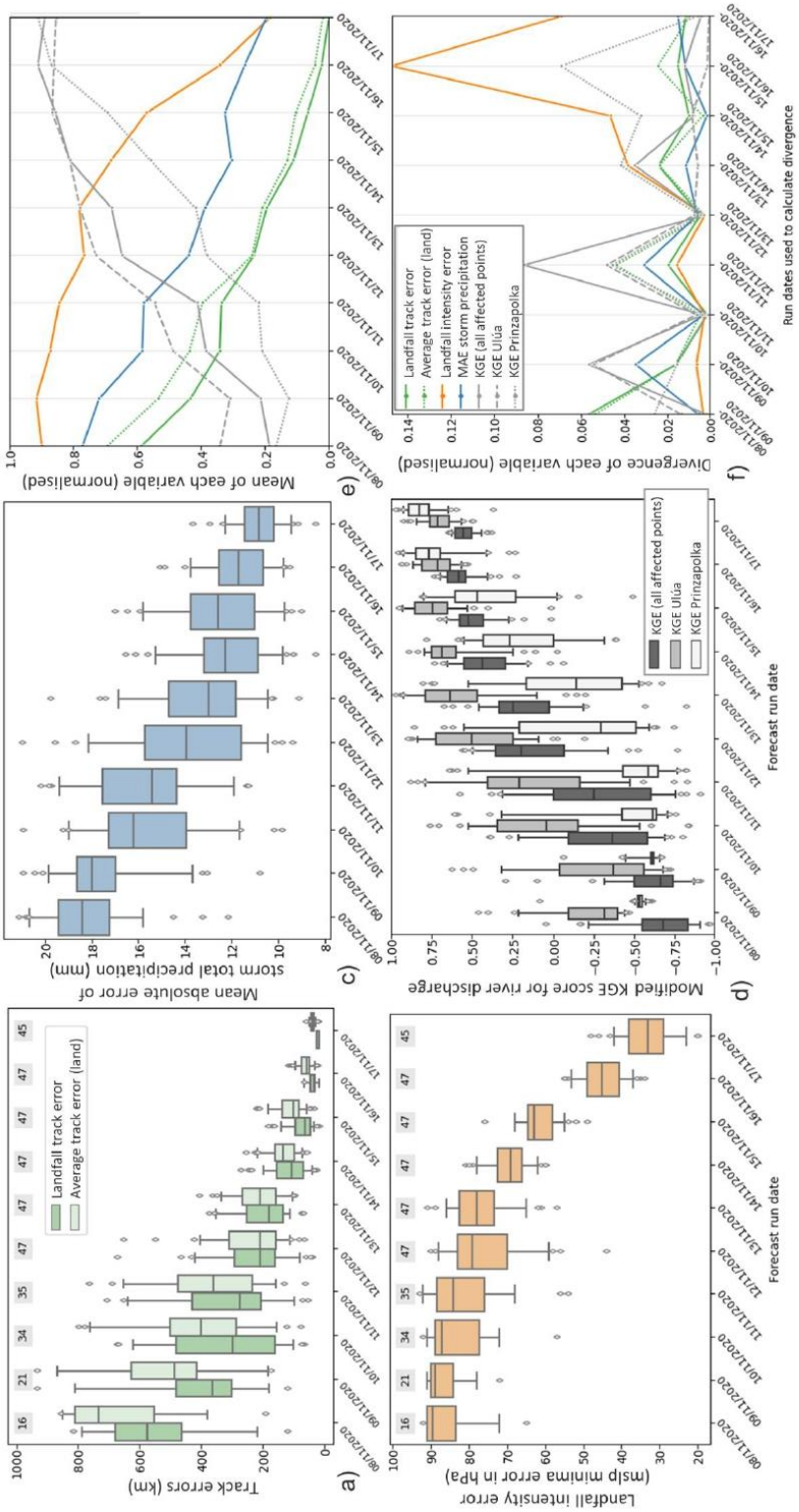


Figure 5: Box plots comparing forecast performance across the ensemble members for each 00UTC forecast run in the lead up to the Hurricane Iota event: a) forecast track error (landfall error and average track error over land); b) forecast landfall intensity (mslp minima) error; c) forecast precipitation error (mean absolute error of the storm total precipitation (within 500km of TC track); d) Discharge forecast performance as measured by the average KGE' score at all severely impacted river points (where 5-year return period is exceeded in the GloFAS reanalysis), and at the River Ulúa / River Prinzapolka outlet river points; e) a comparison of the means of these metrics, normalized by the y-axis range in each subplot and colour-coded from the plots and key as displayed in (f); f) divergence score of these metrics, normalized by the y-axis range and colour coded as shown in the key. In (a) and (b), errors can only be computed for those ensemble members with a tracked TC (numbers overlain in grey).

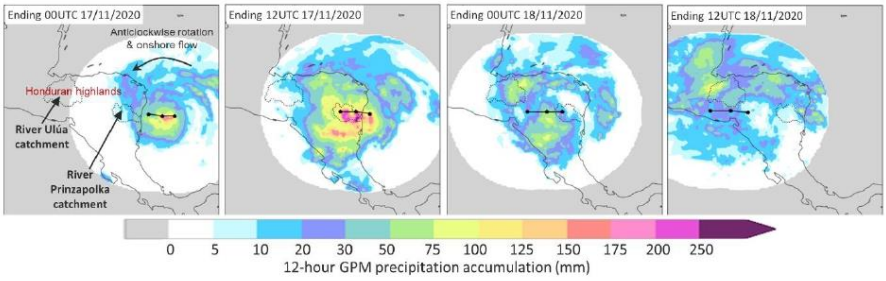


Figure 6: Precipitation accumulation across Central America from Hurricane Iota in four 12-hour periods from 12 UTC 16th November 2020 to 12 UTC 18th November 2020. The river catchments for the River Ulúa and River Prinzapolka are shown in dashed lines, with labels on the left plot.

The right hand two plots on Figure 5 then compare the means of these scores (Figure 5e), and the divergence score (i.e. change in performance from run to run (Figure 5f)). The idea is to objectively study how closely the skill of each part of the forecast is linked to the others. The biggest change in the KGE' of all river points (from 11th to 12th) coincides with changes in the precipitation, landfall position, and intensity errors, and in particular to the biggest change seen in the average track error over land, showing the importance of verifying and improving forecasts of TC track after landfall. Comparing the distribution of the track errors from 11th to 12th in 5a shows there was both an overall reduction in track error and a significant reduction in the spread. But although there are matching improvements in the precipitation and discharge scores, the spread is not significantly reduced and remains relatively high until very short lead times, showing there is greater variation in precipitation and discharge forecasts even after the track forecasts become more confident. For the River Prinzapolka, which was close to the landfall location, the biggest jump in KGE' scores coincided with the biggest change in landfall intensity error (from 15-16th, where the forecasts begin to pick up on the ongoing intensification of the storm). This large improvement in the intensity forecast in these short-range forecasts has only a modest impact on the track, precipitation and overall discharge forecast scores, but has a large impact for the KGE' score for the River Prinzapolka. As the River Prinzapolka catchment is much closer to the landfall point, the rainfall it experiences is predominately from the area close to the centre of the storm e.g. the eye wall, and the intensity of the precipitation in this part of the storm is greater for intense storms.

File generated with AMS Word template 2.0

c. Member-specific analysis

The consecutive forecast runs on 11th and 12th November were found in Figure 5f to have high divergence scores i.e. significant run-to-run changes in ensemble performance. The increase in forecast confidence from 11th to 12th November in Iota's position forecast the day after landfall is notable with a much more concentrated pattern of ensemble member TC positions (Figure 7). The additional member-specific analysis here illustrated in Figure 7 is to colour the forecast positions by their KGE' scores on two rivers points that have been focused on: the Ulúa (left) and the Prinzapolka (right) outlet points. This allows us to provide an impact-relevant hydrological assessment of which ensemble members provided the most accurate flood forecast for each river (those coloured in blue). Both before and after the increase in forecast confidence, the members with a good forecast for the River Prinzapolka all have the TC forecast centre relatively close to the observed location, although there are also several ensemble members with forecast centres close to the observed location that don't have good KGE scores. A broader range of ensemble members based on TC position perform well for the Ulúa catchment. The TC centres in Figure 7 are sized according to the MAE in storm total precipitation to allow an inspection of the relationship between three aspects of the forecast: the TC track (position), precipitation, and river discharge skill. The ensemble members with better overall precipitation forecasts (lower MAE), shown as larger dots, are mainly those that are forecasting the TC centre to be close to its observed position. Several of the ensemble members with TC positions too far to the north are shaded blue in the left plots as they score well for river discharge on the Ulúa, but they have a small size indicating they do not score well overall for precipitation. Their better score for the Ulúa discharge is partly due to their more northerly tracks bringing forecasts of heavy rain from the central part of the storm over the Ulúa catchment. The centre of the storm tracked further south in reality and so the overall precipitation errors in these members are higher as they fail to forecast the precipitation further south.

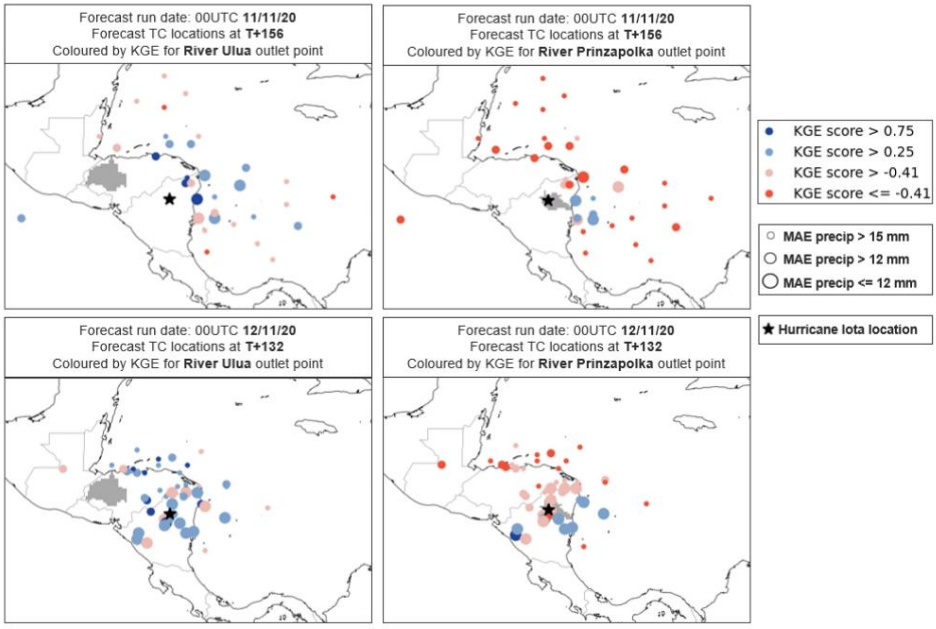
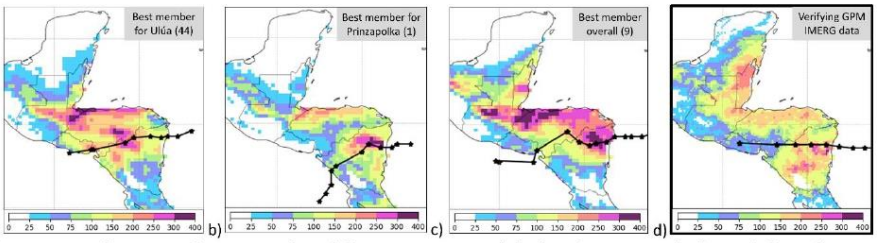


Figure 7: Forecast TC locations in the ECMWF ENS forecasts from 00UTC 11th November 2020 (top row) and 00UTC 12th November 2020 (bottom row). The centres are coloured according to the modified KGE' score for the River Ulúa (left column) and River Prinzapolka (right column) outlet river points as defined in Section 4a. The centres are sized by the mean absolute error in the storm total precipitation field (larger circles = better forecasts). The verifying position of Hurricane Iota at this time is marked with a black star, and the catchment areas for the two river outlet points are shown in grey for reference.

To take this member-specific analysis a step further, the discharge scores from each ensemble member from a particular run can be ranked to determine the 'best' performing members. This allows us to examine the other aspects of forecast performance for those members, to better understand why they led to good flood forecasts. This is illustrated for the 00UTC 12th November run in Figure 8. Interestingly in this case, the landfall position of all three of the identified 'best' members is similar, and is close to the observed landfall location. However, their forecast tracks after landfall are different, with the best performing member for the River Ulúa being closest to the observed track (Figure 8a), while the best member for the River Prinzapolka turns southwest shortly after landfall, and therefore underestimates precipitation totals further north (Figure 8b). Despite their similar landfall locations, the forecast precipitation patterns are very different, with large variation in the precipitation totals even close to the landfall location in all three. When comparing to the verifying IMERG data, the member that had the highest KGE' score when averaged over affected river

File generated with AMS Word template 2.0

418 points across the region (Figure 8c) had the best match in terms of the overall extent of the
419 rain event, but was forecasting significantly greater totals in several areas compared to
420 IMERG. However, rain totals of up to 510mm were recorded locally (Stewart, 2021),
421 indicating that IMERG may be underestimating the most extreme rainfall totals. All three of
422 the 'best' members from this forecast run failed to significantly intensify the storm. Member
423 9 (Figure 8c) was the strongest at landfall (986mb compared to 1001 and 1006mb
424 respectively for the other two 'best' members), which may have contributed to the larger
425 forecast rainfall totals nearer the landfall site, but it was still much weaker than observed
426 (918mb). This analysis shows that the ensemble forecasts of TC rainfall totals depend on
427 much more than having an accurate landfall location, and can vary significantly even between
428 members with similar track forecasts, emphasizing the need for an increased focus on
429 evaluating and understanding ensemble TC rainfall forecasts.
430



431
432 Figure 8: Hurricane Iota forecast track and forecast event precipitation forecast totals (in mm) from the
433 00UTC 12/11/2020 forecast run from the EC ENS ensemble members with the highest KGE' scores for a)
434 River Ulúa outlet point, b) River Prinzapolka outlet point, and c) the average of all severely impacted river
435 points; d) displays the verifying GPM IMERG precipitation (in mm) along with the observed track of
436 Hurricane Iota.
437

438 The correlation of the member scores between different aspects of forecast performance
439 helps to quantify the degree to which the forecast performance aspects are tied to each other
440 (Figure 9). The average track error over land was selected as the track measure to plot as
441 initial investigation (not shown) found a stronger correlation between this measure and both
442 the precipitation error and hydrology skill (KGE') compared to against the landfall error. This
443 further emphasises the earlier finding from Figure 5 of the importance of track after landfall
444 to the wider impacts in the region. Each scatter plot in Figure 9 includes the ensemble
445 members across all lead times allowing us to examine the link between each forecast
446 component using a large sample of hundreds of forecasts. All the relationships have a
447 significant correlation when all lead times are included (shown in blue), with the strongest

448 correlations being between the track error and the river discharge skill, and the precipitation
449 errors and the river discharge skill. Later forecasts have lower errors / greater skill, and
450 reduced spread in the scores, but the shape of the relationship varies considerably between
451 plots. For example, the points in the track vs discharge and rainfall vs discharge plots (Figure
452 9c and 9f) are mainly clustered along the diagonal, whereas in the intensity vs discharge plot
453 (Figure 9e) the points are spread across the upper right half of the plot. This indicates that
454 good forecasts of river discharge are possible even with high intensity errors, but not with
455 high rainfall or track errors.

456 Looking at the correlations at each forecast run time (in tables on Figure 9, with average
457 of these in green) shows the predictability links between each component for particular
458 forecast runs. The correlation between track and intensity error is not significant at any lead
459 time (Figure 9a). There are however significant correlations between track and precipitation,
460 and track and river discharge at all forecast times other than the latest two, by which point the
461 track errors were very small (Figure 9b and 9c). Interestingly, the correlation between the
462 track errors and river discharge skill is stronger than track errors against the precipitation
463 errors, which may be because discharge skill depends on the total catchment rainfall, where
464 the smaller grid point to grid point differences that influence the precipitation errors are
465 balanced out to some degree. The slightly lower correlation with precipitation may also be
466 due to the independence of the precipitation verification data, as opposed to the model-based
467 reanalysis data that is used to verify the discharge forecast, and may also be impacted by
468 potential underestimates of the rainfall totals by GPM IMERG. The intensity error at landfall
469 is not significantly correlated with river discharge skill at any forecast lead time (Figure 9e).
470 Although there is a positive correlation overall between the intensity error and rainfall error,
471 there is actually a counter-intuitive negative correlation at three of the forecast runs, and
472 when you take an average of the correlations at each forecast run (Figure 9d). So, in a
473 particular forecast cycle, the members with better forecasts for TC intensity do not tend to be
474 those with better precipitation forecasts or river discharge forecasts. There are however
475 significant correlations as expected between the precipitation error and the river discharge
476 score (Figure 9f). It is therefore crucial to verify and understand the skill of ensemble TC
477 rainfall forecasts, and focus future research on understanding and improving rainfall
478 prediction in addition to track and intensity in order to advance flood forecasting in TC cases.

File generated with AMS Word template 2.0

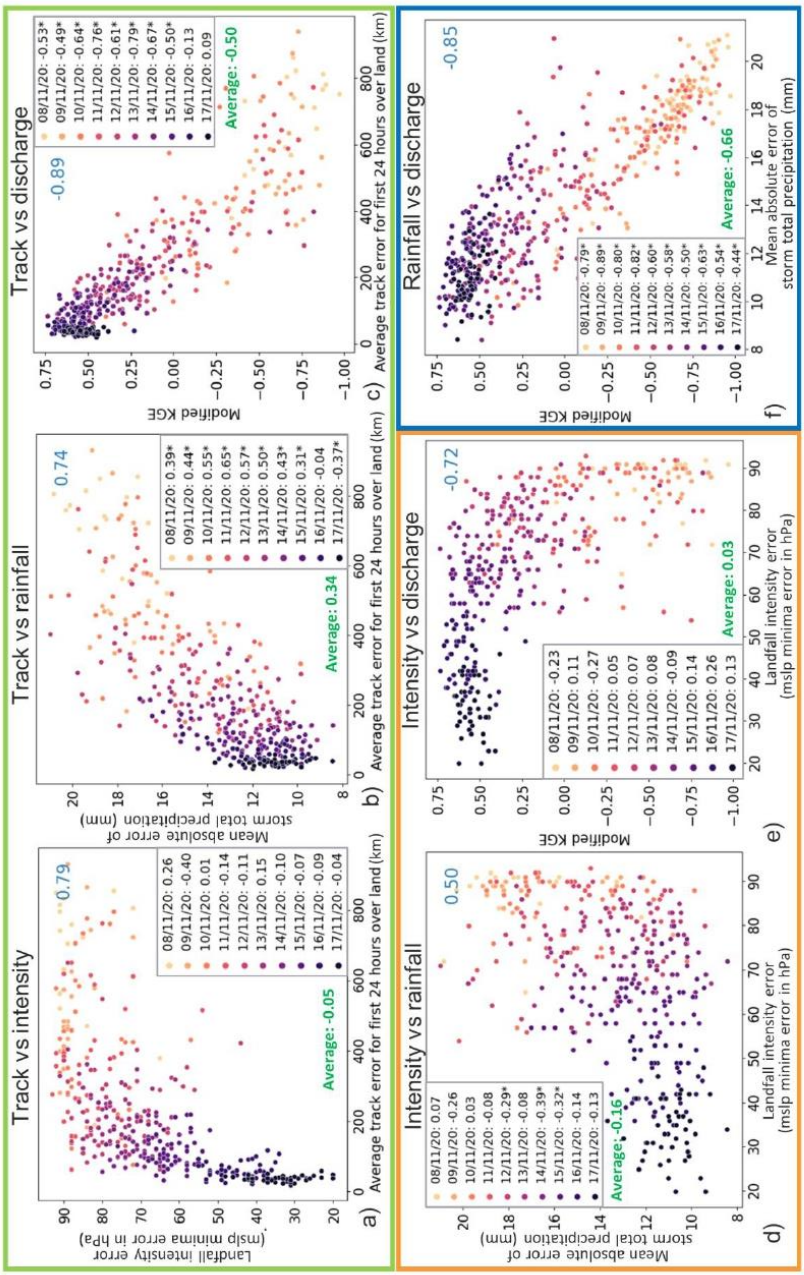
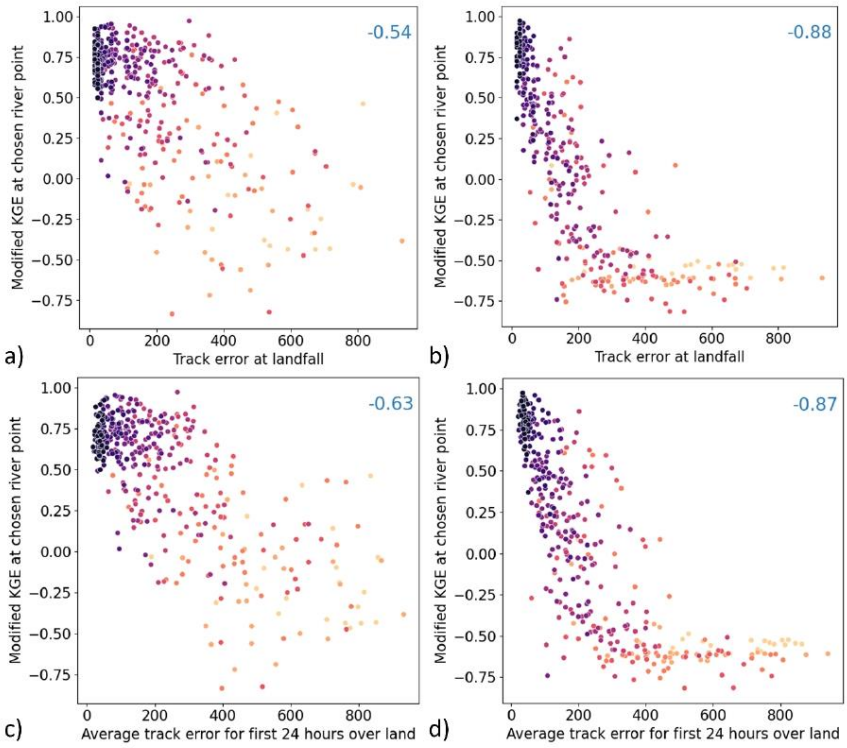


Figure 9: Scatter plots of each error type / score where each dot is one member: a) Average track error over land vs landfall intensity error; b) Average track error over land vs MAE of storm total precipitation; c) Average track error over land vs average KGE' score at all severely impacted river points (where 5-year return period is exceeded in the GloFAS reanalysis); d) landfall intensity error vs MAE of storm total precipitation; e) landfall intensity error vs average KGE'; f) MAE of storm total precipitation vs average KGE'. Overall Spearman correlations are shown on each plot in blue. Members are coloured by their forecast run and the Spearman correlation for each forecast run is shown in the legend, with a star indicating when this is significant (<0.05). The average of these forecast run correlations is shown in green. The green, orange and blue outlines show those relationships that were represented by this colour arrow in Figure 2.



e)

Forecast run (00 UTC)	a) Landfall track error vs KGE River Ulúa	b) Landfall track error vs KGE River Prinzapolka	c) Land track error vs KGE River Ulúa	d) Land track error vs KGE River Prinzapolka
08/11/20	-0.17	-0.29	-0.23	-0.36
09/11/20	-0.32	-0.25	-0.32	-0.27
10/11/20	0.11	-0.37	-0.23	-0.51
11/11/20	-0.51	-0.72	-0.65	-0.62
12/11/20	-0.03	-0.67	-0.06	-0.57
13/11/20	-0.22	-0.73	-0.33	-0.62
14/11/20	-0.08	-0.78	-0.33	-0.41
15/11/20	-0.00	-0.54	-0.03	-0.48
16/11/20	0.05	-0.27	0.17	-0.10
17/11/20	0.00	0.00	-0.03	-0.28
Average of run correlations	-0.12	-0.46	-0.20	-0.42

Figure 10: Scatter plots of the various errors / scores where each dot is one member: a) Track error at landfall vs KGE' at the River Ulúa outlet point; b) Track error at landfall vs KGE' at the River Prinzapolka outlet point; c) Average track error over land vs KGE' at the River Ulúa outlet point; d) Average track error over land vs KGE' at the River Prinzapolka outlet point. Overall Spearman correlations are shown on each plot in blue. The Spearman correlations for each run time (and the average of these) are shown in the table in (e). Members are coloured by their forecast run as shown in the table in (e).

494 Scatter plots of the two track metrics (landfall error and track error over land) against the
495 KGE' river discharge score for the River Prinzapolka reveal a strong 'l-shaped' relationship,
496 with a very quick drop off in forecast skill as track error increases, no forecast skill for higher
497 track errors, and little difference between the plots for the two track error metrics (with just a
498 slightly stronger correlation with landfall track error compared to the average track error over
499 land both overall and in the run-to-run correlations) (Figure 10, b, d, and e). This shows how
500 crucial an accurate landfall and track forecast was to a skilful flood forecast for this river. In
501 the equivalent plots for the River Ulúa (Figure 10, a, c and e), this l-shaped relationship is not
502 observed and there is a more gradual drop in KGE' as track error increases, and so an
503 accurate landfall forecast was less crucial for this river. The correlation between track error
504 and river discharge skill is lower compared to the Prinzapolka, and there is a stronger
505 relationship with average land track error rather than landfall error (overall and in the run-to-
506 run correlations), showing the greater influence of the detail of the forecast track after
507 landfall on how much rain fell in the Ulúa catchment in northern Honduras.

508 **5. Discussion**

509 The ensemble-based methods developed in this study to analyse the cascade of
510 predictability and uncertainty through the various stages of the TC flood forecasting chain
511 provide useful information to both modellers interested in optimising flood forecast skill and
512 to those who prepare and communicate flood forecasts with stakeholders and end-users in TC
513 events.

514 The results showed that shifts in track skill do correspond to shifts in precipitation and
515 discharge skill. The stronger relationship between discharge skill and the mean track error
516 over land compared to the landfall location error (as indicated by the divergence score
517 analysis and by the correlation scores), shows the importance of verifying and improving
518 forecasts of TC behaviour after landfall, and crucially of not only focusing on the landfall
519 location when communicating the forecasts in advance of a TC. The track after landfall is
520 crucial, as the rainfall totals and flood extent and severity will vary depending on the path the
521 cyclone takes after landfall and the translation speed at which it travels.

522 The large changes in TC intensity skill, which generally do not happen until shorter lead
523 times, do not result in significant improvements in precipitation and discharge skill, apart
524 from in catchments close to the landfall location that are impacted by the heavy rains around

525 the eye wall. More widely it is the wider rainfall from the storm that is crucial to flood
526 impacts, which is controlled more by the track (location and translation speed) and other
527 factors such as the size of the storm and the orography than it is by the intensity forecasts
528 (Titley et al., 2021). This also has important implications for forecast communication, as it
529 reinforces the need to focus on more than just the intensity category of the storm when
530 forecasting flood impacts.

531 The average correlation between rainfall and river discharge skill is stronger than just
532 track error against river discharge at most forecast runs and the analysis of a few strong
533 performing members from one of the forecast runs reveals that even similar track forecasts
534 can have very different precipitation patterns. This, combined with the finding that the spread
535 in precipitation forecast errors across the ensemble remains relatively large even after the
536 track spread reduces, highlights the importance of understanding and optimising precipitation
537 forecasts in the ensemble models that are used to drive ensemble flood forecasts.
538 Traditionally ensemble TC forecast verification at numerical weather prediction (NWP)
539 centres has focused on track and intensity (Magnusson et al., 2019; 2021; Titley et al., 2020;
540 Zhou et al, 2022), and the relatively crude precipitation verification done here does not
541 provide all the answers as to which aspects of the precipitation forecasts would benefit from
542 improvement. More detailed verification of ensemble precipitation forecasts in TC cases is
543 recommended, particularly using spatial or object-oriented methods to highlight which
544 specific aspects of rainfall forecasts would benefit from improvement in order to feed into
545 improvements in flood forecasts.

546 This study shows how closely the flood forecast skill is tied to both TC track and
547 precipitation forecast skill and how uncertainty in these forecast aspects feeds through to
548 uncertainty in the flood forecasts. However traditionally forecasts of precipitation and
549 downstream flood hazards in TCs have used deterministic forecast models, leading to
550 overconfident and unreliable flood predictions. The use of an ensemble flood forecasting
551 system, such as the GloFAS system used in this paper, where ensemble forecasts are carried
552 through the forecast chain to create probabilistic flood forecasts, are vital in representing the
553 uncertainty in flood predictions, and preparing for a range of possible flood scenarios when
554 there is an approaching TC. Some of the learning of this study can be directly applied in a
555 flood forecasting capacity by providing additional context and advice to decision makers. For
556 example, Figure 11 shows an example forecast TC centre plot that could be produced in the

File generated with AMS Word template 2.0

run up to an event to show the specific forecast tracks that were likely to lead to significant flooding in a particular river, by colouring the forecast centres by the discharge return period that is forecast to be exceeded. Being able to follow specific members through the forecast chain in this way could help in scenario planning and forecast communication as a supplementary tool alongside the probabilistic ensemble forecasts.

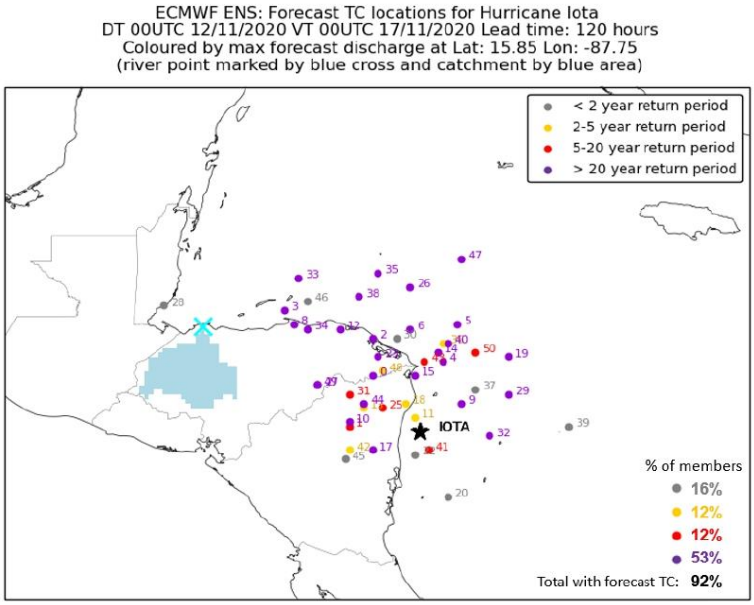


Figure 11: Prototype plot that could be produced in the run up to a TC event to highlight the forecast TC locations of those ensemble members that are forecasting a significant flood event at a particular river point. Here the outlet point on the River Ulúa has been used (cyan cross). The Day 5 ECMWF ENS forecast positions for Hurricane Iota from 00UTC 12th November 2020 forecast run are displayed (valid at 00UTC 17th November 2020 close to the time of landfall). The forecast positions are numbered by the ensemble member number (to allow the individual ensemble members to be followed through the forecast chain), and coloured specifically by the return period that is forecast to be exceeded on the River Ulúa. The thresholds and colours are the same as those used on the GloFAS web interface https://www.globalfloods.eu/.

The main strength of the techniques described in this study is that they allow for an analysis of the links between forecast track, intensity, precipitation and discharge skill across hundreds of sample forecasts for a particular high-impact case. We are using the ensemble as a larger sample of potential outcomes for what are rare events in any given location, and using this sample to investigate the interconnectivity of the various forecast components and give valuable insights as to the controls on TC flood predictability. All the data used in this

579 study are freely available, meaning the analysis is repeatable and scalable, and can be applied
580 to other cases in any TC basin around the world.

581 There are limitations to the methods as described here that could be improved in
582 subsequent case studies. The use of reanalyses as the river discharge verification data set has
583 the advantage of being available at every global river point, and further work is ongoing to
584 address their known limitations (Harrigan et al., 2020b; Zsoter et al., 2020), but if reliable
585 discharge observations were more widespread and openly available this would benefit studies
586 of this kind. The rainfall verification metric used here is a very simple point-based metric and
587 other metrics that remove the point-by-point sensitivity such as the Fractions Skill Score
588 (Roberts and Lean, 2008) could be used. It would also be useful to calculate the forecast
589 catchment-total rainfall in addition to the overall storm-total rainfall when analysing the
590 forecasts for specific rivers. A further limitation of this study is the reliance on GPM IMERG
591 data to verify the TC rainfall. GPM is an incredibly useful resource due to its global coverage
592 and wide accessibility, and the Final Run data used here incorporates additional microwave
593 passes and a monthly gauge correction. However, some of the ground-based observations in
594 Iota indicate potential errors in capturing the true rain rates in some regions, which may
595 impact the robustness of the analysis of the errors of the precipitation forecasts. More detailed
596 investigation of biases in observations of TC precipitation data is recommended, comparing
597 IMERG data with other sources of precipitation 'truth' data such as radar and gauge-based
598 datasets, in order to develop better understanding of observation uncertainty and improve the
599 confidence in TC rainfall verification results.

600 This study has focused on a particular case, Hurricane Iota, and the resultant flooding that
601 occurred in Honduras and Nicaragua. In the event of a cyclone with a similar forecast track
602 impacting the region, the findings of this study could help in giving greater confidence to
603 longer-range flood forecasts for those catchments that are forecasted to be impacted by the
604 wider storm circulation and orographically-enhanced rainfall, whilst understanding that the
605 locations and severity of flooding closer to the potential landfall site will be more difficult to
606 predict until closer to the landfall time. However, it is important to understand that even in
607 the same region, no two cyclones are the same in terms of their track, speed, size and wider
608 environmental conditions, and so care should be taken in applying the understanding gained
609 in this study to future storms. The wider applicability of these results in other geographical
610 contexts is also uncertain because flood severity also depends on the local landscape and

611 catchment characteristics. The purpose of this study was to provide a suggested approach and
612 set of methods to carry out an analysis of the flow of predictability through the flood forecast
613 chain in TC cases. The wider application of these methods across a range of cases would help
614 to inform a) the robustness and wider applicability of the findings of this case; b) how the
615 flow of predictability through the flood forecast chain varies with different geographical and
616 forecast contexts.

617 **6. Conclusion**

618 The work adds a new hydrological perspective to the traditional methods of verifying
619 global ensemble TC forecast performance. It proposes the use of ensemble-based methods to
620 analyse the predictability links between the track, intensity, precipitation and discharge
621 forecast components of the ensemble flood forecast chain in TC cases. These methods are
622 illustrated using Hurricane Iota, which led to widespread and severe impacts from fluvial
623 flooding in Central America in November 2020. We examined the performance of the
624 ensemble members at ten forecast runs in the lead up to the flood event, and found strong
625 correlations between the members' TC track error, precipitation error and river discharge
626 KGE' scores. An analysis of the run-to-run changes in skill for each forecast component
627 found that the largest run-to-run shift in forecast track errors corresponded to the largest
628 change in the discharge KGE' scores.

629 Once the forecasts began to show increased confidence for the TC genesis and track of
630 Iota (around 5 days ahead of landfall) there is a clear signal for a significant flood event in the
631 overall region that was affected. But at the catchment scale, the study highlights how the
632 sensitivity of discharge forecasts to the landfall track and intensity forecasts varies
633 considerably depending on their catchment's location relative to the TC, with a greater
634 sensitivity to precise landfall and intensity forecasts for river catchments close to the landfall
635 location. The stronger relationship found between the flood forecast skill across the impacted
636 region and the mean track error over land, when compared to only the landfall location error,
637 shows the importance of predicting TC behaviour after landfall and not only focusing on
638 landfall location when communicating TC forecasts. Changes in TC intensity only result in
639 significant improvements in discharge skill for river catchments impacted by the heavy rains
640 around the eye wall. The rainfall from the storm's wider circulation is crucial to flood
641 impacts in most of the affected river catchments in the region, and is less sensitive to the TC

File generated with AMS Word template 2.0

intensity, showing the importance of focusing on more than just the category of the storm when forecasting flood impacts.

The spread in precipitation errors and discharge scores is shown to reduce at a lower rate than that of the track forecasts, with relatively large spread found even after the forecast track confidence began to increase. A more detailed look at the precipitation forecasts from some individual ensemble members reveals that the forecast precipitation patterns and total accumulations can vary considerably even when the forecast landfall locations are similar, emphasizing how vital it is to increase the focus on evaluating and understanding ensemble TC rainfall forecasts in order to improve TC flood forecasting.

The application of these methods to additional TC cases is recommended and will allow a broader investigation of the lead times at which ensemble flood forecasting systems such as GloFAS are able to provide useful early warning of river flooding in TC cases, and what forecast and geographical contexts lead to variations in the predictability and the strength of the links between each component of the forecast chain.

Acknowledgments.

Author Helen Titley acknowledges the support of the Weather and Climate Science for Service Partnership (WCSSP) India and South East Asia projects, supported by the U.K. government's Newton Fund. Liz Stephens and Hannah Cloke acknowledge funding from the U.K.'s Natural Environment Research Council (NERC) and Foreign Commonwealth & Development Office (FCDO) (formerly DFID) Science for Humanitarian Emergencies and Resilience (SHEAR) research program Forecasts for Anticipatory Humanitarian Action (FATHUM) project Grant NE/P000525/1. Author Hannah Cloke also acknowledges funding from the U.K.'s Natural Environment Research Council (NERC) The Evolution of Global Flood Risk (EVOFLOOD) project Grant NE/S015590/1.

Data Availability Statement.

The data in this study was accessed from the following available sources: i) ECMWF ENS tracks from the TIGGE CXML archive at the NCAR Research Data Archive (NCEP et al., 2008, doi.org/10.5065/D6GH9GSZ); ii) ECMWF ENS precipitation forecast data from ECMWF Meteorological Archival and Retrieval System (MARS) (ECMWF, 2022); iii) GloFAS river discharge forecasts from the Copernicus Climate Change Service (C3S)

673 Climate Data Store (CDS) (Zsoter et al., 2019, doi.org/10.24381/cds.ff1aef77); iv) IBTrACS
674 data from the National Centers for Environmental Information (NCEI) (Knapp et al., 2018,
675 doi.org/10.25921/82ty-9e16); v) GPM IMERG (final version) precipitation data from NASA
676 (Huffman et al., 2019); vi) GloFAS-ERA5 operational global river discharge reanalysis from
677 the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) (Harrigan et al.,
678 2019, doi.org/10.24381/cds.a4fdd6b9).

679

680 REFERENCES

681 Alfieri, L., P. Burek, E. Dutra, B. Krzeminski,, D. Muraro, J. Thielen, and F. Pappenberger,
682 2013: GloFAS – global ensemble streamflow forecasting and flood early warning,
683 *Hydrol. Earth Syst. Sci.*, 17, 1161–1175, doi.org/10.5194/hess-17-1161-2013

684 Alfieri, L., E. Zsoter, S. Harrigan, F. Aga Hirpa, C. Lavaysse, C. Prudhomme, and P.
685 Salamon, 2019: Range-dependent thresholds for global flood early warning. *Journal of*
686 *Hydrology X*, 4, 100034, doi.org/10.1016/j.hydroa.2019.100034

687 Czajkowski, J. and E. Kennedy, 2010: Fatal tradeoff? Toward a better understanding of the
688 costs of not evacuating from a hurricane in landfall counties. *Popul Environ*, 31(1–
689 3):121–149.

690 Czajkowski, J., G. Villarini, E. Michel-Kerjan, and J. A. Smith, 2013: Determining tropical
691 cyclone inland flooding loss on a large scale through a new flood peak ratio-based
692 methodology. Environ. Res. Lett., 8, 044056, doi.org/10.1088/1748-9326/8/4/044056

693 ECMWF, 2020: IFS Documentation CY47R1 - Part V: Ensemble Prediction System, .
694 ECMWF, Reading. doi.org/10.21957/d7e3lrb

695 ECMWF, 2022: Access to archive datasets, https://www.ecmwf.int/en/forecasts/access-
696 forecasts/access-archive-datasets, accessed 24 May 2022.

697 Emerton, R., H. Cloke, A. Ficchi, L. Hawker, S. de Wit, L. Speight, C. Prudhomme, P.
698 Rundell, R. West, J. Neal, J. Cuna, S. Harrigan, H. Titley, L. Magnusson, F.
699 Pappenberger, N. Klingaman, and E. Stephens, 2020: Emergency flood bulletins for
700 Cyclones Idai and Kenneth: A critical evaluation of the use of global flood forecasts for
701 international humanitarian preparedness and response. *International Journal of Disaster*
702 *Risk Reduction*, 50, doi.org/10.1016/j.ijdrr.2020.101811

703 Gneiting, T. and A. E. Raftery, 2007: Strictly proper scoring rules, prediction, and estimation.
704 *Journal of the American Statistical Association*, 102(477), 359–378,
705 doi.org/10.1198/016214506000001437

706 Golding, B. (ed), 2022: *Towards the "Perfect" Weather Warning: Bridging Disciplinary*
707 *Gaps through Partnership and Communication,* Springer, Switzerland,
708 doi.org/10.1007/978-3-030-98989-7

709 Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez, 2009: Decomposition of the Mean
710 Squared Error and NSE Performance Criteria: Implications for Improving Hydrological
711 Modelling. *Journal of Hydrology*, 377, 80-91, doi.org/10.1016/j.jhydrol.2009.08.003

712 Hakim, G. J. and R. D. Torn, 2008: 'Ensemble Synoptic Analysis'. In L. F. Bosart and H. B.
713 Bluestein (Eds.), Synoptic—Dynamic Meteorology and Weather Analysis and
714 Forecasting: A Tribute to Fred Sanders. Meteorological Monographs. Boston, MA:
715 American Meteorological Society, 147–161. doi.org/10.1007/978-0-933876-68-2_7.

716 Harrigan, S., E. Zsoter, C. Barnard, F. Wetterhall, P. Salamon, and C. Prudhomme, 2019:
717 River discharge and related historical data from the Global Flood Awareness System,
718 v2.1, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), accessed 3
719 February 2022, doi.org/10.24381/cds.a4fdd6b9

720 Harrigan, S., E. Zsoter, H. Cloke, P. Salamon, and C. Prudhomme, 2020a: Daily ensemble
721 river discharge reforecasts and real-time forecasts from the operational Global Flood
722 Awareness System, *Hydrol. Earth Syst. Sci. Discuss.*, doi.org/10.5194/hess-2020-532

723 Harrigan, S., E. Zsoter, L. Alfieri, C. Prudhomme, P. Salamon, F. Wetterhall, C. Barnard, H.
724 Cloke, and F. Pappenberger, 2020b: GloFAS-ERA5 operational global river discharge
725 reanalysis 1979–present. *Earth Syst. Sci. Data*, 12, 2043–2060, doi.org/10.5194/essd-12-
726 2043-2020

727 Huffman, G. J., D.T. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, C. Kidd, E. J. Nelkin, S.
728 Sorooshian, J. Tan, and P. Xie, 2019: Algorithm Theoretical Basis Document (ATBD)
729 Version 06. NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE
730 Retrievals for GPM (IMERG). NASA, https://pmm.nasa.gov/data-access/downloads/gpm

731 IFRC, 2020: Revised emergency appeal: Central America – hurricane Eta and Iota, Red
732 Cross & Red Crescent International. https://reliefweb.int/report/honduras/central-america-

733 hurricane-eta-iota-revised-emergency-appeal-n-mdr43007-revision-no-1, accessed 26
734 August 2022.

735 Jiang, H., C. Liu, and E. J. Zipser, 2011: A TRMM-based tropical cyclone cloud and
736 precipitation feature database. *J. Appl. Meteor. Climatol.*, 50, 1255–1274,
737 doi.org/10.1175/2011JAMC2662.1

738 Luitel, B., G. Villarini, and G. A. Vecchi, 2018: Verification of the skill of numerical weather
739 prediction models in forecasting rainfall from U.S. landfalling tropical cyclones, *J.*
740 *Hydrol.*, 556, 1026–1037, doi.org/10.1016/j.jhydrol.2016.09.019

741 Kling, H., M. Fuchs, and M. Paulin, 2012: Runoff conditions in the upper Danube basin
742 under an ensemble of climate change scenarios, *Journal of Hydrology*, 424-425, 264-277.

743 Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The
744 International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical
745 cyclone best track data, *Bulletin of the American Meteorological Society*, 91, 363-376,
746 doi.org/10.1175/2009BAMS2755.1

747 Knapp, K. R., H. J. Diamond, J. P. Kossin, M. C. Kruk, and C. J. Schreck, 2018: International
748 Best Track Archive for Climate Stewardship (IBTrACS) Project, Version 4. NOAA
749 National Centers for Environmental Information, doi.org/10.25921/82ty-9e16, accessed 9
750 August 2022.

751 Knoben, W., J. Freer and R. Woods, 2019: Technical note: Inherent benchmark or not?
752 Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, *Hydrology and Earth*
753 *System Sciences,* 23, 4323-4331, doi.org/10.5194/hess-23-4323-2019

754 Knutson, T., S. J. Camargo, J. C. L. Chan, K. Emanuel, C. Ho, J. Kossin, M. Mohapatra, M.
755 Satoh, M. Sugi, K. Walsh, and L. Wu, 2020: Tropical Cyclones and Climate Change
756 Assessment: Part II: Projected Response to Anthropogenic Warming, *Bulletin of the*
757 *American Meteorological Society*, 101(3), E303-E322, doi.org/10.1175/BAMS-D-18-
758 0194.1

759 Knutson, T. R., and Coauthors, 2010: Tropical cyclones and climate change, *Nat. Geosci.*, 3,
760 157–163, doi.org/10.1038/ngeo779

761 Knutson, T. R., and Coauthors, 2013: Dynamical downscaling projections of twenty-first-
762 century Atlantic hurricane activity: CMIP3 and CMIP5 model-based scenarios, *J.*
763 *Climate*, 26, 6591–6617, doi.org/10.1175/JCLI-D-12-00539.1

764 Knutson, T. R., J. J. Sirutis, M. Zhao, R. E. Tuleya, M. Bender, G. A. Vecchi, G. Villarini,
765 and D. Chavas, 2015: Global projections of intense tropical cyclone activity for the late
766 twenty-first century from dynamical downscaling of CMIP5/RCP4.5 scenarios, *J.*
767 *Climate*, 28, 7203–7224, doi.org/10.1175/JCLI-D-15-0129.1

768 Lavers, D., S. Harrigan, E. Andersson, D. S. Richardson, C. Prudhomme, and F.
769 Pappenberger, 2019: A vision for improving global flood forecasting, Environ. Res. Lett.,
770 14(12), doi.org/10.1088/1748-9326/ab52b2

771 Lynch, S. L. and R. S. Schumacher, 2014: Ensemble-Based Analysis of the May 2010
772 Extreme Rainfall in Tennessee and Kentucky. Monthly Weather Review, 142 (1), 222–
773 239, doi.org/10.1175/MWR-D-13-00020.1

774 Magnusson, L., and Coauthors, 2019: ECMWF Activities for Improved Hurricane Forecasts.
775 Bull. Amer. Meteor. Soc., 100, 445–458, doi.org/10.1175/BAMS-D-18-0044.1

776 Magnusson, L., and Coauthors, 2021: Tropical cyclone activities at ECMWF. ECMWF
777 Technical Memoranda no. 888. ECMWF, Reading, https://doi.org/10.21957/zzxzzygwv

778 NCEP and Coauthors, 2008: THORPEX Interactive Grand Global Ensemble (TIGGE) Model
779 Tropical Cyclone Track Data. Research Data Archive at the National Center for
780 Atmospheric Research, Computational and Information Systems Laboratory, Boulder,
781 CO., accessed 26 May 2022, doi.org/10.5065/D6GH9GSZ

782 Owens, R. G., and T. D. Hewson, 2018: ECMWF Forecast User Guide. ECMWF, Reading.
783 doi.org/10.21957/m1cs7h

784 Prat, O. P., and B. R. Nelson, 2013: Mapping the world's tropical cyclone rainfall
785 contribution over land using the TRMM Multi-satellite Precipitation Analysis, *Water*
786 *Resour. Res.*, 49, 7236–7254, doi.org/10.1002/wrcr.20527

787 Prat, O. P., and B. R. Nelson, 2016: On the link between tropical cyclones and daily rainfall
788 extremes derived from global satellite observations, *J. Climate*, 29, 6127–6135,
789 doi.org/10.1175/JCLI-D-16-0289.1

790 Rappaport, E. N., 2014: Fatalities in the United States from Atlantic Tropical Cyclones: New
791 Data and Interpretation, *Bull. Amer. Meteor. Soc.*, 95, 341–346

792 Richardson, D. S., H. L. Cloke, and F. Pappenberger, 2020: Evaluation of the consistency of
793 ECMWF ensemble forecasts, *Geophysical Research Letters*, 46,
794 doi.org/10.1029/2020GL087934

795 Roberts, N., and H. Lean, 2008: Scale-selective verification of rainfall accumulations from
796 high-resolution forecasts of convective events, *Mon. Wea. Rev.*, 136 , 78–97,
797 doi.org/10.1175/2007MWR2123.1

798 Senkbeil, J., J. Collins, and J. Reed, 2019: Evacuee Perception of Geophysical Hazards for
799 Hurricane Irma, *Weather, Climate, and Society*, 11(1), 217-227, doi.org/10.1175/WCAS-
800 D-18-0019.1

801 Shen, Y., Y. Du, and G. Chen, 2020: Ensemble Sensitivity Analysis of Heavy Rainfall
802 Associated With Three MCSs Coexisting Over Southern China. Journal of Geophysical
803 Research: Atmospheres, 125 (2), doi.org/10.1029/2019JD031266.

804 Speight, L., E. Stephens, L. Hawker, C. Baugh, J. Neal, H. L. Cloke, S. Grey, H. Titley, K.
805 Marsden, T. Sumner, A. Ficchi, C. Prudhomme, L. Archer, J. Bazo, J. Dambo, S. Dolan,
806 A. L. Hunn, F. Moschini, J. Savage, A. Smith, J. Towner, and M. Wanzala, 2022:
807 Recommendations to improve the interpretation of global flood forecasts to support
808 international humanitarian operations for tropical cyclones, *Journal of Flood Risk*
809 *Management*, Submitted

810 Stein, R. M., L. Dueñas-Osorio, and D. Subramanian, 2010: Who evacuates when hurricanes
811 approach? The role of risk, information, and location, *Social Science Quarterly*, 91(3),
812 816-834

813 Stewart, S. R., 2021: National Hurricane Center tropical cyclone report: Hurricane Iota
814 (AL312020), NOAA, https://www.nhc.noaa.gov/data/tcr/AL312020_Iota.pdf, accessed
815 26 Aug 2022

816 Titley, H. A., R. L. Bowyer, and H. L. Cloke, 2020: A global evaluation of multi-model
817 ensemble tropical cyclone track probability forecasts. Q J R Meteorol Soc. 2020; 146:
818 531– 545. https://doi.org/10.1002/qj.3712

819 Titley, H. A., H. L. Cloke, S. Harrigan, F. Pappenberger, C. Prudhomme, J. C. Robbins, E.
820 M. Stephens, and E. Zsoter, 2021: Key factors influencing the severity of fluvial flood
821 hazard from tropical cyclones, *Journal of Hydrometeorology*, 22, 1801-1817,
822 doi.org/10.1175/JHM-D-20-0250.1

823 Titley, H. A., M. Yamaguchi, and L. Magnusson, 2019: Current and potential use of
824 ensemble forecasts in operational TC forecasting: results from a global forecaster survey,
825 *Tropical Cyclone Research and Review*, 8(3), 166-180,
826 doi.org/10.1016/j.tcrr.2019.10.005

827 Towner, J., H. L. Cloke, E. Zsoter, Z. Flamig, J. M. Hoch, J. Bazo, E. Coughlan de Perez, and
828 E. M. Stephens, 2019: Assessing the performance of global hydrological models for
829 capturing peak river flows in the Amazon basin, *Hydrology and Earth System Sciences*,
830 23 (7). pp. 3057-3080, doi.org/10.5194/hess-23-3057-2019

831 UN OCHA, 2020a: Latin America and The Caribbean: 2020 hurricane seasons - situation
832 report no. 4,
833 https://reliefweb.int/sites/reliefweb.int/files/resources/20201120_CA%20Eta%20SitRep%
834 204%20ENG.pdf, accessed 26 Aug 2022

835 UN OCHA, 2020b: Central America and Mexico: 2020 Hurricane season - situation report
836 no. 5,
837 https://reliefweb.int/sites/reliefweb.int/files/resources/20201126%20CA%20Eta%20SitRe
838 p%205%20%28ENG%29.pdf, accessed 26 August 2022

839 Wang, C. C., B. X. Lin, C. T. Chen, and S. H. Lo, 2015: Quantifying the effects of long-term
840 climate change on tropical cyclone rainfall using a cloud-resolving model: Examples of
841 two landfall typhoons in Taiwan, *J. Climate*, 28, 66–85, doi.org/10.1175/JCLI-D-14-
842 00044.1

843 Whitehead, J. C., R. Edwards, M. Van Willigen, J. R. Maiolo, K. Wilson, and K. T. Smith,
844 2000: Heading for higher ground: factors affecting real and hypothetical hurricane
845 evacuation behavior, *Global Environmental Change Part B: Environmental Hazards*, 2:4,
846 133-142

847 WMO, 2014: HIWeather implementation plan,
848 http://www.hiweather.net/Uploads/ue/file/20190411/1554934497413129.pdf, accessed 26
849 Aug 2022

Wright, D., T. R. Knutson, and J. A. Smith, 2015: Regional climate model projections of rainfall from U.S. landfalling tropical cyclones, *Climate Dyn.*, 45, 3365–3379, doi.org/10.1007/s00382-015-2544-y

Zhang, M. and Z. Meng, 2018: Impact of Synoptic-Scale Factors on Rainfall Forecast in Different Stages of a Persistent Heavy Rainfall Event in South China. Journal of Geophysical Research: Atmospheres, 123 (7), 3574–3593, doi.org/10.1002/2017JD028155

Zhou, X., and Coauthors, 2022: The Development of the NCEP Global Ensemble Forecast System Version 12. Wea. Forecasting, 37, 1069–1084, https://doi.org/10.1175/WAF-D-21-0112.1.

Zsoter, E., S. Harrigan, C. Barnard, F. Wetterhall, P. Salamon, C. Prudhomme, 2019: River discharge and related forecasted data from the Global Flood Awareness System, v2.1, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), accessed 3 February 2022, doi.org/10.24381/cds.ff1aef77

Zsótér, E., H. L. Cloke, C. Prudhomme, S. Harrigan, P. de Rosnay, J. Munoz-Sabater, and E. Stephens, 2020: Trends in the GloFAS-ERA5 river discharge reanalysis. ECMWF Technical Memoranda, 871

A5: Published Article: Track forecast: Operational capability and new techniques - Summary from the Tenth International Workshop on Tropical Cyclones (IWTC-10)

This appendix contains the published version of the following paper (summarised in Section 7.5.1):

Conroy, A., Titley, H., Rivett, R., Feng, X., Methven, J., Hodges, K., Brammer, A., Burton, A., Chakraborty, P., Chen, G., Cowan, L., Dunion, J. and Sarkar, A. (2023) Track forecast: Operational capability and new techniques - Summary from the Tenth International Workshop on Tropical Cyclones (IWTC-10). *Tropical Cyclone Research and Review*, 12 (1): 64–80. doi:10.1016/j.tcrr.2023.05.002.

ESCAP/WMO Typhoon Committee

Available online at www.sciencedirect.com

ScienceDirect

Tropical Cyclone Research and Review 12 (2023) 64–80

www.keaipublishing.com/tcrr

Track forecast: Operational capability and new techniques - Summary from the Tenth International Workshop on Tropical Cyclones (IWTC-10)

Adam Conroy [a,*], Helen Titley [b], Rabi Rivett [a], Xiangbo Feng [c], John Methven [d], Kevin Hodges [d], Alan Brammer [e], Andrew Burton [a], Paromita Chakraborty [f], Guomin Chen [g], Levi Cowan [h], Jason Dunion [i], Abhijit Sarkar [f]

[a] *Bureau of Meteorology, West Perth, Australia*
[b] *Met Office, Exeter, UK*
[c] *National Centre for Atmospheric Science and Department of Meteorology, University of Reading, Reading, UK*
[d] *Department of Meteorology, University of Reading, Reading, UK*
[e] *Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, USA*
[f] *National Centre for Medium Range Weather Forecasting, Noida, India*
[g] *Shanghai Typhoon Institute of China Meteorological Administration, Shanghai, China*
[h] *Joint Typhoon Warning Center, Honolulu, USA*
[i] *Cooperative Institute for Marine and Atmospheric Studies, University of Miami, and Hurricane Research Division, NOAA/Atlantic Oceanographic and Meteorological Laboratory, Miami, USA*

Available online 9 May 2023

Abstract

In this paper, we summarize findings from the Tenth International Workshop on Tropical Cyclones (IWTC-10) subgroup on operational track forecasting techniques and capability.

The rate of improvement in the accuracy of official forecast tracks (OFTs) appears to be slowing down, at least for shorter lead times, where we may be approaching theoretical limits. Operational agencies continue to use consensus methods to produce the OFT with most continuing to rely on an unweighted consensus of four to nine NWP models. There continues to be limited use of weighted consensus techniques, which is likely a result of the skills and additional maintenance needed to support this approach. Improvements in the accuracy of ensemble mean tracks is leading to increased use of ensemble means in consensus tracks.

Operational agencies are increasingly producing situation-dependent depictions of track uncertainty, rather than relying on a static depiction of track forecast certainty based on accuracy statistics from the preceding 5 years. This trend has been facilitated by the greater availability of ensemble NWP guidance, particularly vortex parameter files, and improved spread in ensembles. Despite improving spread-skill relationships, most ensemble NWP systems remain under spread. Hence many operational centers are looking to leverage "super-ensembles" (ensembles of ensembles) to ensure the full spread of location probability is captured. This is an important area of service development for multi-hazard impact-based warnings as it supports better decision making by emergency managers and the community in the face of uncertainty.

© 2023 The Shanghai Typhoon Institute of China Meteorological Administration. Publishing services by Elsevier B.V. on behalf of KeAi Communication Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Tropical cyclones; Track forecasting; Ensembles; Track uncertainty

* Corresponding author.
E-mail address: Adam.Conroy@bom.gov.au (A. Conroy).
Peer review under responsibility of Shanghai Typhoon Institute of China Meteorological Administration.

Production and Hosting by Elsevier on behalf of KeAi

https://doi.org/10.1016/j.tcrr.2023.05.002
2225-6032/© 2023 The Shanghai Typhoon Institute of China Meteorological Administration. Publishing services by Elsevier B.V. on behalf of KeAi Communication Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Forecasts of tropical cyclone impacts from storm surge, freshwater flooding, and extreme wind are subject to uncertainties that vary significantly from case to case and forecast run to forecast run, and to which stakeholders are sensitive in varied ways. Regional Specialized Meteorological Centres (RSMCs) value probabilistic guidance for TCs from ensemble forecast systems, but the "pull-through" of probabilistic information to operational warnings using those forecasts is slow (Titley et al., 2019). RSMCs are taking different approaches to producing probabilistic TC forecast products and using differing methods, and there is an opportunity for better coordination across RSMCs and forecast centers to develop best practices. The 2018 International Working Group on Tropical Cyclones (IWTC-9) recognized the need for improved probabilistic guidance for TCs globally and proposed several recommendations to streamline the use of ensemble forecast guidance and uncertainty information in operational forecast warnings and products (Titley et al. 2019), including the importance of a coordinated interdisciplinary approach in developing best practice.

In this paper we will give an overview over current capability in track forecasting at selected forecast centers. Some of the new techniques developed since IWTC-9 include greater use of Ensemble Prediction Systems (EPS) to express track uncertainty. Discussion will then move towards how we can better harness the available probabilistic guidance in tropical cyclone forecasting, which includes findings from Phase 1 of the Tropical Cyclone-Probabilistic Forecast Products (TC-PFP) project.

2. Methods of producing official forecast tracks

Operational warning agencies typically produce an Official Forecast Track (OFT) based on a consensus of model tracks, most commonly a combination of deterministic tracks and mean tracks from ensemble guidance. Improvements in skill from OFTs are decreasing, and in the future there is likely to be greater focus on probabilistic expression of TC impacts rather than the track itself. Nonetheless, an OFT is likely to remain a component of the tropical cyclone forecast service for some time to come. There may however be an argument to transition towards creating OFTs purely from ensemble data, particularly when this is the same data used to generate probabilistic forecasts of TC hazards.

2.1. Unweighted consensus

Most centers generate a forecast track based on an unweighted consensus of model guidance, usually a combination of deterministic models, with a growing number of centers also including some ensemble mean tracks as part of the consensus. Often there is a standard selection of consensus members selected, although sometimes the forecaster may exclude certain members, or occasionally use multiple runs of the same model (especially for those centers producing forecast tracks prior to TC formation, when not every model will have a track available).

Tropical cyclone forecasts at MetService (New Zealand) are prepared using an in-house tool called CyTrack, which allows the forecaster to create tropical cyclone tracks with a variety of model guidance and satellite imagery layered underneath. After the initial analysis, the forecast track is based on a consensus of the UK, EC and GFS deterministic forecast positions. The forecast positions from the previous forecast track are also included in the consensus, so that changes to forecast policy are made incrementally rather than a large jump when new model data is available. If the model analyses are different from the observed position, then an adjustment may also be made to the short-term track so there is no discontinuity along the track from analysis to forecast.

The Japan Meteorological Agency (JMA) mainly employs a consensus method for TC track forecasts. This approach involves taking the mean of predicted TC positions from multiple deterministic global models, including the Global Spectral Model (the JMA global deterministic model), European Centre for Medium-Range Weather Forecasts (ECMWF), National Centers for Environmental Prediction (NCEP) and United Kingdom Met Office (UKMO).

In a similar fashion, the China Meteorological Administration (CMA) typically produce a consensus track based on ECMWF, NCEP, and the CMA global and regional deterministic models. One of the component CMA models used in the consensus is a deep learning ensemble correction algorithm called TYTEC-AI (TYphoon Track Ensemble Correction-AI prediction technique). TYTEC-AI is a weighted consensus on multiple EPS. The weighting coefficient is applied to each component model (CMA, ECMWF, NCEP and UK ensembles) and is changing from year to year. TYTEC-AI has been found to outperform deterministic models when verified over the 2012–2019 period.

The RSMC La Reunion "official" track forecast is a consensus forecast, which relies on the outputs of all available deterministic numerical models and ensemble systems provided by various modeling centers (e.g. ECMWF, Meteo-France, the UK met Office, or NCEP), as well as knowledge on recent performances of each individual model.

The Joint Typhoon Warning Center (JTWC) has leveraged the consensus approach to tropical cyclone (TC) track forecasting since the late 1990s (Sampson and Schrader, 2000). An unweighted consensus tracker (CONW) consisting of nine members is currently used,[1] made up of global deterministic models and the means of global ensemble system forecasts. The JTWC consensus requires a minimum of two of the nine members be present in order to generate the CONW tracker. The Naval Research Laboratory (NRL) and JTWC annually review the performance and reliability of various models to assess the sensitivity of CONW accuracy to each member and to optimize overall accuracy of the consensus. For example, in 2019, the UK Met Office global ensemble mean (UEMN) was added to the track forecast consensus, and NCEP's HWRF model and the U.S.

[1] The nine members of CONW are GFS – NCEP, ECMWF, UKMET, GALWEM – U.S. Air Force, NAVGEM – U.S. Navy, JMA Global Spectral Model, ECMWF Ensemble Mean, GFS Ensemble Mean and UKMET Ensemble Mean.

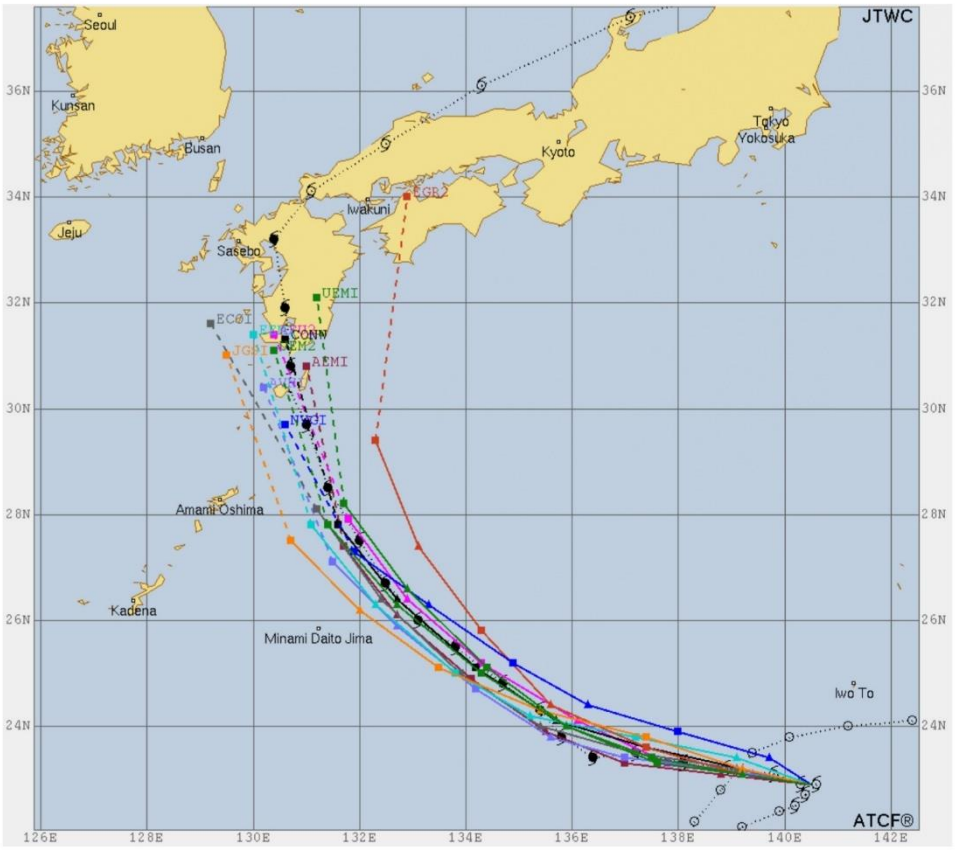


Fig. 1. Track forecasts from CONW members for Tropical Storm 16W initialized at 1200 UTC 14 September 2022 out to 72-h lead time, displayed in the Automated Tropical Cyclone Forecast (ATCF) system. The verifying track is plotted with black tropical cyclone symbols. (Image courtesy JTWC).

Navy's COAMPS-TC model were removed from the consensus. Interpolated versions of each model tracker are used when computing the consensus tracker, which shift the raw model track forecast in space and time to match the most recent best track position analyzed by JTWC. An example of the current CONW members as displayed in the Automated Tropical Cyclone Forecast (ATCF) system is shown in Fig. 1.

Forecasters at JTWC use the track forecast consensus (CONW) as a reasonable starting point for producing a TC forecast. This is done by displaying CONW and its members in a similar fashion to Fig. 1 within ATCF, allowing the forecast spread and the individual model predictions to be analyzed visually. After examining the guidance envelope, the forecaster then selects forecast positions from the map, taking into account known model biases and the previous forecast.

The Australian Bureau of Meteorology (ABoM) creates an unweighted consensus of a standard set of NWP models as a basis for the Official Forecast Track.[2] The consensus is created by a vector motion average of the component models at each timestep, which reduces discontinuities in the track as model availability changes. Once a consensus is created, the forecaster may make minor changes, particularly to bias the first 12 h of the OFT to recent motion.

Some customers have a requirement for an OFT prior to the formation of a tropical system. These are typically generated in a similar fashion to the normal OFT process, with the key differences being that the forecast starts at a future timestep

[2] The standard members of the ABoM consensus (subject to availability) are: ACCESS-G deterministic (Australian global model), ACCESS-G ensemble mean (2022 upgrade AG4 is the preferred model), ECMWF deterministic, EC ensemble mean, GFS deterministic, GFS ensemble mean, JMA deterministic, UK deterministic and UK ensemble mean.

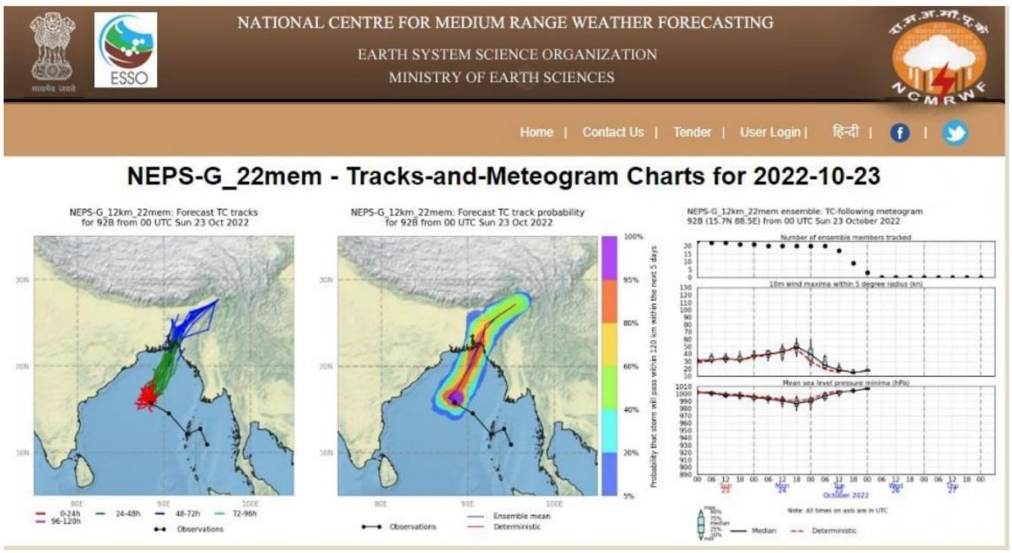


Fig. 2. Sample of a tropical cyclone forecast from NCMRWF (India).

(with no analysis position), and with less available guidance to generate the track from.

2.2. Weighted consensus

A weighted consensus track is one where a weighting coefficient is applied to each component model in the consensus, usually based on past performance. While this can increase the skill of the forecast track, it can require a large maintenance overhead, as whenever significant changes are made to the model, there needs to be a recalculation of the weighting coefficients. This usually needs to be done each season and can even be updated during the season. Because of this overhead, a weighted consensus is not feasible for some centers.

Deterministic track and intensity forecasts produced by the National Hurricane Center in the Atlantic and East Pacific basins come from an ensemble of inputs in a corrected consensus method. The Hurricane Forecast Improvement Program (HFIP) Corrected Consensus Approach (HCCA) technique produces skillful track and intensity forecasts utilizing up to 14 dynamical or statistical model forecasts to produce a consensus output (Simon et al. 2018). As part of this approach the models are weighted based on past performance and the error statistics are updated during the season, revealing expected performance of models as well as the uncertainty between skillful models.

2.3. Use of ensemble mean positions in track forecasting

While not common, there are forecast centers that present tropical cyclone forecasts which use purely ensemble data, including for the forecast track itself. The National Centre for Medium Range Weather Forecasting (NCMRWF – India) have moved towards displaying an ensemble average track for their tropical cyclone forecast tracks. The ensemble mean positions are calculated from the NEPS-G ensemble (the global ensemble prediction system of NCMRWF). This ensemble has a horizontal resolution of 12 km, and consists of 23 members (1 control, 11 perturbed on the most recent run, and 11 lagged members from the run 12 h previous). Systems are tracked using the bi-variate method adopted by the Met Office (UK). Fig. 2 shows a sample of a forecast issued on 23 October 2022. As well as the track, the forecast also shows a probability of location, and meteograms depicting the ensemble spread of 10 m wind and MSLP. The deterministic track shown on this graphic is NCUM, which is the global unified model of NCMRWF.

There remains debate as to whether a forecast track should even be produced as part of TC warning services, as it can draw focus to a single scenario, when the focus should be shifting to considering the range and likelihood of possible tracks. However, it is likely that forecast tracks will continue to be produced for some time yet, given how deeply embedded they are in TC forecast services. As more centers move towards using ensemble spread to generate a cone of uncertainty, or a depiction of forecast confidence, further consideration should be given to the suitability of using this ensemble guidance for generating the official forecast track. At ABoM, some comparisons have been made between super ensemble means and official forecast tracks (a "super ensemble" is defined as a combination of multiple ensemble runs, either of different

ensemble models, different ensemble run times, or a combination of both. Investigations are ongoing but early results are promising. Further work should be done to compare super ensemble means to official forecast tracks on a normalized, global dataset, and with the inclusion of more ensemble prediction systems. An ensemble mean having comparable skill to existing methods would also bring with it the benefits of being more stable, and being consistent with probabilistic outputs from the same ensemble dataset.

3. Skill trends in official forecast tracks

Landsea and Cangialosi (2018) posed the question, "have we reached the limits of predictability for tropical cyclone track forecasting?" Errors in track position had decreased by two thirds in the Eastern North Pacific and North Atlantic basins during the period from 1990 to 2016. This was primarily due to "global modeling advances, data assimilation improvements, dramatic increases in observations primarily derived from satellite platforms, and use of ensemble forecast techniques". It was pointed out that these improvements cannot continue indefinitely, and that there exists a "limit of predictability" in the atmosphere due to chaotic growth of errors with lead time.

Fig. 3 shows track errors in the North Atlantic and Eastern North Pacific relative to a climatology and persistence model (CLIPER5), which is used to normalize errors and account for track variability. This shows that overall there has been very little improvement in track forecasting skill from 2012 to 2016 in these basins. "The sample of 5 years is too short for meaningful statistical significance testing, but it does suggest there has been a slower rate of improvement or perhaps no additional advances" during this period.

Whilst it is difficult to determine when we will reach the limits of predictability (or indeed, whether we are already approaching it), knowledge of this will be critical to all users of tropical cyclone track forecast information. Regardless, as the rate of improvement in track forecasting slows, Landsea and Canglialosi suggest the emphasis should change to providing well-calibrated probabilistic forecasting information for tropical cyclone impacts. This includes finding new and improved techniques for characterizing track uncertainty (discussed in next section) and for generating probabilistic guidance for tropical cyclone impacts such as wind, storms surge, and rainfall. Ultimately, a full understanding of the spatial and temporal risk of these impacts will be of most importance to improving decision making.

A similar slowing improvement of track forecast performance since 2014 is evident in the western North Pacific. Fig. 4 shows trends in mean position error (PE) for five forecast centers in the western North Pacific (CMA, JMA, JTWC, KMA – Korea Meteorological Administration, and HKO – Hong Kong Observatory). Fig. 5 shows similar plots for global and regional NWP track forecasts. Both these figures show very little improvement in skill since 2014.

If the 96 and 120 h PEs are included (see Fig. 6), then it appears there are still some better skill improvements being made at longer lead times. Yu et al. (2021) suggest that this improving trend may continue for some time yet, particularly as improved analysis of position and structure of TCs result in better initialization in NWP guidance, as well as reduced initial position errors in forecast tracks feed into lower errors for the duration of the forecast track. However, it should be noted that verification scores routinely show ECMWF outperforming other global models for track despite not using synthetic observations, and hence having larger position errors at analysis time.

4. Characterizing track uncertainty

In the past, the uncertainty in tropical cyclone location has most commonly been based on fixed values such as the historic forecast error. However, a static area like this fails to take into account any situationally dependent information. With the increased availability and skill of ensemble data, it is possible to more accurately calculate the probability of tropical cyclone location through the forecast period (Titley et al., 2020). Being able to represent this probability accurately not only increases the value of the track forecast, but also can feed into other impact-based products that rely on the location of the tropical cyclone. This will ultimately enable users of the forecast products to make better decisions.

The area that depicts track uncertainty has typically been referred to as an uncertainty area or a cone of uncertainty. However, as there is a move towards using dynamical methods to determine track uncertainty, it makes sense for the name of the area representing this uncertainty to match the method. For a forecast timestep, there exists a probability distribution of where the system will be located at that time (a probability of location). The aim of using dynamic methods is to attempt to create an areal representation of this probability distribution at a pre-selected confidence level. For example, if you select a confidence level of 80%, then the system would be expected to be located within that area 80% of the time. A suitable name for this expression of uncertainty at the particular forecast time is the 'Forecast Confidence Area' (FCA). And if we sum together all the FCAs from the analysis time to that particular forecast time, the swath created would be the 'Forecast Confidence Cone' (FCC). Not only does this express the uncertainty with more positive language, the terminology is consistent with the method used to generate it.

4.1. Statistical methods

For centers that use a statistical method in generating a spatial depiction of the track uncertainty, the area is usually based on an average forecast error over a number of previous seasons of forecasts produced by the center. This has the advantage of being predictable and easy to calculate, but it misses out on capturing situationally dependent information. Many centers use this method, although some do allow for forecaster manipulation to take into account model guidance (discussed in 'Statistical-dynamical methods').

The Joint Typhoon Warning Center (JTWC) takes a slightly different approach to other centers. Fig. 7 shows an

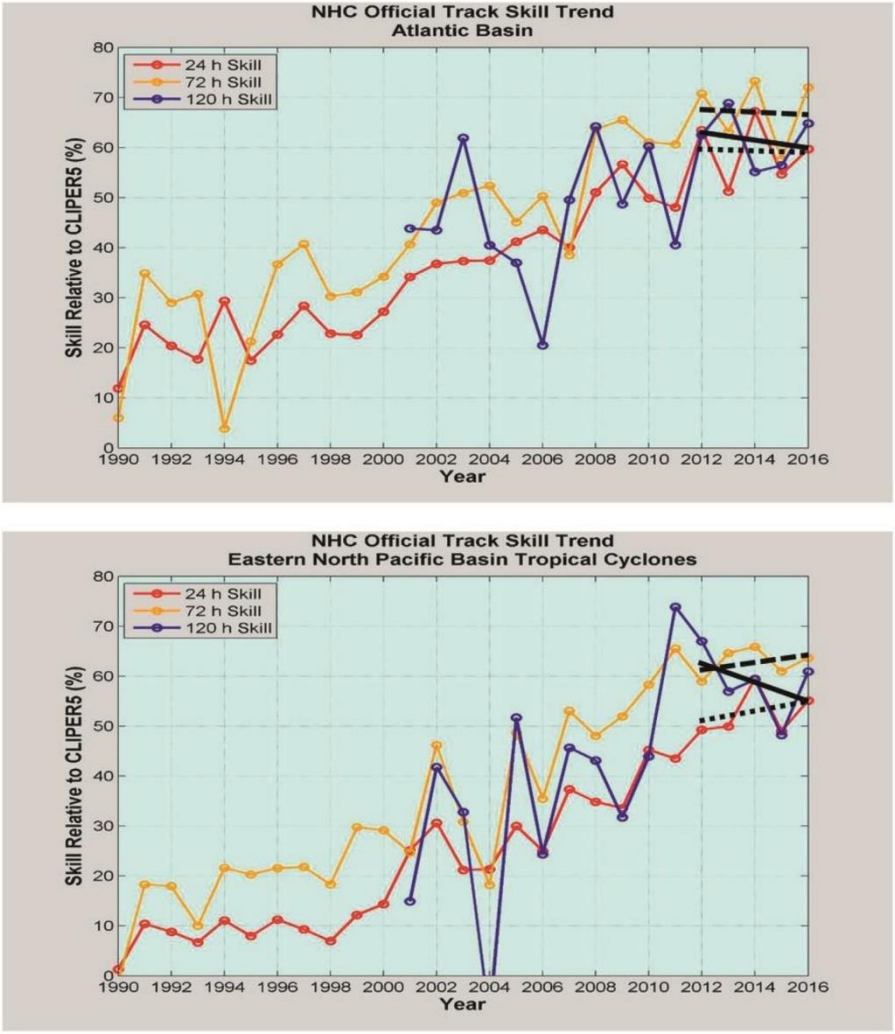


Fig. 3. 1990–2016 track forecast skill (improvements relative to CLIPER5) from the NHC for (top) the Atlantic basin and (bottom) the eastern North Pacific basin for the 24- (red), 72- (yellow), and 120-h (blue) predictions. The 2012–16 best fit linear trend is appended for each time series in black: 24 h, dotted; 72 h, dashed; and 120 h, solid. (Bulletin of the American Meteorological Society 99, 11; 10.1175/BAMS-D-17-0136.1).

example of a forecast graphic produced by the Joint Typhoon Warning Center – JTWC (USA). The TC forecast graphic includes a hatched region known as the "danger area" or "ship avoidance area." This is defined at each forecast point as a circle centered on that point with a radius equal to the JTWC 5-year mean track forecast error plus the maximum forecast radius of 34 kt winds at that point. This region functions as an error swath within which gale-force winds may occur, based on historical track error. It does not currently account for uncertainty or asymmetry in the wind radii forecast, or for asymmetry in track forecast uncertainty. As a result of adding the forecast gale radius to the climatological area, the swath appears larger than a climatologically based cone such as that used by the National Hurricane Center (NHC – USA), even though it is based on a similar level of uncertainty.

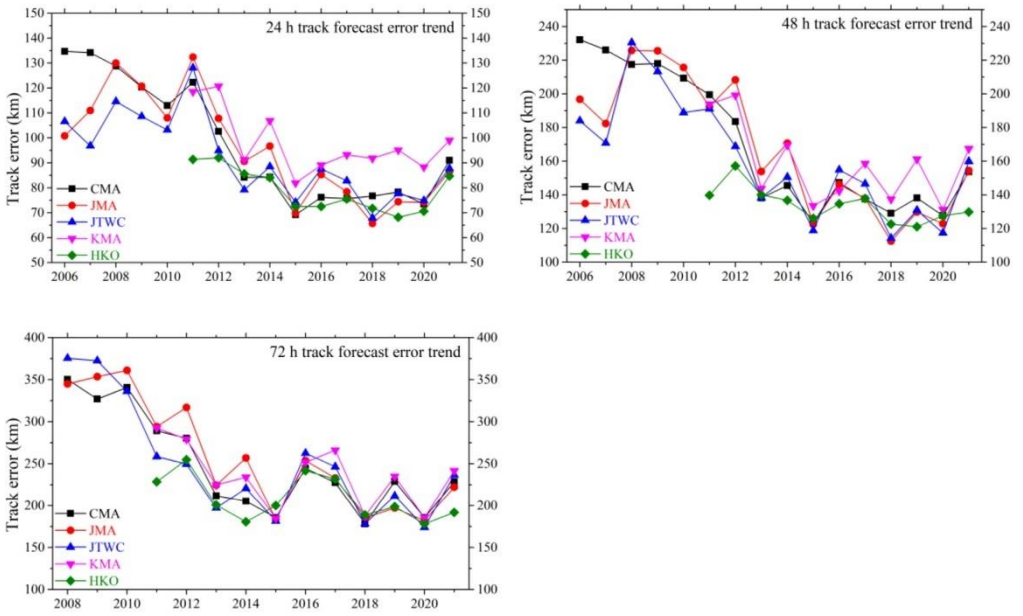


Fig. 4. Position error trends for each official typhoon prediction agency at the lead times of 24, 48, and 72 h (Figure courtesy of Chen and Yang (2022)).

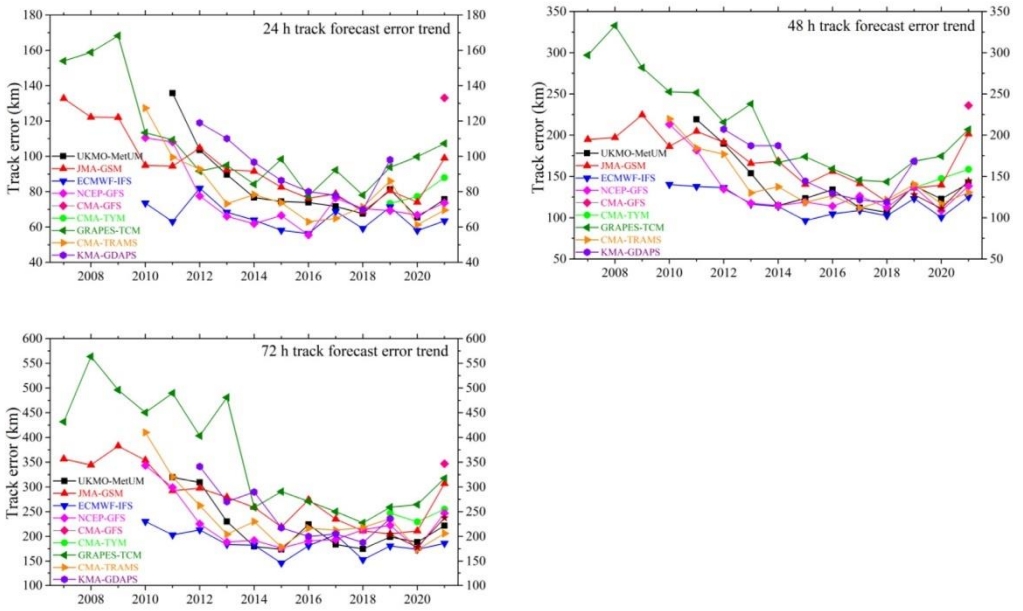


Fig. 5. Position error trends for global and regional NWP models from 2007 to 2021 at the lead times of 24, 48, and 72 h (Figure courtesy of Chen and Yang (2022)).

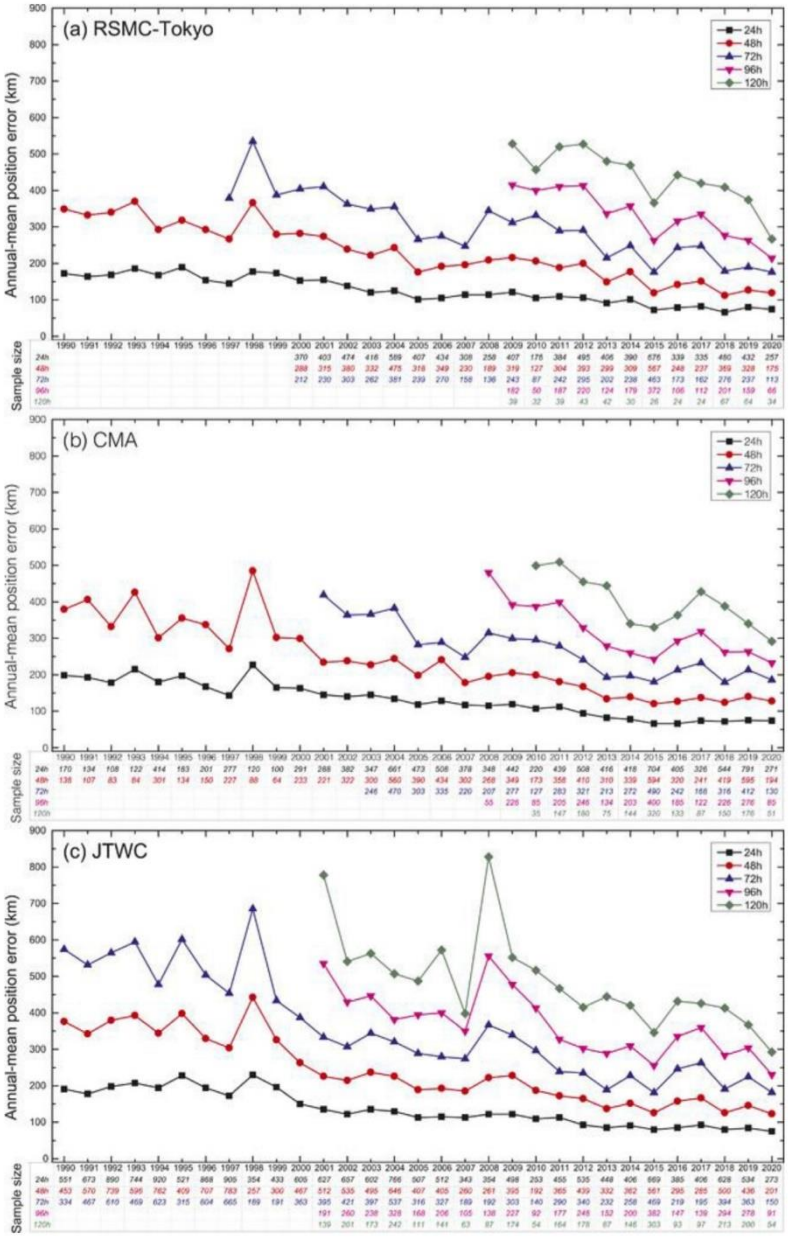


Fig. 6. The annual-mean position error (km) of TC track forecasts from 1990 to 2020 at (a) RSMC-Tokyo, (b) CMA, and (c) JTWC. The sample sizes are listed at the bottom of each chart. Note that the sample size is not available for RSMC-Tokyo before 2000. (Figure courtesy of Yu et al. (2021)).

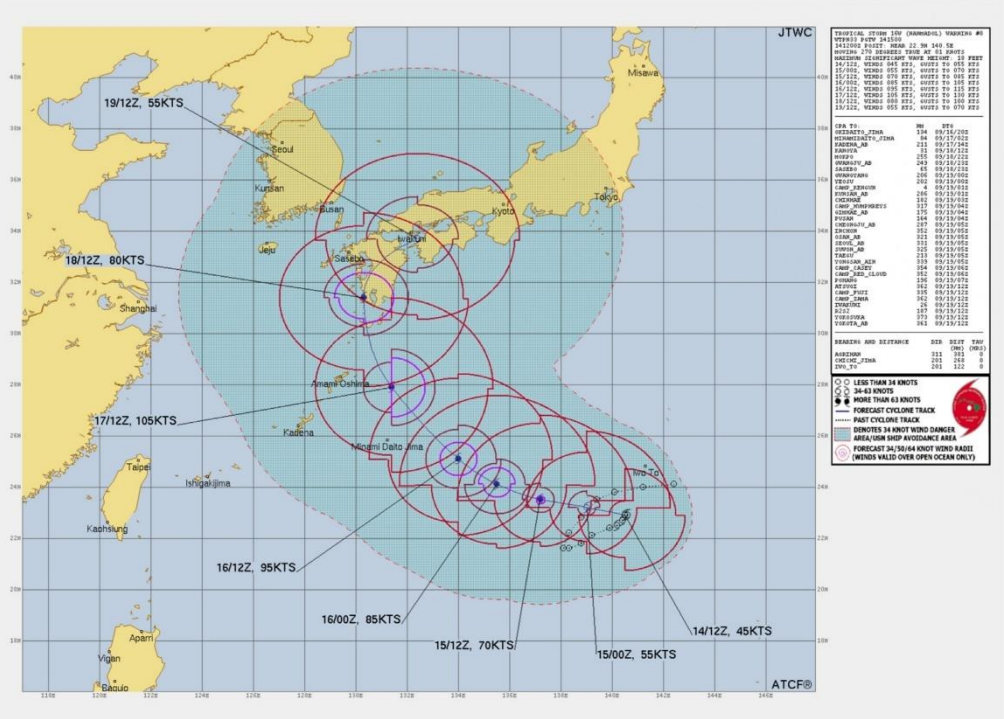


Fig. 7. An example of a JTWC forecast graphic for Tropical Storm 16W, issued at 1500 UTC 14 September 2022. The TC forecast positions are denoted by TC symbols along a solid blue curve. The red, dark red, and purple bounding areas centered on each forecast position denote the forecast radii of 34 kt, 50 kt, and 64 kt winds, respectively. Forecast intensity is annotated for each forecast point in text, and the ship avoidance area is indicated by the blue hatched region. (Image courtesy of JTWC).

4.2. Statistical-dynamical methods

In some cases, the statistically generated uncertainty is combined with information from NWP in order to improve the skill and account for situationally dependent information. This can simply be done by subjectively altering the area generated by climatological error to account for apparent spread of model guidance, or by more elaborate methods.

At some centers, dynamically information is incorporated into the track uncertainty in a subjective fashion. The MetService (NZ) estimate position uncertainty subjectively to encompass the spread of the UK, EC and GFS deterministic forecast positions, then the cone of uncertainty is adjusted to include the majority of the EC ensemble tracks. Other ensemble data from the MOGREPS and GEFS is also assessed, as well as deterministic guidance from NAVGEM, CMC, ACCESS and HWRF, to determine how the forecast policy matches the overall range of model outcomes. This context is useful for communicating track uncertainty in briefings to stakeholders, particularly at longer time frames, so a 'worst case' scenario can be assessed alongside the most likely.

The Australian Bureau of Meteorology (ABoM) used a similar process prior to the 2021/22 tropical cyclone season. The 'first guess' at position uncertainty was a cone generated based on the climatological error at each time step. The forecaster would then overlay deterministic and ensemble track guidance, and manually adjust the area to better reflect NWP guidance. This has since been updated and is detailed in the following section.

4.3. Predominantly dynamical methods

Some centers have already adopted processes for expressing forecast uncertainty by predominantly dynamical methods, and others have plans to make such changes in the future. Typically, this process will generate a Cone of Uncertainty/Forecast Confidence Area using ensemble forecast data as the basis, with some statistical input whenever there is insufficient ensemble data. There can also be bias applied early in track to the analysis position uncertainty.

Unlike RSMC New Delhi, which uses static cones of uncertainty, NCMRWF (India) forecast tracks currently show a color shaded probability of location on their TC forecasts which is based on the NCMRWF global ensemble NEPS-G (shown earlier in Fig. 2). This area is uncalibrated and based on only one model.

At RMSC Tokyo (JMA) the uncertainty at each forecast timestep is expressed via a probability circle, which shows the range into which the center of a TC is expected to move with 70% probability at each validation time. Starting in June 2019, the radius for all forecast times up to 120 h is determined by the multiple ensemble method, which is solely according to the confidence level based on the cumulative ensemble spread calculated using multiple ensemble prediction systems (EPSs) consisting of ECMWF, NCEP and UKMO global EPSs in addition to JMA GEPS (Fukuda and Yamaguchi 2019). The radii are based on the confidence level at each forecast time, which is classified as 'high', 'medium' or 'low', and is based on the degree of ensemble spread. When this method started in 2019, the confidence level was determined from calculating all ensemble spread values from the period 2016 to 2018 and dividing them based on degree of spread into three groups (high, medium, and low) such that the ratio is 4:4:2. A standard radius is used at each forecast timestep based on the confidence level assigned. The radii used for each confidence level is reviewed every few years, but at the time of writing the data from 2016 to 2018 continues to be used.

Since the 2011/12 tropical cyclone season, RSMC La Reunion have incorporated dynamical information into their Cone of Uncertainty (referred to by La Reunion as a 'forecast confidence area'). This has been built around the official RSMC La Reunion track forecast (75% confidence level circles for each lead time) and based on the spread of the ensemble forecast members of the ECMWF ensemble prediction system (EPS). A new version of the forecast confidence area was implemented in 2020 (for the cyclone season 2020–2021) based on SPICy scenarios, developed as part of the *Système de Prévision des Inondations en contexte Cyclonique* (SPICy) project.

In the framework of the SPICy project a methodology was developed at RSMC La Reunion (Meteo-France) in order to generate alternate probabilistic scenarios of tropical cyclones track and intensity forecasts around the official forecast.

Initially two different techniques were considered, one only using climatological errors, and the other purely dynamical using the ECMWF ensemble system (EPS) outputs. A hybrid approach was eventually implemented after being assessed as generally more preferable. A first set of climatology-built scenarios is generated using the statistical distribution of RSMC La Reunion forecast errors. This initial set is then modulated using real-time information provided by the ensemble prediction system (EPS) of the European Centre for Medium-range Weather Forecasts (ECMWF).

The method (described in Bonnardot et al., (2019)) was first developed using five TC seasons of climatological data, from 2011/12–2015/16, including 826 forecasts. These were classified into 45 classes from which alternative scenarios are built around the official RSMC La Reunion forecast. The real time ECMWF ensemble forecast members are then realigned on the RSMC forecast by first calculating differences between each member and the ensemble mean. Each of those scenarios is then assigned to the nearest climatological class and a probability of the previously built climatological scenarios is determined based on the number of realigned ECMWF ensemble members assigned to it. This leads to a decrease in the number of scenarios as climatological classes with zero realigned members assigned are given a probability of zero. In this way, the final ensemble is modulated using the real time ECMWF forecast information to produce a cost-efficient set of approximately 15–30 alternative scenarios (with associated probabilities) that can be delivered within the operational timeframes required.

However, since the article was written the method has slightly evolved. In its latest version, all climatological classes are kept but the initial probability (climatological one) is modulated with respect to the number of ECMWF members associated to each climatological class. Therefore, the final number of scenarios is unchanged at the end of the process (45 classes - but with more classes associated to a zero probability). Those weighted alternate scenarios are then used to determine the confidence circles for each lead time (with a calibrated 75% probability).

For the 2021/22 tropical cyclone season, ABoM changed its method for expressing track uncertainty from a climatologically based cone of uncertainty to a dynamically generated area. For each forecast timestep, an 80% FCA is produced based on a calibrated super ensemble of guidance. An 80% FCA represents the area within which the system will be located at a particular time, to a confidence level of 80%. In other words, it represents the region in which the tropical system center is expected to be located for 80% percent of cases. The 80% confidence level was chosen so that this new method produced similar sized areas when compared to the previous, more subjective, method. The individual FCAs are combined into swaths (FCCs). The FCCs are for the time range starting at analysis time and ending at forecast hour +24, +48, +72 and + 120. The FCCs are the swath of the 80% FCAs in the time range. Due to the skill of the analysis position and its influence on short term forecast uncertainty, the output is weighted to the OFT uncertainty at analysis time, changing to be all super ensemble guidance by forecast hour 36.

Last season, the FCA for super ensemble guidance was derived using an Adaptive Bandwidth Kernel Density Estimate (KDE) to generate a Probability of Location grid from all the guidance fixes, then finding the contour where the sum of the grid cells within the contour gave a FCA which had 80% of analyzed systems within it. This means the FCA's are calibrated.

Included in the forecast system is the option to weight towards the historical error method. This was used when ensemble guidance (either due to model performance or issues with the vortex tracker) was problematic or when the system was weakening and the number of guidance tracks decreased below a certain threshold.

Verification performed when developing this system indicated that a super ensemble of ECMWF EPS, MOGREPS and GEFS resulted in increased spread and once calibrated, was

better performing than any single or two model combination. A lagged super ensemble (including the previous run in the super ensemble along with the current run) further increased the spread and, once calibrated, resulted little change in performance relative to the not lagged super ensemble. However, a lagged super ensemble has the benefits of maintaining some stability between forecasts and increasing the number of ensemble members. For the KDE method in particular, more members mean less uncertainty in the region of the FCA, with simulations indicating a lag of 6 h (2 runs of ECMWF, MOGREPS and GEFS) was optimum.

Forecaster and user feedback from the 2021/2022 season was that the FCA's from the KDE method had "precision exceeding accuracy". The perception was that some FCAs had wavey edges, where they should be straighter or more uniformly curved. In response to this, for the 2022/23 season the FCAs are to be generated using a Gaussian Mixture Model (GMM) instead of the KDE.

The GMM is superior to the existing KDE method in a number of ways:

- GMM appears to outperform the KDE when the number of data points is in the range of the data being used. That is, it gives a more consistent and better estimation.
- GMM delivers shapes which are more regular, which is a desirable property from a customer point of view (See Fig. 8).

For the 2023/24 season, it is expected that forecasters will also have the ability to quality control the super ensemble track guidance prior to the generation of the FCAs. It is also intended that Confidence Areas will be generated for each system of interest at the 00Z and 12Z timesteps out to seven days, even prior to system genesis. This will form the basis of the Tropical Cyclone Outlook (TC genesis product) and will result in there being a smooth transition from outlook products to warning products as a system develops.

5. WMO TC-PFP Tropical Cyclone-Probabilistic Forecast Products (TC-PFP) seamless GDPFS Pilot Project

The Tropical Cyclone-Probabilistic Forecast Products (TC-PFP) effort is a WMO Seamless GDPFS Pilot Project established in response to recommendations from the 2018 IWTC-9 in Honolulu, Hawaii. The main goal of TC-PFP is to coordinate across RSMCs, TCWCs, and other forecast centers to help identify best practice guidance for probabilistic TC forecasts incorporating a value cycle approach. TC-PFP is being implemented in 3 phases: Phase 1 (TC formation and position) began in 2020; Phase 2 (TC intensity and structure) will begin in 2023; and Phase 3 (rainfall and storm surge) will start in 2024. Phase 1 efforts are described in detail by Dunion et al. (2023; anticipated TCRR submission). They surveyed numerous forecast centers around the world and found that many are developing their own probabilistic forecast

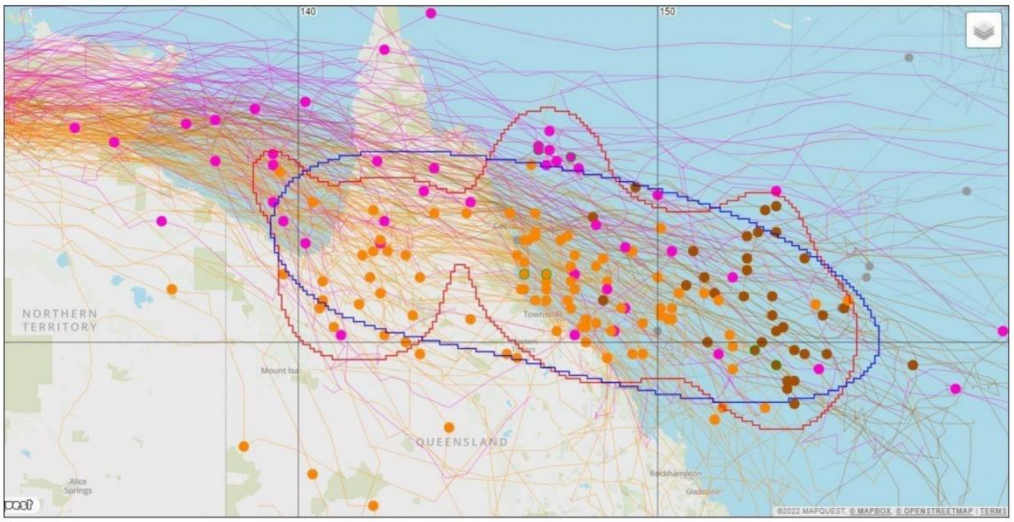


Fig. 8. A comparison of Forecast Confidence Area generation methods. The red line is an uncalibrated KDE calculation, the blue line is an uncalibrated GMM calculation. Note how the GMM produces a more desirable shape whilst still covering a similar sized area. This image is for the system that would eventually develop into Tropical Cyclone Seth. The displayed tracks are from a lagged super ensemble of ECMWF, MOGREPS and GEFS, with the runs from 00Z 26 December 2021, and 18Z 25 December 2021. The time shown is 00Z 31 December 2021 (i.e. 120 h from the most recent ensemble run) (Image courtesy of Australian Bureau of Meteorology).

techniques for predicting TC track and intensity. Several common challenges that many centres face when developing probabilistic TC forecasts were also identified by these communications. Dunion and coauthors, (2023) present recommendations related to probabilistic forecasts of TC formation and position that included enhancing communication between centres, improving accessibility and standardization of probabilistic forecast data, and optimizing product design and communication to end users through a value-cycle approach. It is anticipated that TC-PFP Phase 1–3 efforts should be closely coordinated with IWTC-10 and future IWTC workshops and will be an ongoing process that WMO could help steward.

6. Importance of vortex tracking and clustering algorithms and use of super ensembles

6.1. Vortex trackers

For tropical cyclone guidance, vortex trackers have a strong influence on the characteristics and useability of the data for operations as well as verification. Inconsistency between how systems are tracked between different ensemble prediction systems is an issue that needs to be overcome when developing probabilistic guidance that incorporates multiple ensemble guidance sources.

Whether the tracker is "pre-formation", or "post-formation" has an influence on data availability. Post-formation means vortices are only tracked if initialized with an initial position, which is usually a manual analysis. Pre-formation means vortices are found and tracked in the grids without needing a manual analysis position to initiate the tracker. Pre-formation trackers produce data prior to the formation of a system and therefore have a more uniform amount of data across the 10 days of the forecast (the common length of guidance). Pre-formation trackers can be used to produce pre-formation forecast tracks, genesis, and outlook products. Post-formation trackers only produce data once the system has formed and produce much less data towards the end of the 10 days due to the average life of tropical cyclones meaning that many no longer exist towards the end of the forecast period. Several trackers combine the tracking of both pre-formation and post-formation tracks, with the option of applying different thresholds for each, and are the preferred type.

The threshold at which vortices are tracked also influences data availability and suitability for use in a super ensemble. For example, the GFDL tracker will track very weak systems, and this makes difficult to use pre-formation as it can be hard to distinguish between weak, shallow systems and ones that have the potential to develop. The MOGREPS tracker (Heming, 2017) has a higher threshold for tracking vortices, which means less overall tracks. Fig. 9 compares the data availability from EC (ECWMF EPS) and UK (MOGREPS) and the US (GEFS) over a three-and-a-half-year period. The EC and UK are pre-formation tracks, and the US is post-formation. A data point refers to a unique combination of tropical system, base time and forecast time. The US has less datapoints at forecast hour 12 because it is post-formation; it needs an analyzed system to initiate the tracker. Due to the typical life span of a system at tropical cyclone intensity, this also explains the bigger decrease in data points for the US as the forecast hour increases. The UK

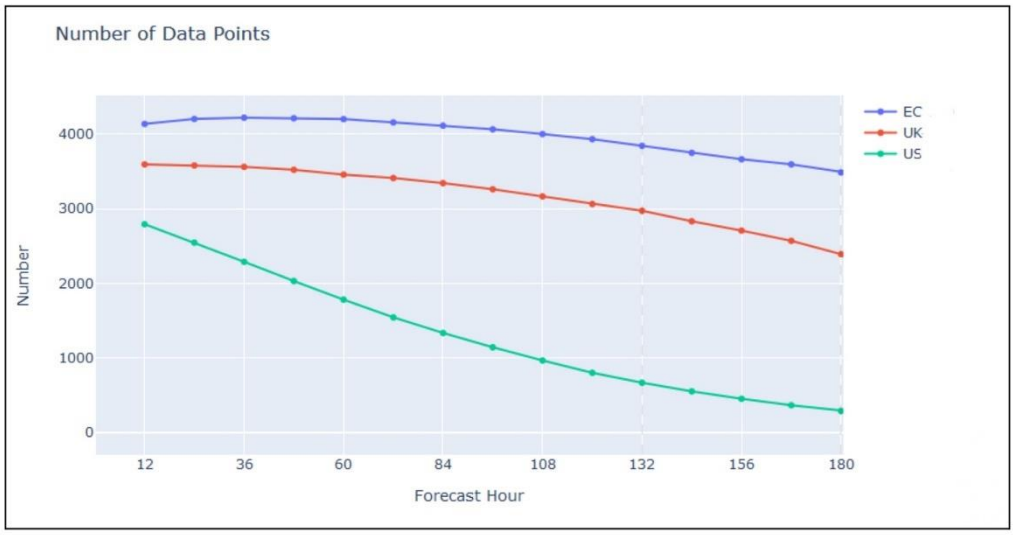


Fig. 9. Chart showing the number of data points available from each ensemble. The data is from January 2019 to September 2022 inclusive; the base times are 00Z and 12Z and the UK data includes the previous base time (18Z for the 00Z run, and 06Z for the 12Z run). The data is for analyzed tropical systems that appear in the track data and the IBTrACS dataset.

(MOGREPS) has less data points than the EC due to the tracker having a higher threshold to track a vortex and also having less members (36 vs 51). Less members means less overall data points that have more than 10 available fixes.

A study by Xiangbo Feng, John Methven and Kevin Hodges from the University of Reading aimed to compare four different tracking methods. This was performed to support the TC-PFP project. The study looked at western North Pacific tropical cyclones in the 2020 typhoon season and applied four different trackers to the same ECMWF EPS data.

The tracking methods compared were:

- Met Office (MO) operational tracker (Heming, 2017). Tracks provided by Helen Titley (Met Office).
- Geophysical Fluid Dynamics Laboratory (GFDL) operational tracker (Marchok, 2021). Tracks provided by Tim Marchok.
- ECMWF (EC) operational tracker (Vitart et al., 2012; Magnusson et al., 2021). Tracks are downloaded from the TIGGE TC database, and are for named systems only.
- Reading (RD) tracker (Hodges et al., 2017).

The study found that the estimate of ensemble spread is relatively insensitive to the tracker used (less than 10 km difference between trackers in the first 6 days of the forecast).

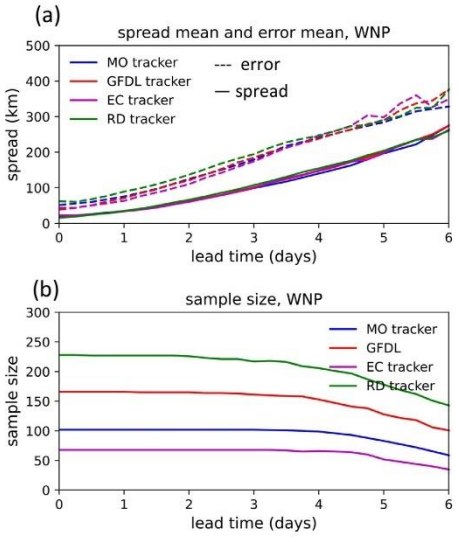


Fig. 10. (a) Ensemble spread (solid lines) and position error for the ensemble mean (dashed lines) averaged over TC forecasts for the western North Pacific basin in 2020 (b) The number of forecasts (i.e., different start dates). in (a). The calculation only includes the forecasts, in which i) ensemble members are matched to observations within the first three days of the forecast and with the mean of four consecutive track points ≤4° away from observation, and ii) the minimum number of matched members over the ensemble forecast members is 30 of 51.

However, the difference in error of the ensemble mean was more pronounced, varying by up to 20–25 km between trackers across the first 6 days. Fig. 10 shows this data. The error in ensemble mean track is also significantly greater than spread showing that the EC ensemble is under-dispersive. In these calculations only forecasts with named storms at t = 0 were included, enabling unambiguous comparison of features identified by different trackers. In each forecast, the error is measured using the great circle distance of ensemble mean position points to the corresponding observed track points (from IBTrACS). The ensemble spread is calculated by calculating the square root of the average of the squared distances (i.e., standard deviation) between track points in the ensemble members to the ensemble mean positions, and then the statistics are composited over all forecasts during the season containing named storms.

Due to the differences in feature identification (both in variables and thresholds used), there was a significant difference in the number of tracks identified by each tracker, even for named storms. The comparison between trackers has been made without normalization of this data (i.e. not every timestep of every storm compared is identified in all four track datasets). This is due to the fact there would not be a suitable sample size for some trackers, and also the assumption that the cases where all four trackers have a position would mostly be confined to mature storms (where the differences in trackers would be less pronounced). By not normalizing the tracks, the data is more comparable to what is currently available in an operational environment.

The differences in mean error between trackers may be related to differences in how the trackers calculate position, and also the sample size differences between trackers (the trackers that have a lower threshold for storm intensity will track weaker systems, which would be likely to introduce larger errors to the ensemble mean).

6.2. Clustering algorithms

When using multiple ensemble sources, or when using preformation tracks, there is a need to ensure that tracks from different ensemble members are each associated with a particular storm. The group of tracks from different ensemble methods diagnosed to represent the same storm are referred to as a cluster. At the Australian Bureau of Meteorology (ABoM), the clustering algorithm applied to ensembles is built upon the method described in Aijaz et al. (2019). This method consists of looping through all ensemble members for each forecast time. When the start of a track from a particular member is found, it is compared to existing clusters, and if it close enough to one then it is added to that cluster. If it is too far away from an existing cluster, then a new cluster is created. At the end of this looping process, a set of clusters will have been created for both existing systems and for systems that have yet to form.

As an overview, the clustering algorithm:

- Aims to minimize the total (sum) distance of a member's tracks to the clusters.

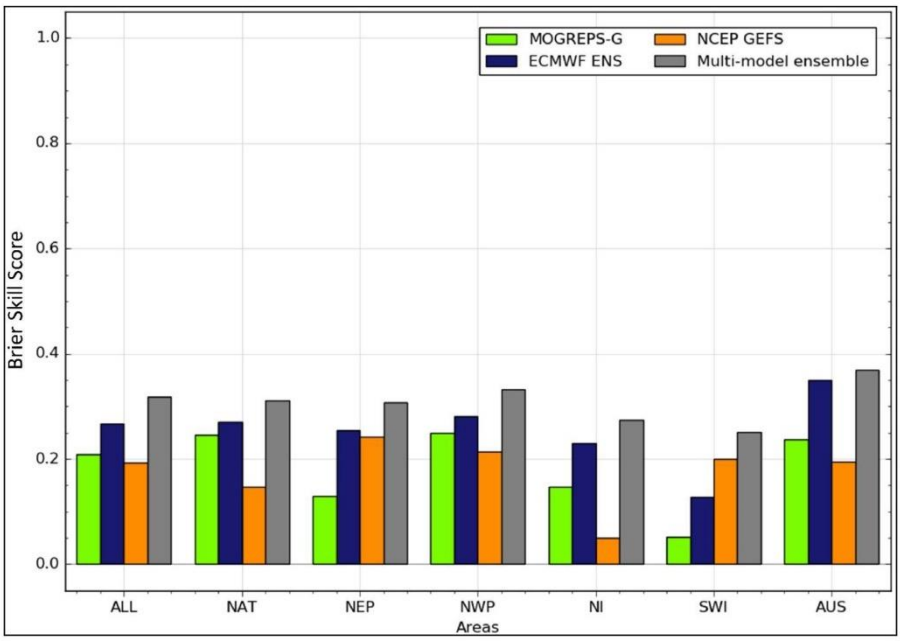


Fig. 11. BSS for the overall track probability forecasts for all named tropical cyclones (left) and for named tropical cyclones in each of the individual six basins, for each individual ensemble and the multi-model ensemble. The reference forecast used in the skill score calculation is the consensus (mean) track of the three deterministic forecasts NAT = North Atlantic, NEP = eastern North Pacific, NWP = western North Pacific, NI = Northern Indian Ocean, SWI = SW Indian Ocean, AUS = Australian region) (image courtesy of Titley et al. (2020)).

- Has a distance which is trajectory based, using a version of a Fréchet Distance, which weights towards the start of the track. The spatial distance is between a track and the mean of the cluster's tracks. The actual distance is a function of the spatial distance and the inverse of the number of unique ensemble members in the cluster, which preferences clusters with more tracks.
- Ensures a member can only have one track per cluster, unless they do not "conflict" (i.e. do not overlap with multiple positions at the same forecast timestep).
- Uses a large distance threshold when determining if a track is "close enough". Typically, it is the "one track per cluster" rule which creates new clusters.

The method can also be used to "cluster" different models and base times together by clustering the mean tracks of each ensemble. This allows for combination into a super ensemble, even prior to system formation, and also for continuity across different forecast issue times. A clustering method that works on systems prior to formation allows for the equivalent of post-formation information to be made available in a pre-formation environment, which allows for system-based forecast products to be issued prior to tropical cyclone formation. For users of the forecast products, this can allow for longer range responses to tropical cyclone events, which in turn can improve decision making.

6.3. Evaluation of multi-model ensemble track probability forecasts

Ensemble forecast systems provide a useful way to assess situationally dependent forecast uncertainty. However, single-model ensembles tend to be under-spread and can fail to provide a range of solutions that covers the full probability distribution. A multi-model ensemble approach can help solve the problem of under-spread, with a set of independent and skillful ensemble providing a more complete representation of the uncertainty (in a similar fashion to how a consensus of deterministic models shows greater skill than each individual member of the consensus). Titley et al. (2020) evaluated the performance of track probability forecasts from a multi-model ensemble of MOGREPS-G, ECMWF ENS and NCEP GEFS across the period from January 2017 to December 2018, and found that this multi-model ensemble increased the skill and

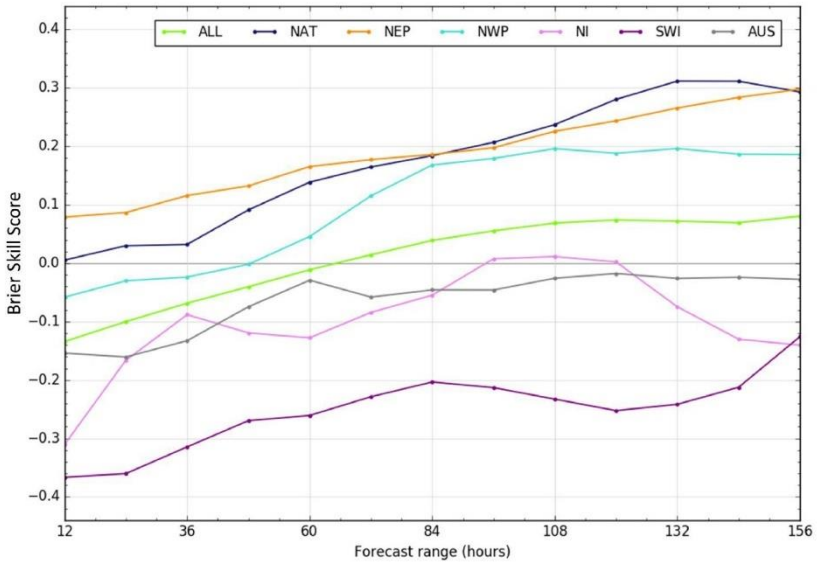


Fig. 12. Brier skill score (BSS) for the MOGREPS-G/ECMWF ENS/NCEP GEFS multi-model ensemble, for all storms (in green) and split into the six tropical cyclone basins, for the 24-h track probability forecasts centered around each forecast range. The reference forecast is the multi-model consensus of the parent deterministic models (image courtesy of Titley et al. (2020)).

value of probabilistic forecasts over the individual ensemble models. This was found to be applicable over all basins, as can be seen in Fig. 11.

Fig. 12 shows how the BSS varies over time compared to the reference multi-model consensus of the parent deterministic models across all basins and on the global dataset. It shows that the comparative skill of the multi-model ensemble becomes greater as the forecast lead time increases. On a global scale, the multi-model ensemble shows positive skill over the multi-model consensus beyond 60 h lead time. It needs to be noted that for some basins, the BSS remained mostly negative throughout the forecast period, even though for these basins the BSS was positive when considered across the entire 7 day forecast period. This is a reflection of the 7-day verification being more forgiving of along track errors (i.e. the track position may be quite accurate, but the timing of movement along the track has more error). The sample sizes were smallest in these basins, and a case-by-case assessment for the worst performing basin (the SW Indian Ocean) found that the deterministic errors were relatively low for the period studied when compared to previous seasons. This would make it more difficult for the multi-model ensemble to outperform the reference consensus.

7. Summary and conclusions

The generation of a forecast track via a consensus of model guidance is a well-established process used by all agencies surveyed for this report. The accuracy of official forecast tracks constructed in this manner continues to improve, though the rate of improvement has slowed in the last four years and it is possible we are reaching predictability limits, at least at shorter lead times. However, the established consensus methodology only depicts a single scenario. There is a growing need to have an accurate understanding of the full range of possible scenarios, especially when considering impact-based forecasts and warnings.

Recent advances and new techniques in track forecasting have centered around improving the depiction of location probability for a system. The best way to do this comes from harnessing the information from skillful ensemble prediction systems. Individual ensemble systems tend to be under spread, but the use of an ensemble of ensembles (a “super ensemble”) can overcome this limitation. Challenges remain in developing a semi-standardized set of best practices for publicly-available probabilistic TC genesis and TC track forecasts and how to tailor both the products and associated communication to suit the requirements of the users.

As agencies move towards the inclusion of dynamic location probability in TC forecasts, consideration should be given to reducing the prominence of the forecast track itself (and even moving away from the concept of a forecast track and towards a true probabilistic representation of potential TC impacts).

Vortex parameter files are an important vehicle for sharing NWP data, allowing forecast centers to implement a super

ensemble approach without needing the technical capability to run vortex tracking algorithms on NWP grids. The content of a vortex parameter file is dependent on the algorithm used. Many tracking algorithms are sensitive to model resolution, and some perform better than others with respect to tracking weak circulations. Further work is needed to support optimization within a standardized framework so that all operational centers are able to deliver to their communities the full value of a super-ensemble-based probabilistic approach to track forecasting.

Acknowledgments

The authors would like to thank their respective affiliate organizations for supporting their participation in IWTC-10. They would also like to thank the WMO World Weather Research Programme and WMO Tropical Cyclone Programme for jointly organizing IWTC-10 and co-chairs J. Courtney and R. Rogers for their leadership and direction.

Additional thanks go out to Philippe Caroff for information on SpiCy, and Takuya Hosomi for information on the JMA forecast process.

Acronyms used in the report

ABoM Australian Bureau of Meteorology
ACCESS-G Australian Community Climate and Earth-System Simulator - Global (ABoM)
ATCF Automated Tropical Cyclone Forecast
AUS Australian
BSS Brier Skill Score
CLIPER Climatology and Persistence Model
CMA China Meteorological Administration
CMC Canadian Meteorological Centre
COAMPS Coupled Ocean/ Atmosphere Mesoscale Prediction System
CONW JTWC track forecast consensus
COU Cone of Uncertainty
EC/ECMWF European Centre for Medium-Range Weather Forecasts
ENS Ensemble
EPS Esnsemble Prediction System
FCA Forecast Confidence Area
FCC Forecast Confidence Cone
GALWEM Global Air-Land Weather Exploitation Model
GDPFS Global Data-Processing and Forecasting System
GEFS/GEPS Global Ensemble Forecast System (USA)
GEPS Global Ensemble Prediction System
GFDL Geophysical Fluid Dynamics Laboratory
GFS/US Global Forecast System
GMM Gaussian Mixture Model
GPCE Goerss probability consensus estimate
HCCA Hurricane Forecast Improvement Program (HFIP) Corrected Consensus Approach
HCCA HFIP Corrected Consensus Approach
HFIP Hurricane Forecast Improvement Program
HFIP Hurricane Forecast Improvement Program
HKO HKO – Hong Kong Observatory
HWRF Hurricane Weather Research and Forecasting
IBTrACS International Best Track Archive for Climate Stewardship
JMA Japan Meteorological Agency
JTWC Joint Typhoon Warning Center
KDE Kernel Density Estimate
km kilometres
KMA Korea Meteorological Administration
kt knots
MOGREPS Met Office Global and Regional Ensemble Prediction System (UK)
NAT North Atlantic
NAVGEM Navy Global Environmental Model (US)
NCEP National Center for Environmental Prediction
NCEP National Centers for Environmental Prediction
NCMRWF National Centre for Medium Range Weather Forecasting
NCMRWF National Centre for Medium Range Weather Forecasting (India)
NCUM NCMRWF Global Unified Model
NEP Eastern North Pacific
NEPS-G NCMRWF Global Ensemble Prediction System
NHC National Hurricane Centre
NI Northern Indian Ocean
NWP Numerical Weather Prediction
NWP Western North Pacific
NZ New Zealand
OFT Official Forecast Track
PE Position Error
RD Reading
RSMC Regional Specialized Meteorological Centre
SPICy Système de Prévision des Inondations en contexte Cyclonique
SWI Southwest Indian Ocean
TC Tropical cyclone
TC-PFP Tropical Cyclone Probabilistic Forecast Products (project)
TCWC Tropical Cyclone Warning Centre
UK/UKMET/UKMO/MO United Kingdom Met Office
USA/US United States of America
WMO World Meteorological Organisation

References

Aijaz, S., Kepert, J.D., Ye, H., Huang, Z., Hawksford, A., 2019. Bias correction of tropical cyclone parameters in the ECMWF ensemble prediction system in Australia. Mon. Wea. Rev. 147 (11), 4261–4285. https://journals.ametsoc.org/view/journals/mwre/147/11/mwr-d-18-0377.1.xml.

Bonnardot, F., Quetelard, H., Jumaux, G., Leroux, M.-D., Bessafi, M., 2019. Probabilistic forecasts of tropical cyclone tracks and intensities in the southwest Indian Ocean basin. Q. J. R. Meteorol. Soc. 145, 675–686. https://doi.org/10.1002/qj.3459.

Chen, G.M., Yang, M.Q., 2022. Verification of tropical cyclone operational forecast in 2021. Tech. Rep. TLFDP-WMO. http://www.tlfdp.net/#/main/verification-reports.

Dunion, J., coauthors, 2023. Recommendations for improved tropical cyclone formation and position probabilistic forecast products (anticipated submission). tropical cyclone Research and review.

Fukuda, J., Yamaguchi, M., 2019. Determining Probability-Circle Radii of Tropical Cyclone Track Forecasts with Multiple Ensembles. RSMC Tokyo – Typhoon Center. *Technical Review No. 21* (April 2019).

Heming, J.T., 2017. Tropical cyclone tracking and verification techniques for Met Office numerical weather prediction models. Met. Apps 24, 1–8. https://doi.org/10.1002/met.1599.

Landsea, C.W., Cangialosi, J.P., 2018. Have we reached the limits of predictability for tropical cyclone track forecasting? Bull. Amer. Meteorol. Soc. 99, 2237–2243. https://doi.org/10.1175/BAMS-D-17-0136.1.

Magnusson, L., Majumdar, S., Emerton, R., Richardson, D., Alonso-Balmaseda, M., Baugh, C., Zsoter, E., 2021. Tropical Cyclone Activities at ECMWF. European Centre for Medium Range Weather Forecasts. *Technical Memo* 888.

Marchok, T., 2021. Important factors in the tracking of tropical cyclones in operational models. J. Appl. Meteorol. Climatol. 60, 1265–1284. https://doi.org/10.1175/JAMC-D-20-0175.1.

Sampson, C.R., Schrader, A.J., 2000. The automated tropical cyclone forecasting system (version 3.2). Bull. Amer. Meteorol. Soc. 81, 1131–1240. https://doi.org/10.1175/1520-0477(2000)081<1231:TATCFS>2.3.CO;2.

Simon, A., Penny, A.B., DeMaria, M., Franklin, J.L., Pasch, R.J., Rappaport, E.N., Zelinsky, D.A., 2018. A description of the real-time HFIP corrected consensus approach (HCCA) for tropical cyclone track and intensity guidance. Wea. Forecast. 33, 37–57. https://doi.org/10.1175/WAF-D-17-0068.1.

Titley, H.A., Bowyer, R.L., Cloke, H.L., 2020. A global evaluation of multi-model ensemble tropical cyclone track probability forecasts. Quart. J. Roy. Meteorol. Soc. 146 (726), 531–545.

Titley, H.A., Yamaguchi, M., Magnusson, L., 2019. Current and potential use of ensemble forecasts in operational TC forecasting: results from a global forecaster survey. Trop. Cyclone Res. Rev. 8 (3), 166–180. https://doi.org/10.1016/j.tcrr.2019.10.005. ISSN 2225-6032.

Vitart, F., Prates, F., Bonet, A., Sahin, C., 2012. New tropical cyclone products on the web. ECMWF Newsl. 130, 17–23.

Yu, H., Chen, G.-M., Zhou, C., Wong, W.K., Yang, M.-Q., Xu, Y.-L., Chen, P.-Y., Wan, R.-J., Hu, X.-R., 2021. Are we reaching the limit of tropical cyclone track predictability in the western North Pacific? Bull. Amer. Meteorol. Soc. 103, E410–E428. https://doi.org/10.1175/BAMS-D-20-0308.1.

A6: WCSSP India report: The predictability of fluvial flooding from recent Indian cyclones in the Global Flood Awareness System (GloFAS)

This appendix contains the full WCSSP India report summarised in Section 7.5.2, with the following reference:

Met Office WCSSP India report: Titley, H.A. (2021). M3.2.1.6: The predictability of fluvial flooding from recent Indian cyclones in the Global Flood Awareness System (GloFAS) (WCSSP India project report).

Met Office

M3.2.1.6: The predictability of fluvial flooding from recent Indian cyclones in the Global Flood Awareness System (GloFAS)

Helen Titley

www.metoffice.gov.uk

© Crown Copyright 2021, Met Office

Met Office

Contents

Contents 2
Executive summary 3
1. Introduction 4
2. Data and methods 5
3. Cyclone case studies 7
3.1. Cyclone Bulbul (6 Nov - 11 Nov 2019) 7
3.2. Cyclone Amphan (16 May – 21 May 2020) 8
3.3. Cyclone Nisarga (1 June – 4 June 2020) 10
3.4. Cyclone Nivar (23 Nov – 27 Nov 2020) 11
3.5. Cyclone Burevi (30 Nov – 5 Dec 2020) 13
4. Ideas for further work and possible areas for collaboration 15
References 17
Appendices 18

© Crown copyright 2021, Met Office

Met Office

Executive summary

This report assesses the ability of ensemble river discharge forecasts from the Global Flood Awareness System (GloFAS) to forecast the fluvial flooding from five Indian tropical cyclones: Cyclones Bulbul, Amphan, Nisarga, Nivar and Burevi. It introduces a new score to assess the forecast discharge peaks across the area affected by the cyclone in terms of the amplitude and timing of the flood peaks. The study adds a different angle on more traditional methods of verifying the global ensemble forecast performance for tropical cyclones, by looking at the forecasts specifically through a hazard-based hydrological lens rather than the more traditional track and intensity forecast metrics. The main findings are as follows:

- GloFAS forecasts were largely able to successfully distinguish between those events that had relatively high levels of fluvial flooding (Nivar and Burevi) and those where the flooding was much less severe (Nisarga and Bulbul).
- In Cyclone Amphan, the forecast had lower scores for forecasting the discharge peaks until the short- range, largely due to the very high number of false alarms, where the forecast predicted the 2-year return period would be exceeded at river points but this was not the case in the reanalysis. The track of the storm looks to be the key control on predictability in this case, with many of the EC ensemble members being to the SW of the observed centre, causing the flooding to be displaced to the west of observed, and also heavy precipitation to be more widespread resulting from slower track forecasts. The ability to correctly forecast the speed of the storm also hinders the predictability of the flooding in the Burevi case, where the stalling of the cyclone in the Gulf of Mannar was key to correctly forecasting the heavy precipitation accumulations. The translation speed of the storm was found to be a key factor influencing the severity of the fluvial flood hazard in tropical cyclone cases in a recent global analysis of landfalling tropical cyclones (Titley et al., 2021), matching the findings in this study.
- The overall forecast performance was better at longer lead times for the two more southerly events impacting Tamil Nadu and Andhra Pradesh (Nivar and Burevi), with good performance 6-8 days prior to the discharge peaks compared to only 2-3 days prior to the discharge peaks in the northern events impacting West Bengal (Bulbul and Amphan).
- There appears to be a boost to the predictability of the fluvial flooding around the time that several of the cyclones make landfall, showing the importance of correctly forecasting the landfall location in order to correctly forecast the area affected by the heaviest precipitation.
- The more detailed evaluation of Cyclone Nivar illustrates that even when the landfall location is well forecast, the details of the precipitation forecast can influence the predictability of the flooding, highlighting the importance of further research to evaluate the ability of the driving ensemble precipitation forecasts to forecast the spatial distribution of rainfall in tropical cyclone cases.

The evaluation methods introduced in this paper will continue to be developed, and used to assess how the uncertainty in tropical cyclone track, intensity and precipitation forecasts, combine with the uncertainty in the hydrological model to influence the final predictability of the flood forecasts. Future work will also explore the use of spatial verification techniques on

© Crown copyright 2021, Met Office

ensemble forecasts of tropical cyclone precipitation. This work would benefit from collaboration with Indian partners to access observations of river discharge or flooding impacts, compare GloFAS forecasts with regional or local hydrological forecasts for these events, and to work together to compare tropical cyclone precipitation verification across global ensembles.

This report starts by introducing the motivation of the work and the data and methods used. It then details the evaluation of the flood forecasts in each cyclone case, before ending with ideas for further work and collaboration.

1. Introduction

Global ensemble forecast models have been shown to have good skill in forecasting the track probabilities of tropical cyclones worldwide (Titley et al., 2020), but less well-studied is their ability to predict the hazards resulting from tropical cyclones, which in the case of fluvial flooding can extend far from the landfall location traditionally focussed on in operational tropical cyclone warnings. This report investigates the predictability of fluvial flooding from tropical cyclones in India, using forecasts and reanalysis data from the Global Flood Awareness System (GloFAS, https://www.globalfloods.eu/, Figure 1). GloFAS is designed to provide a global overview of upcoming flood events to decision makers such as humanitarian organizations as part of the Copernicus Emergency Management Service for floods (https://emergency.copernicus.eu/).

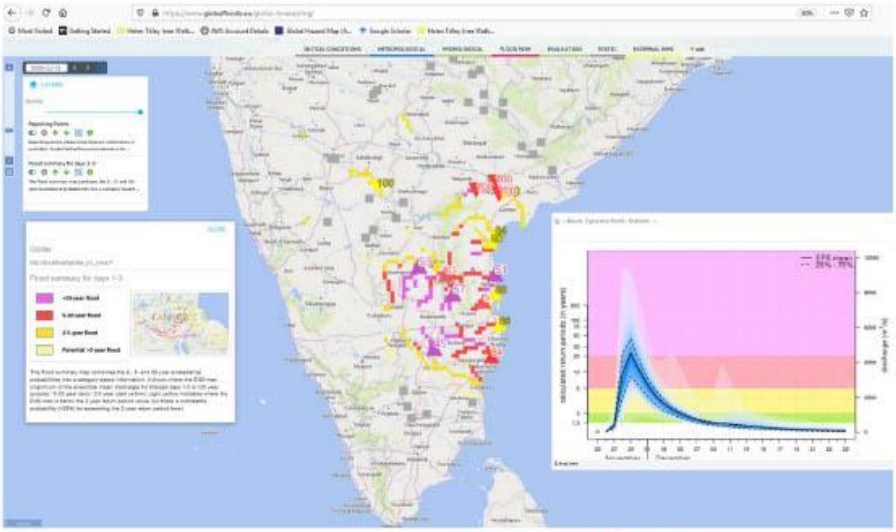


Figure 1. A screenshot from the GloFAS map viewer (https://www.globalfloods.eu/).

© Crown copyright 2021, Met Office

Met Office

In GloFAS, ensemble meteorological forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) are processed by the revised land surface hydrology scheme (HTESSEL) to create land surface runoff fields, with additional hydrological processes such as flow routing provided by the LISFLOOD hydrological model in order to forecast river discharge (Harrigan et al., 2020a). GloFAS provides a daily global gridded discharge forecast out to 30 days with a horizontal resolution of 0.1°. GloFAS forecasts have been used to advise humanitarian agencies and disaster risk reduction activities on the potential fluvial flood hazard ahead of recent tropical cyclone landfalls in Central America and in Mozambique (Emerton et al., 2020). This report assesses the potential of the GloFAS forecasts to help predict fluvial flood hazard for Indian cyclones.

The methodology involves exploring the chain of predictability through the forecasting process for recent landfalling cyclones in India, to evaluate the influence on the overall predictability and uncertainty of fluvial flooding from cyclones from the track, intensity, precipitation, and hydrological components of the forecasting chain. In this study the river discharge forecast hydrographs are scored for each member against discharge reanalyses from each forecast run in the lead up to the event. Comparing the evolution of the hydrological forecast skill across forecast runs to the evolution of the TC track, intensity and precipitation forecasts can help to identify the key factors controlling the predictability and uncertainty in the forecasts for each case, and uncover whether there are common features across all cases. The EC ENS and GloFAS forecasts for five Indian cyclones cases from 2019 and 2020 are analysed, to investigate how well GloFAS predicts the levels of fluvial flooding experienced in each case, and which aspects of the forecasts particularly influenced the predictability. This is followed by ideas for further work, and suggestions for areas where collaboration with Indian partners may be useful.

2. Data and methods

This report covers five landfalling Indian cyclones, listed in Table 1. These five cyclones have been selected as there is GloFAS v2.1 data available for these dates so that the forecasts are consistent with the verifying GloFAS-ERA5 v2.1 reanalysis (Harrigan et al., 2020b). Verifying river discharge against observations is challenging given the relative sparsity of river discharge observations and difficulties accessing this data. Therefore, this study uses the GloFAS-ERA5 global river discharge reanalysis as a proxy for river discharge observations. Although GloFAS-ERA5 is partially dependent on model-derived precipitation and soil moisture, the groundwater and river routing parameters were calibrated against daily river discharge from 1287 observation stations worldwide (Hirpa et al., 2018) and a recent comprehensive global evaluation found it was skilful against a mean flow benchmark in 86% of catchments (Harrigan et al., 2020b). The use of GloFAS-ERA5 reanalysis data allows an analysis of the flooding across the whole area affected by each TC, rather than just at isolated points on the river network.

© Crown copyright 2021, Met Office

Met Office

Cyclone name	Dates
Bulbul	6 November 2019 – 11 November 2019
Amphan	16 May 2020 – 21 May 2020
Nisarga	1 June 2020 – 4 June 2020
Nivar	23 November 2020 – 27 November 2020
Burevi	30 November 2020 – 5 December 2020

Table 1: The Indian cyclones that are included in this report.

A new score is proposed for use in scoring discharge hydrographs in tropical cyclone cases, here termed the Titley score. The Titley score assesses the discharge forecast from each ensemble member according to how well they forecast the timing and amplitude of the peak in the hydrograph following the passage of the tropical cyclone. The discharge forecast from each GloFAS ensemble member at each river point is scored against the reanalysis discharge using the following calculation:

Titley score = Timing score * Amplitude score

Timing score = (ndays - abs(fc_max_index - an_max_index)) / ndays)
ndays: the number of days in the hydrograph (15 in this study)
fc_max_index: the index of the date with the discharge peak in the forecast
an_max_index: the index of the date with the discharge peak in the reanalysis

Amplitude score = ((2*an_max_value) – abs(fc_max_value – an_max_value)) / (2*an_max_value)
an_max_value: the discharge value at the peak in the reanalysis
fc_max_value: the discharge value at the peak in the forecast

Before calculating the Titley score, the GloFAS forecasts are masked so that only those points that were within 500km of the cyclone's track are scored. The forecasts are then filtered so that only river points with an upstream area greater than 1000km2 are included in the score calculations. This is because GloFAS is designed for use in medium to large river catchments. In order to focus on the forecasts where a significant flood event was either forecast or observed, a final filter is applied so that only those points where the 2-year return period is exceeded in either the forecast or the reanalysis are included in the score calculations. The return period thresholds were taken directly from GloFAS, where they are determined from the GloFAS-ERA5 river discharge reanalysis, by fitting a Gumbel extreme value distribution on the annual maxima time series over the 1979-2018 period (Alfieri et al.,2019; Harrigan et al., 2020a; Zsoter et al., 2020). The Titley score is then used to score the discharge forecast from each ensemble member at each river point against the reanalysis discharge. The amplitude component and timing component of the Titley score each range from 0:1, where 1 is perfect, and as the Titley score is a product of these two component scores, it too ranges from 0:1, where 1 is a perfect forecast of the discharge peak. The average Titley score across all the included river points is then calculated for each ensemble member from each forecast, and a box plot is produced which summarises the ensemble performance in predicting the fluvial flooding across several forecast runs.

© Crown copyright 2021, Met Office

Met Office

In addition to calculating the Titley score at all river points where the 2-year return period is exceeded in either the forecast or the reanalysis, these points are categorised as follows:

Correct event: Return period threshold is exceeded in both the reanalysis and the forecast
Missed event: Return period threshold is exceeded in the reanalysis and not in the forecast
False alarm: Return period threshold is exceeded in the forecast and not in the reanalysis

The hydrological performance is also assessed spatially by comparing the points at which key return periods were exceeded in the reanalysis with maps displaying the forecast probability for exceeding the return period thresholds. These maps are shown alongside plots of the precipitation associated with the storm (observed via GPM and forecast by EC ENS), and with forecast centre plots, in order to analyse which aspects of the forecasts contributed to the predictability of the flooding in that case. Hydrographs for key rivers in each cyclone case are also included. Each of the five cases will now be considered in turn.

3. Cyclone case studies

3.1. Cyclone Bulbul (6 Nov - 11 Nov 2019)

Very Severe Cyclonic Storm Bulbul made landfall near Sagar Island in West Bengal, around 18:30UTC on 9th November 9th 2019. The storm produced heavy rain over many parts of West Bengal, including over 100mm over Kolkata. Bulbul rapidly weakened into a deep depression on 10th November, as it began to move over Bangladesh. In West Bengal, approximately 3.5 million people were directly affected by the cyclone, with 14 fatalities, and over 500,000 homes and 1.5 million hectares of crops damaged or destroyed. In Bangladesh 25 people died and 18,000 houses were destroyed. The fatalities were largely due to wind impacts.

Appendix 1A shows the observed and forecast data for Cyclone Bulbul. Only a relatively small area in West Bengal and Bangladesh exceeded the 2-year GloFAS flood return period in the GloFAS reanalysis. These areas are shown by GPM to have experienced the highest storm total precipitation (> 250mm). 1 day ahead of landfall, the position of the storm is largely well forecast, albeit with several of the forecast centres being slightly to the west of the observed landfall location. There are high forecast probabilities for significant rainfall accumulations in the area affected, and c. 80% probability shown in the GloFAS forecasts for the affected area. However, the forecasts from 3 days ahead of landfall see the forecast centres mostly displaced to the south west of observed, and the probability of high rainfall accumulations are lower over the Ganges delta. The GloFAS probability is now lower than 20% across much of the area that was shown to have exceeded the 2-year return period in the reanalysis. This becomes even more pronounced in the forecast 5 days before landfall, when there are very few forecast centres close to the landfall location, which results in very low probabilities for heavy rainfall or significant flooding over the area that was affected. Appendix 1B shows hydrographs from two river points, on the Mani River in West Bengal, and the Arial Khan River in Bangladesh. Both are relatively small rivers with the chosen points having upstream areas between 1,000-2,000km2. Both points exceed the 2-year return period in the GloFAS reanalysis. At the

© Crown copyright 2021, Met Office

Met Office

shortest lead-time shown, 2-3 days before the discharge peaks, most of the members are predicting the 2-year threshold to be exceeded, with many predicting the discharge peak to be higher than observed. However, the run from 48 hours earlier shows a much-reduced signal for severe flooding, with only six members forecasting the 2-year return period to be exceeded on the Mani River, and no members for the Arial Khan river point.

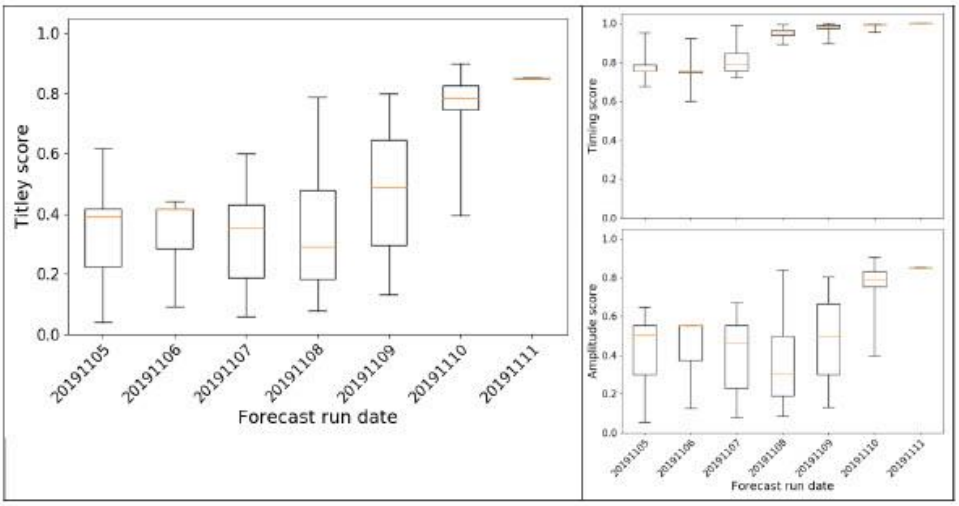


Figure 2: Cyclone Bulbul: Area-averaged Titley score, and its Timing and Amplitude components, in the GloFAS forecasts from 05/11/2019-11/11/2019 (landfall was late on 9th). NB. These plots display the average score across all river points where the 2-year return period is exceeded in the forecast or reanalysis. In this case there were 10 river points where the 2-year return period was exceeded in the reanalysis, along with a varying number of false alarm points.

Figure 2 shows the average Titley Score for each of the ensemble members on each of the forecast runs in a box-whisker plot. On the right are the same scores broken down by the timing and amplitude components. The timing of peak discharge values is very well forecast from 8th but the amplitude and overall score stay relatively low until there is an improvement on 9th, and then another big improvement for the 10th November run (which is also the time of landfall). The median Titley score first exceeds 0.5 in 00UTC 9th run, which is 2-3 days ahead of the peak discharge.

3.2. Cyclone Amphan (16 May – 21 May 2020)

Cyclone Amphan was the strongest tropical cyclone to strike the Ganges Delta since Sidr in 2007. Around 1200UTC on 20th May 2020, Amphan made landfall as a Very Severe Cyclonic storm near Bakkhali, West Bengal with winds of 155 km/hr. As it moved further inland, Amphan rapidly weakened. Causing over US$13 billion of damage, Amphan is the costliest cyclone ever recorded in the North Indian Ocean. Ahead of the storm's landfall, approximately 4.2 million people were evacuated in coastal India and Bangladesh. 128 fatalities are attributed to

© Crown copyright 2021, Met Office

Met Office

Amphan, with 98 of these in India. In West Bengal at least 86 people died, and the state government estimated that the storm caused at least US$13.5 billion in damage and directly affected 70 percent of the state's population. An estimated storm surge of 5 m inundated a wide swath of coastal communities including many of the islands in the Sundarbans. Wind gusts of 190 km/hr were recorded by the Alipore observatory, Kolkata, West Bengal, damaging homes and uprooting trees and electric poles. At least 19 people were killed in Kolkata, where 236mm of rain was recorded, triggering widespread flooding.

Appendix 2A shows the observed and forecast data for Cyclone Bulbul. There are two main areas where the GloFAS reanalysis indicates flooding occurred, in Eastern West Bengal, and northern Bangladesh. This matches well with the short-range GloFAS forecasts from 00UTC 20/11/20. The forecasts from 18/11/20 meanwhile are over-forecasting the area exceeding the 2-year return period, indicating flooding too far to the west and north in West Bengal. This is likely due to the displacement in the forecast centres to the west of observed. There is also some indication that the forecast track is slower in the EC ENS, which may be worsening the over-forecasting of the flooding in some members as a slower storm will give higher precipitation accumulations. At longer lead times the forecast track uncertainty is greater, with most members to the SW of observed, leading to false alarms for flooding in the state of Odisha. Appendix 2B shows forecast and reanalysis discharge hydrographs from two river points, on the Haldi River and the Hugli (Hooghly) River in West Bengal. The Haldi is a much smaller river and a tributary to the much larger Hugli. Only the Haldi river point exceeds the 2-year return period in the reanalysis, with only a small rise seen in the larger river. In the forecasts from 19/05 and 17/05, however, many members can be seen to be over forecasting the amplitude of the discharge peak in both rivers, and forecasting the larger river to also experience flooding. In the earliest forecast shown (15/05/20), most members are not forecasting significant flooding, but there are still several members indicating the possibility of a flood event larger than shown in the reanalysis.

Figure 3 shows the average Titley Score for each of the ensemble members on each of the forecast runs in a box-whisker plot. On the right are the same scores broken down by the timing and amplitude components. The timing of peak discharge values is well forecast throughout, but the low amplitude scores dominated the overall score, which remains low until a huge improvement for the 00UTC 20th run, on the day of landfall, and 2-3 days ahead of the peak discharge values. In this event the most significant factor influencing predictability was the displacement of the main area of rain slightly to the west of observed, which led to many false alarms where flooding was not shown in the reanalysis.

© Crown copyright 2021, Met Office

Met Office

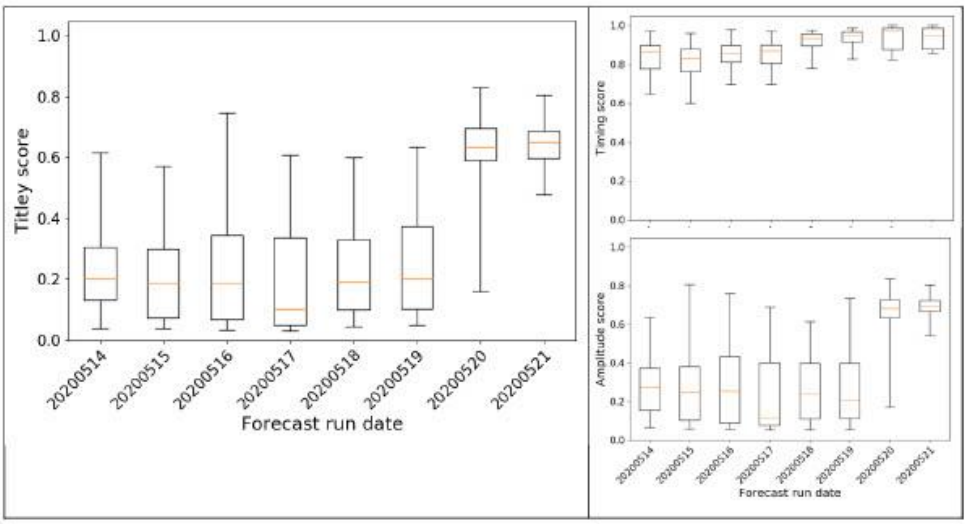


Figure 3: Cyclone Amphan: Area-averaged Titley score, and its Timing and Amplitude components, in the GloFAS forecasts from 14/05/2020-21/05/2020 (landfall was on 20th). NB. These plots display the average score across all river points where the 2-year return period is exceeded in the forecast or reanalysis. In this case there were 34 river points where the 2-year return period was exceeded in the reanalysis, along with a varying number of false alarm points.

3.3. Cyclone Nisarga (1 June – 4 June 2020)

Severe Cyclonic Storm Nisarga made landfall at 0700UTC 3rd June 2020 near the town of Alibag, in Maharashtra, at peak intensity, with nearby Ratnagiri recording a wind speed of 110 km/hr. Nisarga caused 6 deaths, with over 5,000 hectares of land damaged. The State Government of Maharashtra put the total damage from Nisarga at US$800 million.

Appendix 3A shows the observed and forecast data for Cyclone Nisarga. The GloFAS reanalysis is indicating that only a very few points exceeded the 2-year return period, and not for points with upstream area greater than 1000km2. The GloFAS forecasts at short range show low probabilities of flooding occurring in some areas along the Maharashtra coast, but this is less than in the short range forecasts in the other cases, indicating that it is correctly highlighted a lower level of risk for fluvial flooding in this case. At longer lead times very low probabilities of flooding are forecast. It is notable that the EC ENS precipitation forecasts extend the area of high precipitation too far to the south along the coast, and underestimate the precipitation in northern Maharashtra in the earlier forecasts, where many members have the forecast centre for Nisarga displaced to the south. Appendix 3B shows forecast and reanalysis discharge hydrographs from the Savitri River, close to the landfall site in Maharashtra. At all lead times the GloFAS forecasts are consistent with the reanalysis in showing that the 2-year return period will not be exceeded on this river in the Nisarga event (although a handful of members forecast a flood event much later in the forecast, unrelated to the cyclone).

© Crown copyright 2021, Met Office

Met Office

The Titley score plots are not included for this event as there were no river points where the reanalysis discharge exceeded the 2-year return periods. Overall, GloFAS successfully predicts that there will not be a significant flood hazard for this tropical cyclone.

3.4. Cyclone Nivar (23 Nov – 27 Nov 2020)

On 25th June 2020, Cyclone Nivar reached its peak intensity of 120km/hr, making it a Very Severe Cyclonic Storm. It then weakened ahead of making landfall in Marakkanam, Tamil Nadu, close to Puducherry. It then gradually weakened and entered back into the Bay of Bengal as a Deep Depression, before making a second landfall on the Andhra Pradesh coast on 26th June. The Cyclone brought very heavy rains over northern Tamil Nadu ahead of and after landfall. Many areas including Madipakkam, Velachery, Adambakkam and suburbs around Tambaram and low-lying regions along the river Adyar were flooded. India's central water commission said that several rivers were above the danger mark including the Araniyar river in the Tiruvallur district of Tamil Nadu. Several areas were flooded in Puducherry, while in Andhra Pradesh, the rainfall made significant impact on the districts of Chittoor, Prakasam, Kadapa and Nellore with 112,000 people affected. Five people were reported dead in Tamil Nadu, and eight in Andhra Pradesh.

Appendix 4A shows that a large area of eastern Tamil Nadu and Andhra Pradesh exceeded the 2-year return period in the GloFAS reanalysis, making this cyclone the most severely affected by flooding of the five being covered in this report. The forecast from 25/11/20 gives high probabilities of flooding in the affected regions, with a good match also shown between the EC ENS precipitation forecasts and the GPM observations. At earlier lead times, although there is still a significant forecast signal for flooding, the probabilities are lower for some of the inland areas that were affected. A week ahead of landfall (00UTC 19/11/20) the signal for flooding has largely disappeared from GloFAS, with the precipitation being too far to the south. By this point the cyclone is only present in a third of members with few members close to the observed location. Appendix 4B displays forecast and reanalysis discharge hydrographs from two of the most severely affected sites, the Cheyyuru River, inland in Andhra Pradesh (a tributary of the Penna River), and the Palar River near the coast, in Tamil Nadu. The timing of the peaks in both rivers (27th and 28th November respectively) is well forecast from the runs on the 24th and the 22nd. The amplitude of the peak is within the ensemble spread, but there is a large range in the amplitude of the forecast peak in the GloFAS members. At longer lead times e.g. 20th November, only one forecast member correctly forecasts the 20-year return period to be exceeded at each site.

Figure 4 shows the average Titley Score for each of the ensemble members on each of the forecast runs in a box-whisker plot. The scores are higher at longer lead times than was seen in the Bulbul and Amphan cases, with the median first going above a score of 0.5 in the 00UTC 21st November run, four days ahead of landfall, and 6 days ahead of the discharge peaks. There is a further jump in forecast skill from 00UTC 26th November, just after the storm has made landfall.

© Crown copyright 2021, Met Office

Met Office

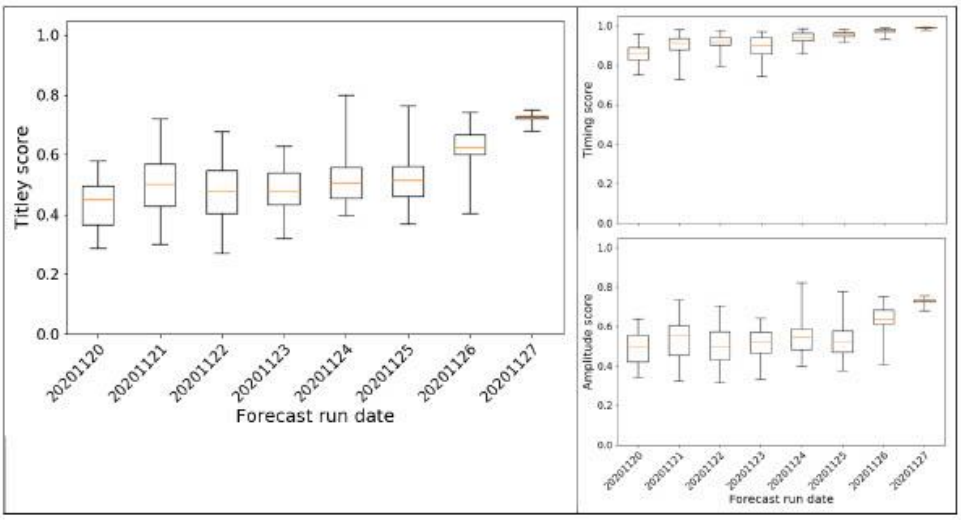


Figure 4: Cyclone Nivar: Area-averaged Titley score, and its Timing and Amplitude components, in the GloFAS forecasts from 20/11/2020-27/11/2020 (landfall was late on 25th). NB. There were a total of 184 river points where the reanalysis discharge exceeded the 2-year return periods, and these are included along with a varying number of false alarm points.

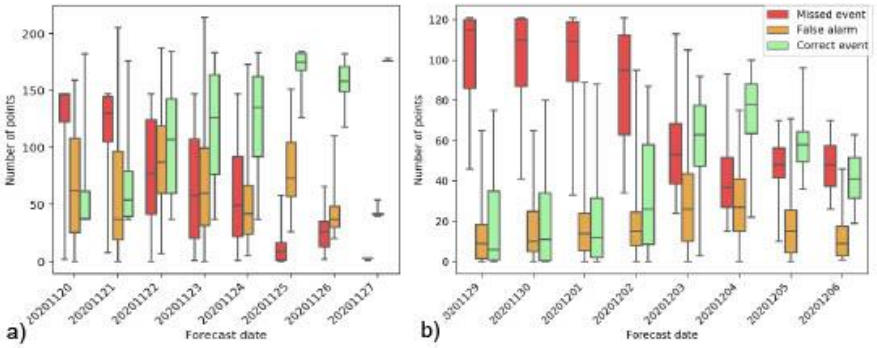


Figure 5: Number of missed event, false alarm, and correct event points in each of the ensemble members across forecast runs for the Cyclone Nivar case (a) and the Cyclone Burevi case (b).

Out of the five cases considered in this report, Cyclone Nivar led to the most widespread flooding hazard and impacts. Hence it is interesting to examine the forecasts for this event in more detail. Figure 5a shows the number of missed event, false alarm and correct event river points across the ensemble for each forecast run, using a 2-year discharge return period exceedance as the event definition. At longer forecast runs it is interesting to note that the ensemble is still forecasting significant flooding, but there are more missed events and false alarms than correctly forecast points, showing that there is uncertainty in the track and its

© Crown copyright 2021, Met Office

Met Office

associated precipitation. This means that many members are predicting the flooding to be displaced from where it was observed. As the forecast lead times get shorter, the number of river points classed as misses and false alarms reduce and the number of correct events increase, becoming higher than the misses and false alarms on the 00UTC 22nd November run.

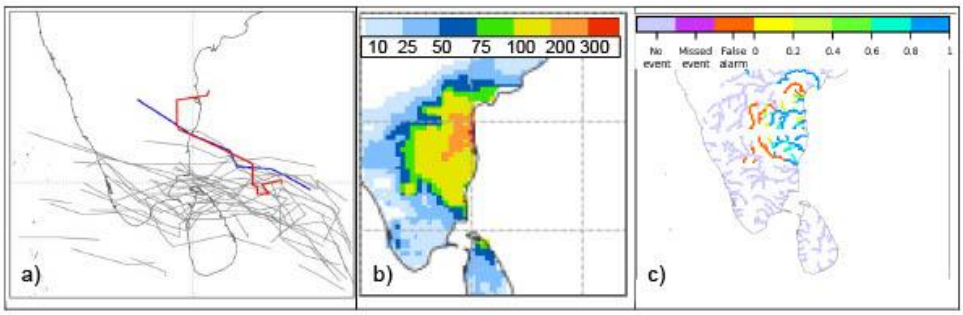


Figure 6: Detailed information on one of the best performing GloFAS members (member 6) from 00UTC 20/11/2020: a) Forecast track for member 6 (blue) against observed (red) and other members (grey); b) precipitation forecast for first 10 days of the forecast; c) mapped Titley scores for member 6 for this run.

Another interesting way to examine the forecasts is to use the Titley scores of the individual ensemble members to identify members with particularly good and poor forecasts. In this case, the hydrographs from 00UTC 20th November (Appendix 4B) showed that member 6 produced a good forecast at both river points. The detailed forecasts for this member are examined in Figure 6. The forecast track is one of the closest to the observed, whereas most of the members have tracks too far to the south. The precipitation is a reasonable match for the GPM totals, although the area over 200mm is all in the north, whereas GPM indicates a more complex pattern. The map of the Titley scores for member 6 shows that there are no 'missed event' points, but there are some false alarms, particularly towards the north of the affected area where precipitation was over-estimated, and further inland, perhaps due to the track continuing inland when in fact Nivar re-emerged back into the Bay of Bengal.

3.5. Cyclone Burevi (30 Nov – 5 Dec 2020)

Cyclonic Storm Burevi made landfall in Sri Lanka on December 2nd 2020, the first tropical cyclone to make landfall in Sri Lanka since 2014. After weakening over Sri Lanka, Burevi exited into the Gulf of Mannar early on December 3rd. Burevi stalled just west of Sri Lanka as it became stuck in a col between two subtropical ridges before eventually dissipating on December 5. In Sri Lanka, more than 75,000 people were evacuated ahead of the storm, which damaged around 3,000 houses. In southern India, prolonged and heavy rainfall led to flooding in Tamil Nadu and Puducherry. Burevi caused at least 11 fatalities.

The top-left image in Appendix 5A shows that a large area of Tamil Nadu exceeded the 2-year return period in the GloFAS reanalysis, making this cyclone one of the cases with the most associated fluvial flooding in India, despite not make landfall in the country. Most of the

© Crown copyright 2021, Met Office

Met Office

precipitation that resulted in fluvial flooding occurred while it was stalled in the Gulf of Mannar. Interestingly, in this case there is not such a close match between the areas most affected by flooding (south central Tamil Nadu) and the areas with the heaviest precipitation according to GPM (coastal eastern Tamil Nadu). The flooded areas have high forecast probabilities from GloFAS on 03/12/20, but these are much lower in the earlier forecast runs (note the large purple area shown on the earliest forecast run is the flooding from the previous storm, Cyclone Nivar). The stalling of the cyclone, which is so key to the heavy precipitation accumulations, is not well forecast in the earlier forecast runs of EC ENS, as shown by the large along-track spread in the centre locations. Appendix 5B displays forecast and reanalysis discharge hydrographs from two of the most severely affected sites, the Vaippra River in southern Tamil Nadu, and the Amaravati River, inland in Tamil Nadu and a tributary of the Kaveri River. Even at short lead times, the timing of the peak at the Vaippra River is too early in many forecast members, with more spread in general in the timing of the peak in both rivers than seen in the other cases. At longer lead times only a handful of members forecast the peak to be as high as in the reanalysis.

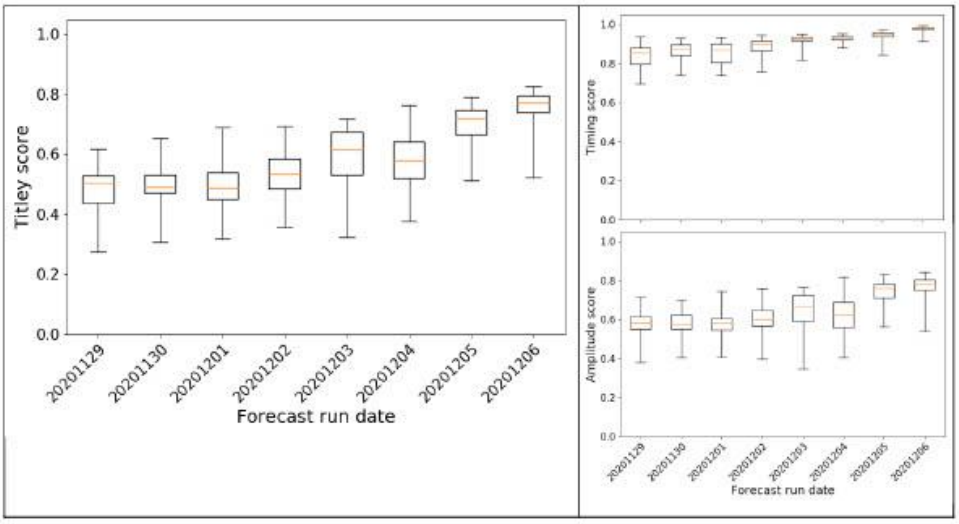


Figure 7: Cyclone Burevi: Area-averaged Titley score, and its Timing and Amplitude components, in the GloFAS forecasts from 29/11/2020-06/12/2020 (stalled on 3rd). NB. There were a total of 121 river points where the reanalysis discharge exceeded the 2-year return periods, and these are included along with a varying number of false alarm points.

Figure 7 shows the average Titley Score for each of the ensemble members on each of the forecast runs in a box-whisker plot. Again, the scores are relatively high compared to the Bulbul and Amphan cases, with the median value close to or above 0.5 throughout. Unlike some of the other cases there aren't any sudden jumps in the scores, perhaps because this storm didn't make landfall in India, which seems to have resulted in forecast improvement in the other cases.

© Crown copyright 2021, Met Office

Met Office

Compared to Cyclone Nivar, the Burevi case has fewer false alarms and more missed events (see Figure 5b), and the correct events 'overtake' misses at a shorter lead time. This may be due to the difficulties the EC ENS forecasts had in correctly predicting the stalling of Burevi after it had passed Sri Lanka, meaning it was harder for the precipitation accumulations to be correctly forecast and the downstream flood hazard to be predicted at longer lead times.

4. Ideas for further work and possible areas for collaboration

This work is ongoing, and the main components of the evaluation method will continue to be developed:

- The box-and-whisker-style summary of the performance of GloFAS in the run-up to the event will be supplemented by similar box-and-whisker plots for TC track, intensity, speed, and precipitation forecasts. A comparison of these plots will allow an assessment of how the uncertainty in each component filters down into the final uncertainty and predictability of the flood forecasts.
- Cyclone tracks and centre plot can be coloured by their resulting hydrological forecasts or verification scores, and the member-specific analysis will be further developed to highlight which members had particular good or poor hydrological skill on a particular run. This will allow for a more detailed evaluation of what aspects of these forecasts led to their hydrological solutions, and how these vary from place to place and from run to run.

Each of these aspects adds a different angle on more traditional methods of verifying the ECMWF ENS global ensemble forecast performance for tropical cyclones, by looking at the forecasts specifically through a hydrological lens rather than the more traditional track and intensity forecast metrics.

Further work is then planned in the following two areas:

1) This work has highlighted the vital importance of the precipitation forecasts for tropical cyclone being correct in terms of intensity and spatial distribution of the rainfall in order to maximise hydrological forecast skill. Very few studies have been carried out to verify spatially the precipitation forecasts from tropical cyclones from ensemble forecast data. Work will now start exploring the use of spatial verification techniques on ensemble forecasts of tropical cyclone precipitation.
2) The tropical cyclone flood forecasts and reanalysis, along with the static GloFAS return-period inundation maps, will be utilised in a wider piece of work to compare methods for how we describe and calculate impact for a range of high-impact rainfall and tropical cyclone flood-related events in India in order to support district-level impact-based forecasting.

Four possible areas are highlighted where collaboration between Indian and UK partners would be very beneficial to this work:

i) In this study the 'truth' or verifying value of discharge was taken from the GloFAS-ERA5 reanalysis, as utilising discharge observations is very challenging due the variability in

© Crown copyright 2021, Met Office

availability and consistency of discharge observation data. The use of reanalysis allows a comparison between events in a consistent way, and enables an evaluation of flooding across the whole area affected by each TC, rather than just at isolated points on the river network. However, the reanalyses are partially dependent on model-derived precipitation and soil moisture. An additional complication is that on managed rivers, where dams and reservoirs are in operation, the discharge levels and flooding will be impacted my water management decisions on these rivers. It would be useful to be able to collaborate with Indian partners to facilitate access to observations of river discharge on some of the key rivers included in this report. This would allow a comparison of the reanalysis to more independent observations. If satellite images of flood extent were available it would also be interesting to compare these to the inundation maps from the 2-year and 5-year return period inundation maps that are available for GloFAS, as a further way of validating the forecasts and reanalyses.

ii) If any partners have records of any impacts of fluvial flooding in cyclone cases, it would be very interesting to compare this impact data to the GloFAS forecasts and reanalysis.

iii) It would be very interesting to compare any other fluvial flood models run in-country to the GloFAS forecasts if they have available forecast data for these cyclone cases. Local regional flood models, if available, may run at a significantly higher resolution than the 0.1 degree global resolution of GloFAS, and so may be able to better model flooding at the local level. The advantage of GloFAS is that their flood forecasts are openly available to humanitarian agencies and disaster risk reduction agencies across the world, but a comparison of the forecast skill against higher resolution local models would be very beneficial. If the Flood Hazard Impact Model – India (FHIM-India) , developed by the project of the same name and led by the UK Centre for Ecology and Hydrology (UK CEH; project started on September 2019 and ended March 2021), could also be run for these events that would also be a very valuable comparison for both systems.

iv) As stated above, further work is planned to carry out more detailed verification of the precipitation forecasts from the ECMWF ENS forecasts that are the key driver of the flood forecasts in GloFAS. As a side-project from this main work, it would be very interesting to utilise the precipitation verification methods being developed to carry out a comparison of the ECMWF ENS precipitation forecasts with the precipitation forecasts from the MOGREPS-G ensemble, and from the NEPS ensemble forecast system run by NCMRWF.

© Crown copyright 2021, Met Office

Met Office

References

Alfieri, L., E. Zsoter, S. Harrigan, F. Aga Hirpa, C. Lavaysse, C. Prudhomme, and P. Salamon, 2019: Range-dependent thresholds for global flood early warning. Journal of Hydrology X, 4, 100034.

Emerton, R., and Coauthors, 2020: Emergency flood bulletins for Cyclones Idai and Kenneth: the use of global flood forecasts for international humanitarian preparedness and response. International Journal for Disaster Risk Reduction, 50, 101811

Harrigan, S., E. Zsoter, H. Cloke, P. Salamon, and C. Prudhomme, 2020a: Daily ensemble river discharge reforecasts and real-time forecasts from the operational Global Flood Awareness System, Hydrol. Earth Syst. Sci. Discuss. [preprint], https://doi.org/10.5194/hess-2020-532, in review.

Harrigan, S., and Coauthors, 2020b: GloFAS-ERA5 operational global river discharge reanalysis 1979–present, Earth Syst. Sci. Data, 12(3), 2043-2060.

Hirpa, F. A., P. Salamon, H. E. Beck, V. Lorini, L. Alfieri, E. Zsoter, and S. J. Dadson, 2018: Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data, J. Hydrol., 566: 595–606.

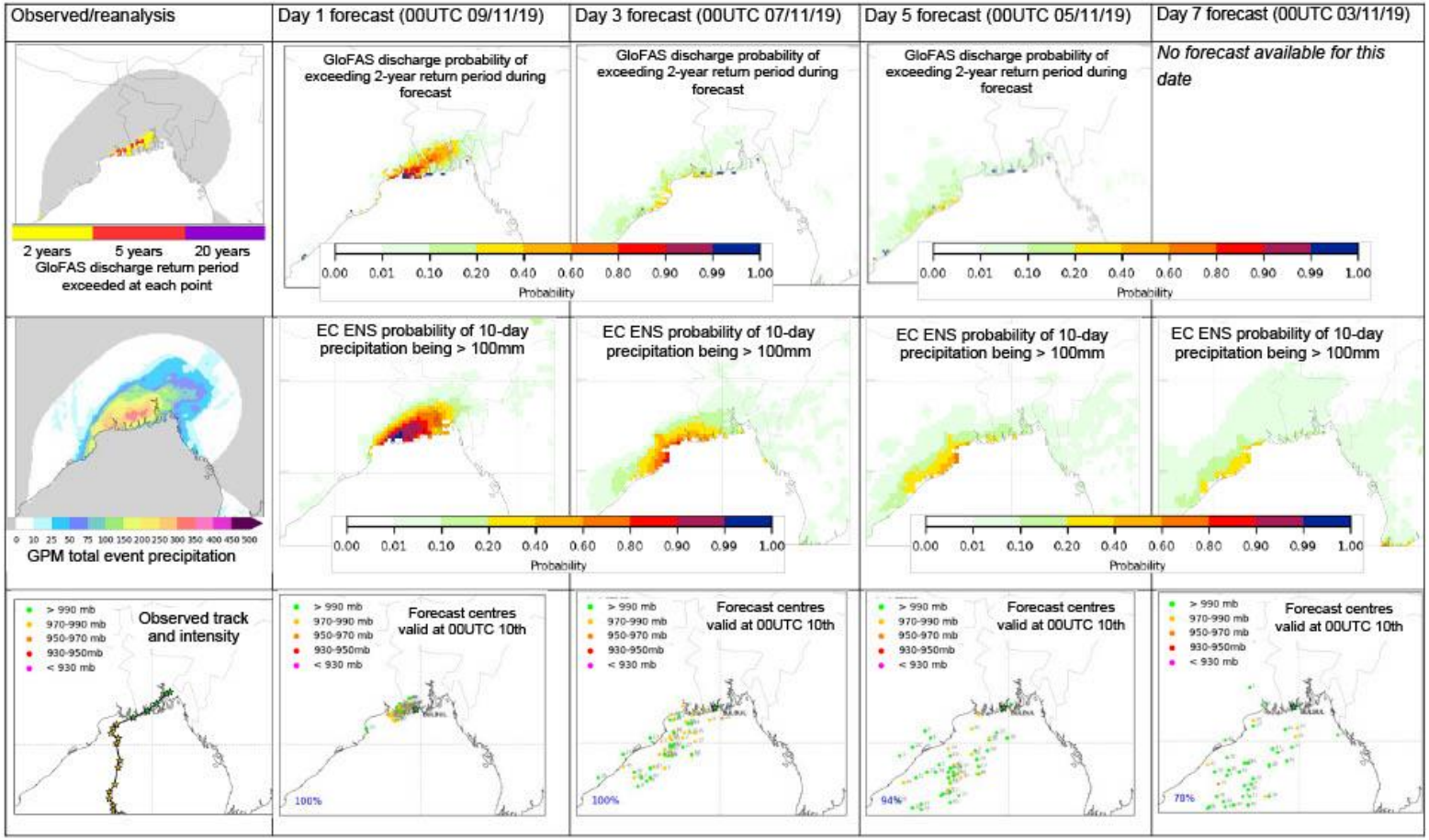
Titley, H. A., R. L. Bowyer, and H. L. Cloke, 2020: A global evaluation of multi-model ensemble tropical cyclone track probability forecasts. Q J R Meteorol Soc., 146: 531–545.

Titley, H. A., H. L. Cloke, S. Harrigan, F. Pappenberger, C. Prudhomme, J. C. Robbins, E. M. Stephens and E. Zsoter, 2021: Key factors influencing the severity of fluvial flood hazard from tropical cyclones. Journal of Hydrometeorology. Submitted.

Zsótér, E., H. L. Cloke, C. Prudhomme, S. Harrigan, P. de Rosnay, J. Munoz-Sabater, and E. Stephens, 2020: Trends in the GloFAS-ERA5 river discharge reanalysis. ECMWF Technical Memoranda, 871.

© Crown copyright 2021, Met Office

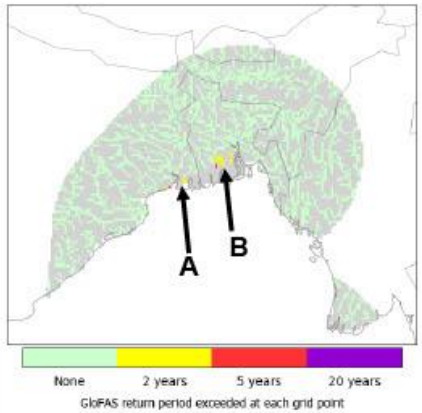
Appendix 1A: Cyclone Bulbul (6 Nov - 11 Nov 2019): GloFAS probability plots, EC ENS precipitation probability maps, and forecast centre plots.

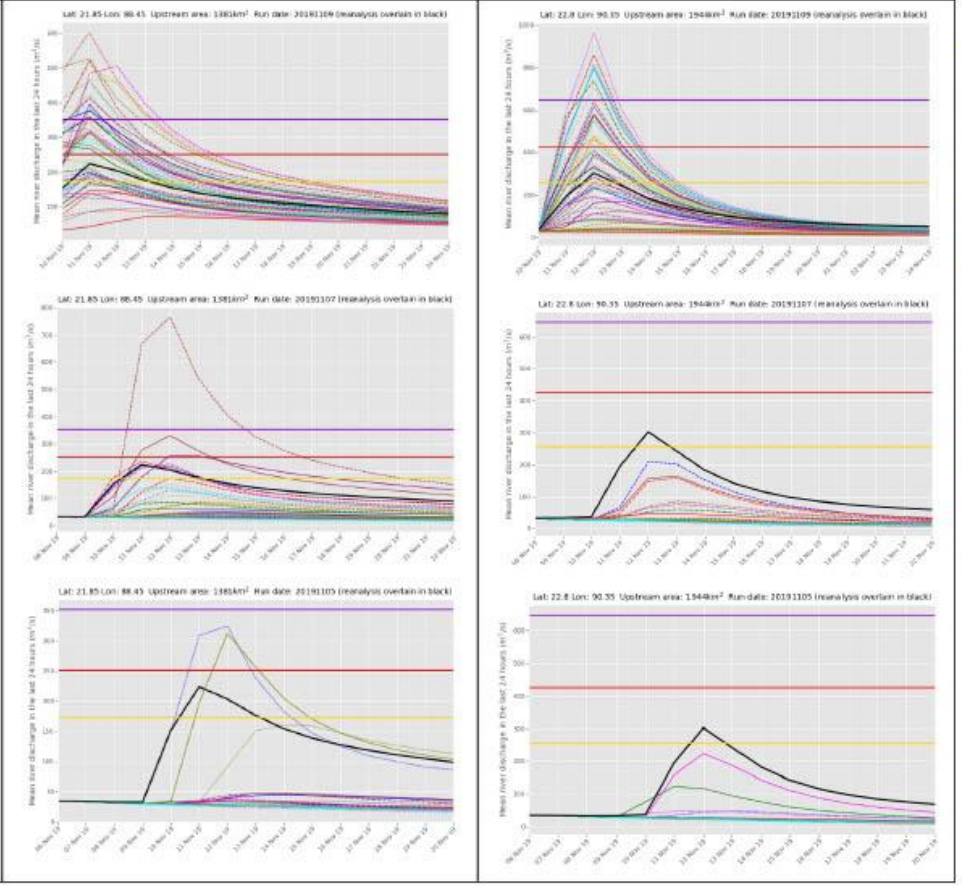


© Crown copyright 2021, Met Office

Appendix 1B: Cyclone Bulbul (6 Nov - 11 Nov 2019): Hydrographs from selected rivers.

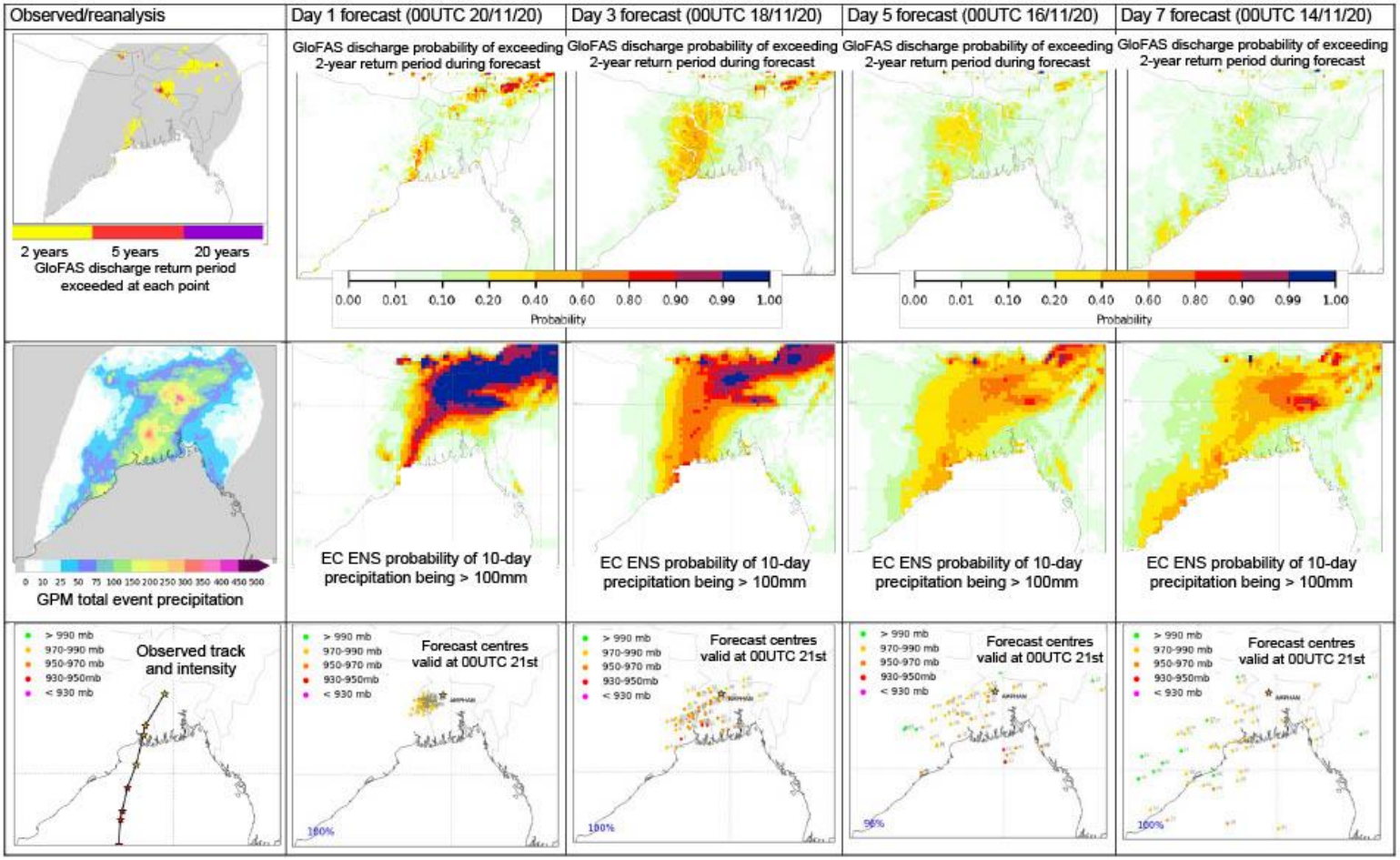
Observed river points (upstream > 500km^2) exceeding return period thresholds	Mani River, West Bengal **(A)**	Arial Khan River, Bangladesh **(B)**





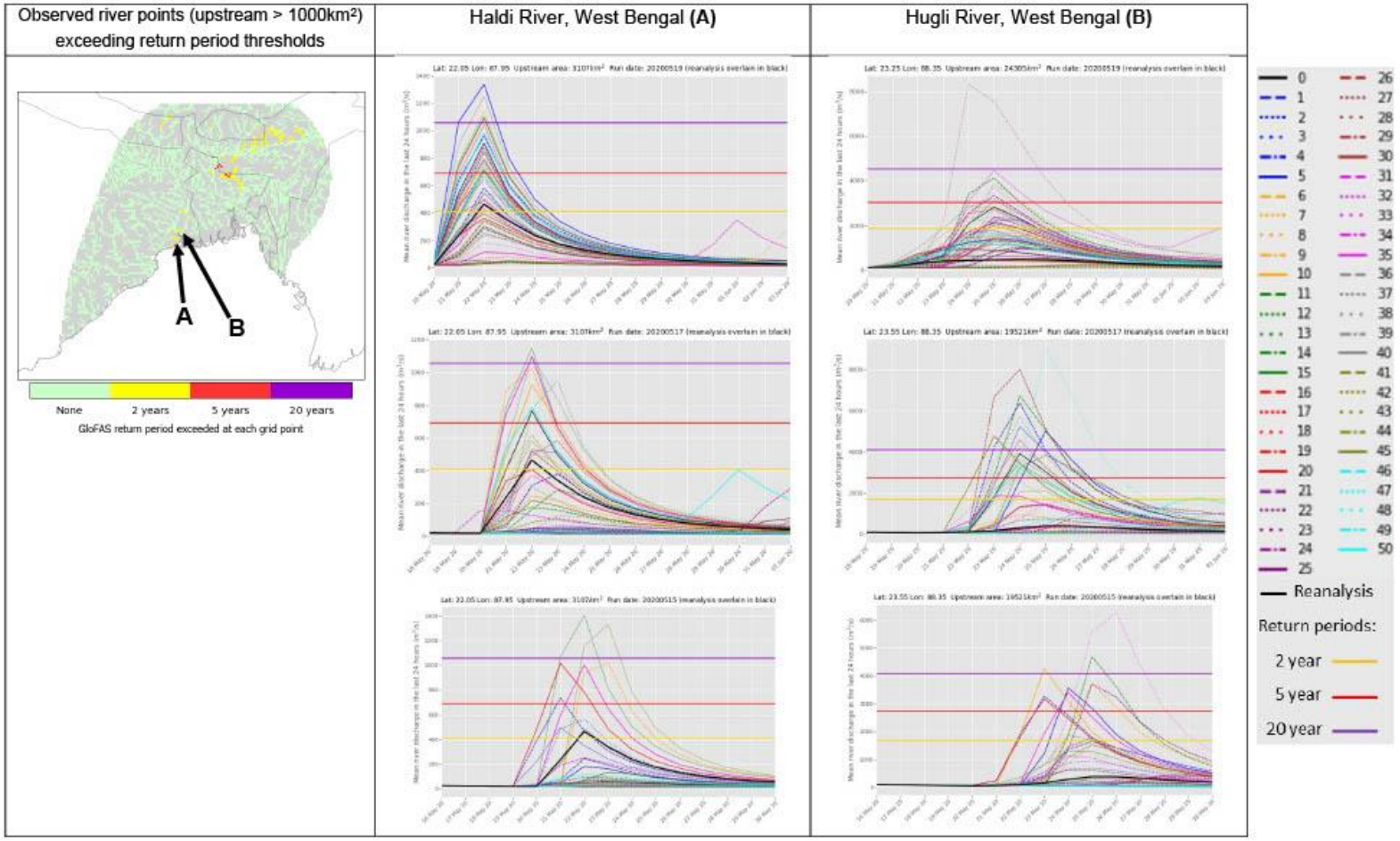
© Crown copyright 2021, Met Office

Appendix 2A: Cyclone Amphan (16 May – 21 May 2020): GloFAS probability plots, EC ENS precipitation probability maps, and forecast centre plots.



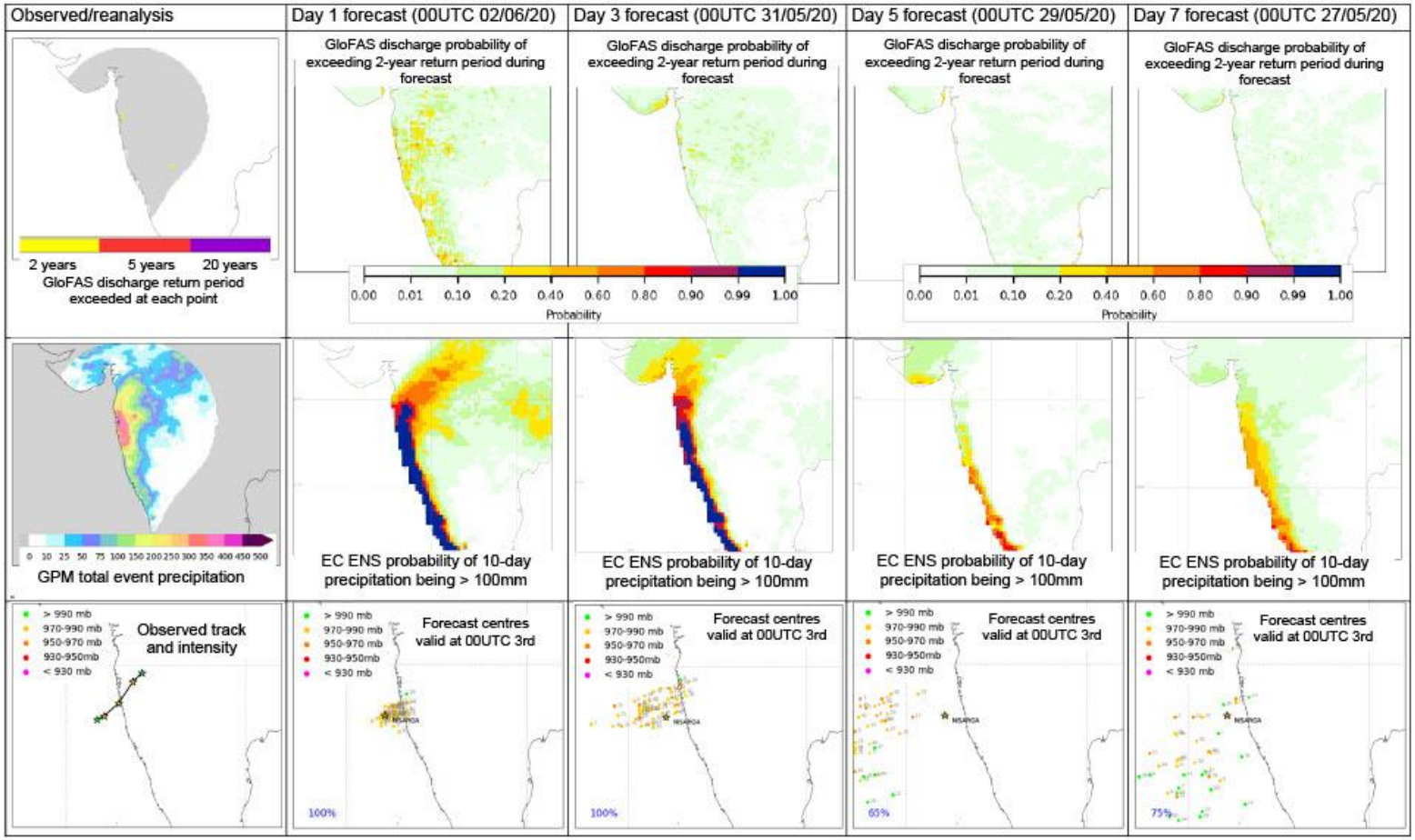
© Crown copyright 2021, Met Office

Appendix 2B: Cyclone Amphan (16 May – 21 May 2020): Hydrographs from selected rivers.



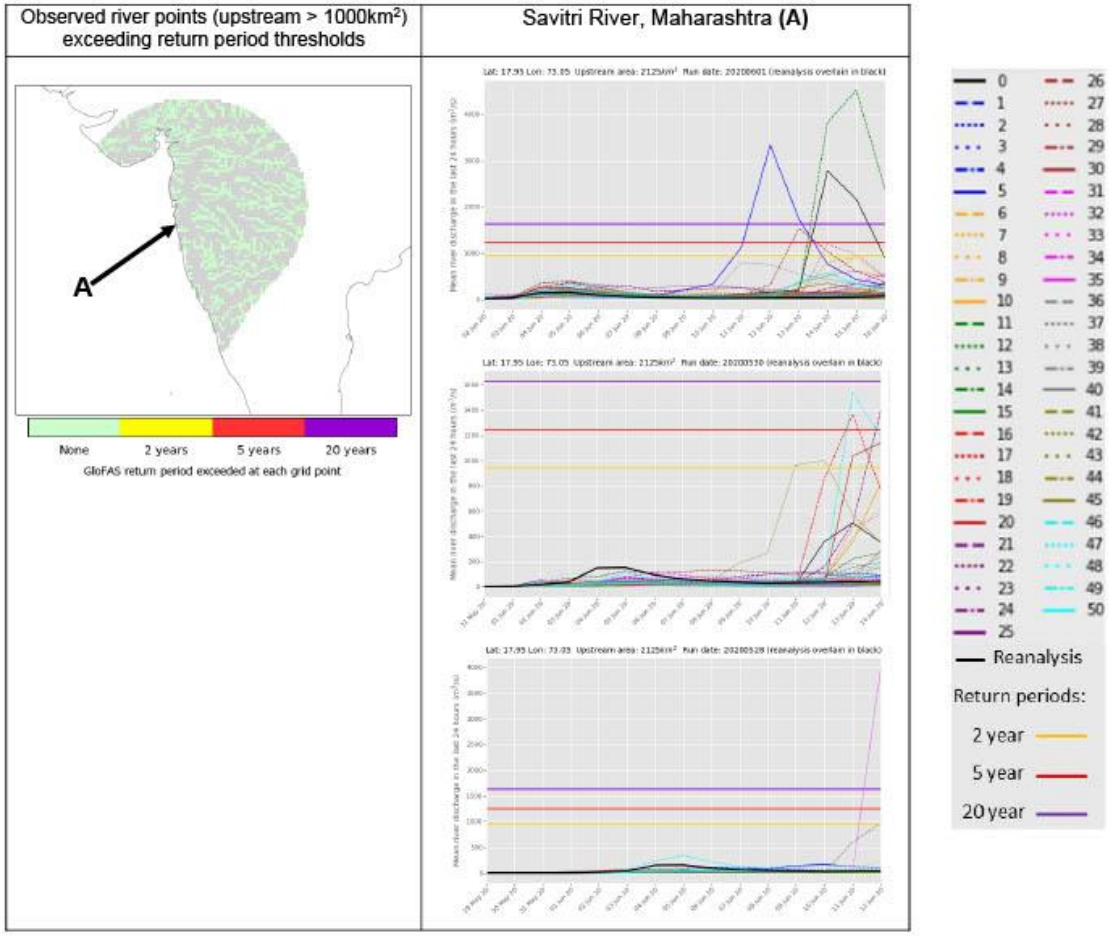
© Crown copyright 2021, Met Office

Appendix 3A: Cyclone Nisarga (1 June – 4 June 2020): GloFAS probability plots, EC ENS precipitation probability maps, and forecast centre plots.



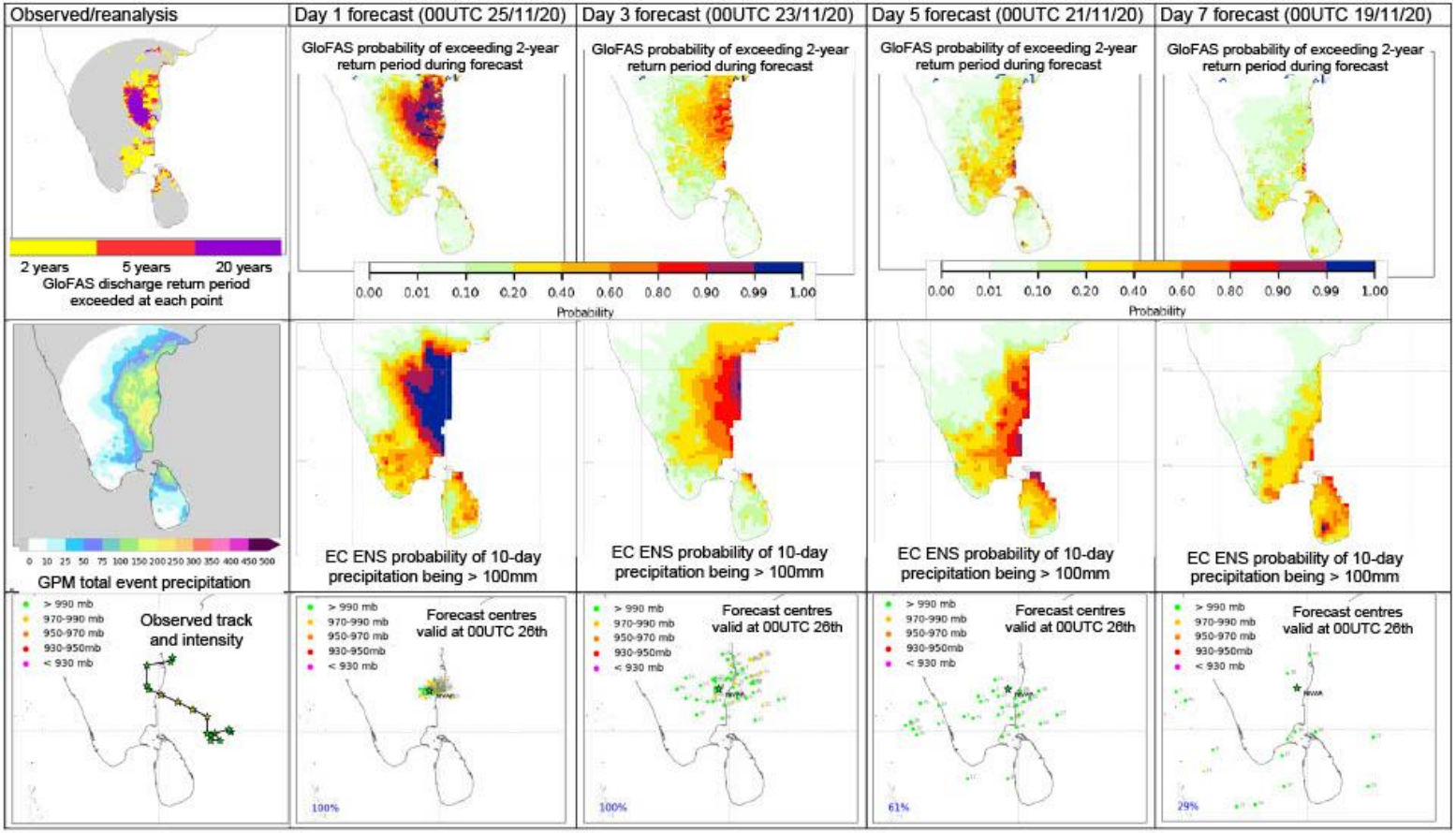
© Crown copyright 2021, Met Office

Appendix 3B: Cyclone Nisarga (1 June – 4 June 2020): Hydrographs from selected rivers.

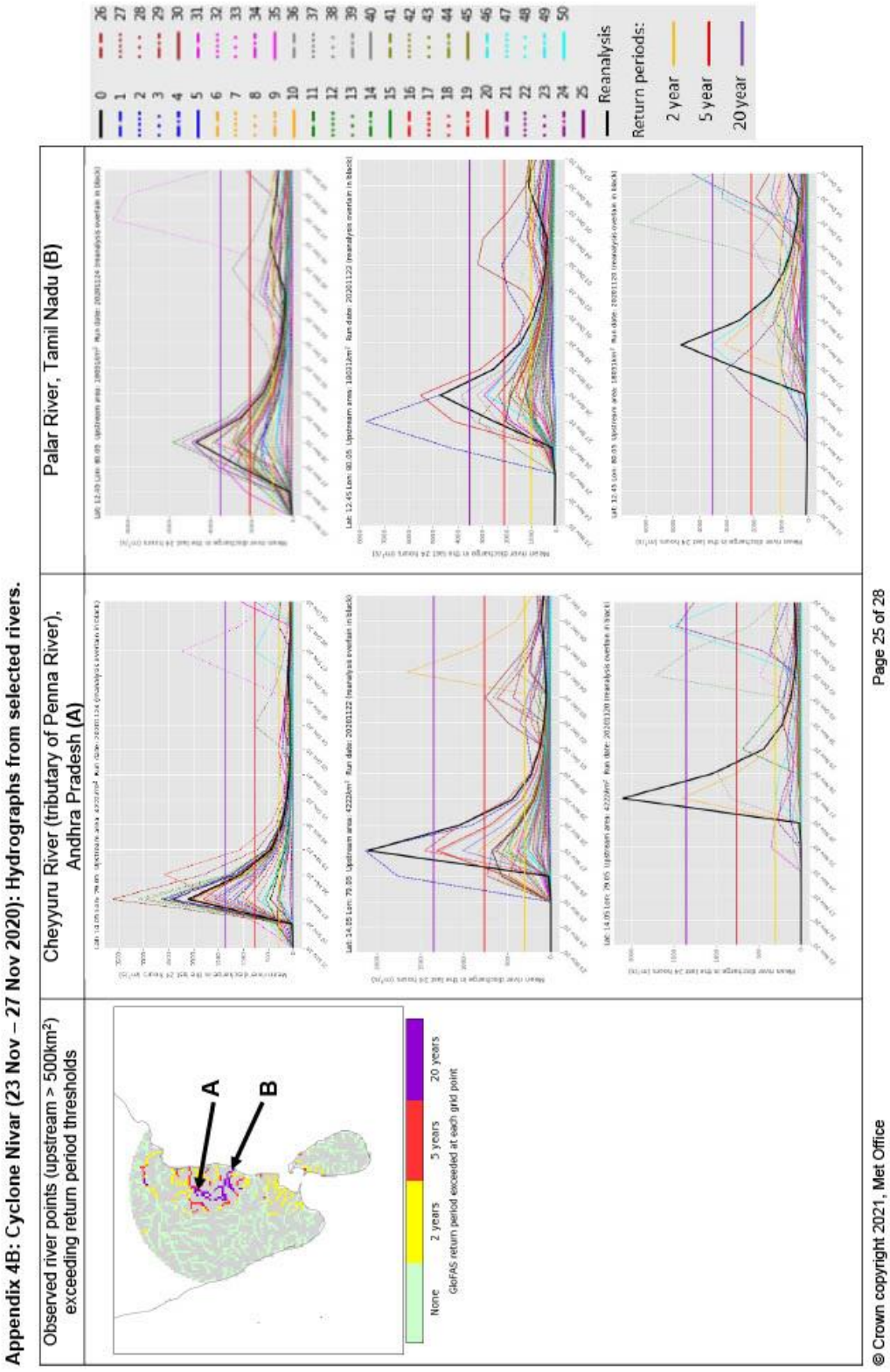


© Crown copyright 2021, Met Office

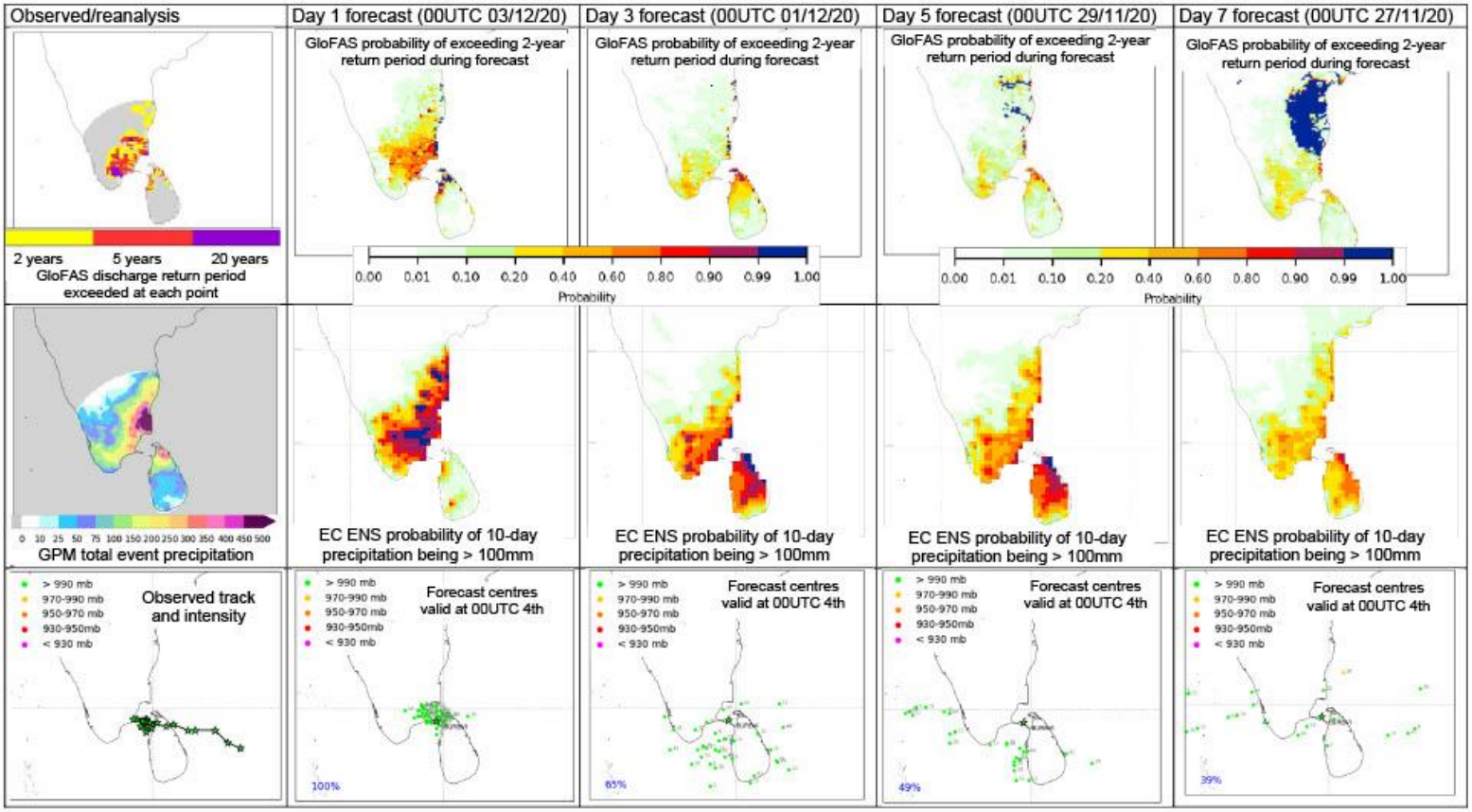
Appendix 4A: Cyclone Nivar (23 Nov – 27 Nov 2020): GloFAS probability plots, EC ENS precipitation probability maps, and forecast centre plots.



© Crown copyright 2021, Met Office

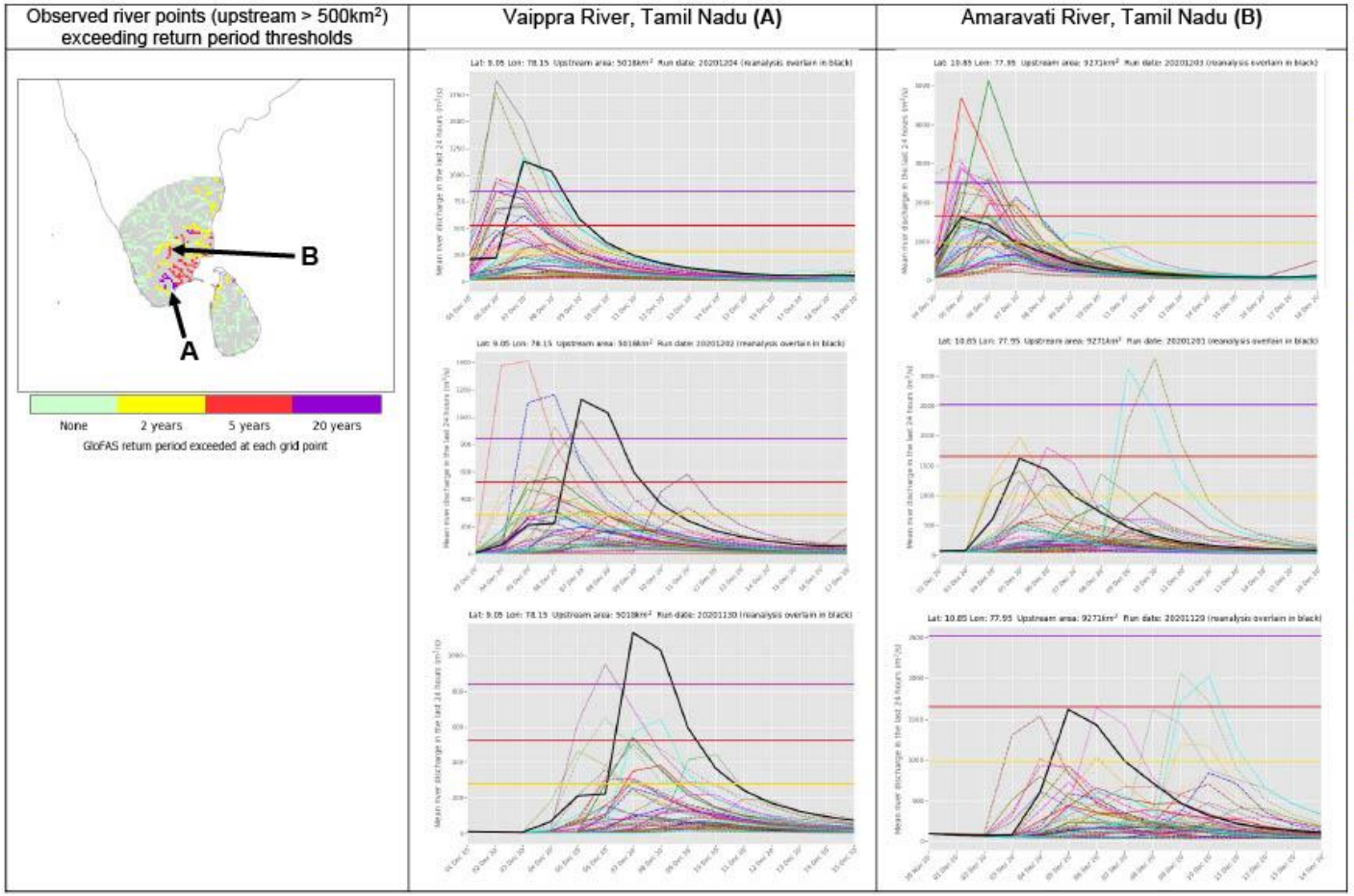

Appendix 4B: Cyclone Nivar (23 Nov – 27 Nov 2020): Hydrographs from selected rivers.
Observed river points (upstream > 500km²) exceeding return period thresholds
Cheyyuru River (tributary of Penna River), Andhra Pradesh (A)
Palar River, Tamil Nadu (B)
A
B
None
2 years
5 years
20 years
GloFAS return period exceeded at each grid point
Reanalysis
Return periods:
2 year
5 year
20 year
© Crown copyright 2021, Met Office
Page 25 of 28

Appendix 5A: Cyclone Burevi (30 Nov – 5 Dec 2020): GloFAS probability plots, EC ENS precipitation probability maps, and forecast centre plots.



© Crown copyright 2021, Met Office

Appendix 5B: Cyclone Burevi (30 Nov – 5 Dec 2020): Hydrographs from selected rivers.



© Crown copyright 2021, Met Office

A7: WCSSP India report: Exploring the use of feature-based verification techniques on ensemble forecasts of tropical cyclone rainfall

This appendix contains the full WCSSP India report summarised in Section 7.5.2, with the following reference:

Titley, H.A., North, R., Ashrit, R. and Dube, A. (2022). M4.1.2: Exploring the use of feature-based verification techniques on ensemble forecasts of tropical cyclone rainfall (WCSSP India project report).

Met Office

M4.1.2: Exploring the use of feature-based verification techniques on ensemble forecasts of tropical cyclone rainfall

Helen Titley and Rachel North (Met Office)

Raghavendra Ashrit and Anumeha Dube (NCMRWF)

www.metoffice.gov.uk

© Crown Copyright 2022, Met Office

Contents

1. Introduction 3
2. Data 4
2.1. Cyclone cases 4
2.2. Forecast data 5
2.3. Observation data 5
3. Verification methods and code 6
3.1. SAL 6
3.2. MODE 7
3.3. CRA 8
3.4. Additional ensemble verification metrics 9
4. Results 10
4.1. SAL 10
4.1.1. Cyclone Nivar 10
4.1.2. Cyclone Burevi 12
4.1.3. Cyclone Tauktae 12
4.1.4. Cyclone Gulab 13
4.1.5. Summary 14
4.2. MODE 16
4.2.1. Cyclone Nivar 17
4.2.2. Cyclone Burevi 20
4.2.3. Cyclone Tauktae 23
4.2.4. Cyclone Gulab 24
4.2.5. Summary 26
4.3. CRA using 'excRa' 27
4.3.1. Cyclone Nivar 27
4.3.2. Cyclone Burevi 29
4.3.3. Cyclone Tauktae 31
4.3.4. Cyclone Gulab 32
4.3.5. Summary 34
5. Summary, discussion, and recommendations for future work 35
References 38
Appendix 1: Development of 'cRa and excRa' 40

1. Introduction

The focus of investigation into tropical cyclone (TC) predictive skill has traditionally been on track and intensity forecasts, but there is increasing awareness of the importance of correctly forecasting hazards and impacts in order to narrow the gap between science and decision-making for early action based on forecast warnings and guidance. When forecasting hazards such as flooding due to heavy rainfall, it is imperative that the information on forecast uncertainty provided by ensemble prediction systems is pulled through into forecasts for those hazards, but there have been very few studies that investigate the ability of global ensemble forecast models to provide useful forecasts of the rainfall associated with landfalling TCs.

Traditional grid-to-grid precipitation verification methods generally offer little diagnostic information or reasoning on the performance of the forecast model that could be fed back into the model development process. Feature-based verification methods identify and isolate precipitation features in forecast and observation fields, and then verify attributes of these object pairs such as their intensity, area, shape, structure, and centroid location. These provide more relevant and useful information on model performance that can guide model developers on where to focus resource to improve those aspects of the precipitation forecasts that would lead to the most improvement in the forecasts.

There are two main aims of this study:

1) To identify what spatial and object-orientated verification techniques are available to help verify precipitation and how these can be adapted and applied to ensemble rainfall forecasts in TC cases.
2) To analyse the ability of global ensemble forecast models to predict rainfall in TC cases and investigate how this varies in terms of rainfall location, area and intensity.

In this study, three feature-based verification methods are selected to explore the use of such methods in the verification of ensemble tropical cyclone rainfall forecasts. These are the SAL (Structure, Amplitude and Location) method (Wernli et al. 2008; 2009), the Method for Object-Based Diagnostic Evaluation (MODE) (Davis et al. 2006; 2009) and the Contiguous Rain Area (CRA) method (Ebert and McBride, 2000; Ebert and Gallus, 2009). These object-based techniques were originally developed and applied to deterministic forecasts but their use in ensemble forecast verification is also gaining popularity (e.g. Gallus, 2010; Radanovics et al., 2018, Ji et al., 2020; Dube et al, 2022).

Object-based verification methods have not been extensively used in TC rainfall validation, aside from Marchok et al. (2007), who compared several deterministic NWP models for landfalling TCs in the US in terms of rainfall pattern, mean rain intensity and volume, and intensity distribution. Of the three selected methods in this paper, to the authors' knowledge only the CRA method has been applied to rainfall forecasts specifically from tropical cyclones (Chen et al., 2018; Yu et al., 2020), and only in deterministic forecast studies. Therefore, this study offers the first opportunity to explore the use of object-based methods on ensemble forecasts of tropical cyclone rainfall. Three ensembles are included in the study: the ECMWF Ensemble Prediction System (EC ENS), the Met Office global ensemble (MOGREPS-G), and the NCMRWF Global Ensemble Prediction System (NEPS-G). The rainfall forecasts from these three models are examined for four recent landfalling cyclones in India: Cyclone Nivar (November 2020); Cyclone Burevi (December 2020); Cyclone Tauktae (May 2021) and Cyclone Gulab (September 2021).

The remainder of this report is set out as follows: Section 2 gives an overview of the four cases being examined in this study, and the forecast and observational datasets that are being used. Section 3 introduces the verification methods and code for the three feature-based verification methods (SAL, MODE and CRA). Section 4 presents the results from each of these methods, and section 5 contains a discussion of the key findings and ideas for future work.

2. Data

2.1. Cyclone cases

This report covers four landfalling Indian cyclones in the 2020 and 2021 Indian cyclone season. The dates and tracks of these cyclones are shown in Figure 1, alongside the event total rainfall in each case. These four cyclones have been selected as they all led to impacts from heavy rain or flooding and cover the range of cyclone types experienced in India. Cyclones Nivar and Burevi occurred in the post-monsoon cyclone season, while Cyclone Tauktae was in the pre-monsoon period, and Cyclone Gulab in the monsoon season.

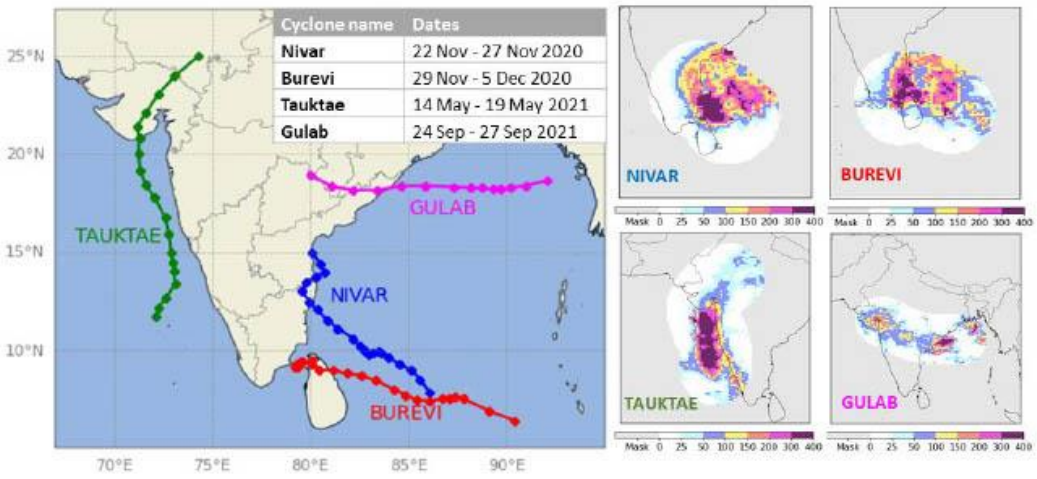


Figure 1: The dates and tracks of the four Indian cyclones included in this report (left), alongside the storm total rainfall, as calculated from GPM IMERG (final product).

Cyclone Nivar reached its peak intensity of 120km/hr on 25th November 2020, making it a Very Severe Cyclonic Storm. It weakened ahead of making landfall in Marakkanam, Tamil Nadu, close to Puducherry, then entered back into the Bay of Bengal as a Deep Depression, before making a second landfall on the Andhra Pradesh coast on 26th November. The cyclone brought very heavy rains over northern Tamil Nadu ahead of and after landfall. Many areas including Madipakkam, Velachery, Adambakkam and suburbs around Tambaram and low-lying regions along the river Adyar were flooded. Several areas were flooded in Puducherry, while in Andhra Pradesh, the rainfall led to significant impact in the districts of Chittoor, Prakasam, Kadapa and Nellore with 112,000 people affected. Five people were reported dead in Tamil Nadu, and eight in Andhra Pradesh.

Cyclone Burevi made landfall in Sri Lanka on December 2nd 2020, the first tropical cyclone to make landfall in Sri Lanka since 2014. After weakening over Sri Lanka, Burevi exited into the Gulf of Mannar early on December 3rd. Burevi stalled just west of Sri Lanka as it became stuck in a col between two subtropical ridges before eventually dissipating on December 5th. In Sri Lanka, more than 75,000 people were evacuated ahead of the storm, which damaged around 3,000 houses. In southern India, prolonged and heavy rainfall led to flooding in Tamil Nadu and Puducherry. Burevi caused at least 11 fatalities.

Cyclone Tauktae was the strongest tropical cyclone to make landfall in the Indian state of Gujarat since the 1998 Gujarat cyclone. Tauktae originated from an area of low pressure in the Arabian Sea on May 14th, then took a northward turn and strengthened while travelling parallel the coast of the Indian states of Kerala, Karnataka, Goa and Maharashtra, becoming an extremely severe cyclonic storm by May 17th. After making landfall in Gujarat, Tauktae gradually weakened as it turned north-eastward inland. Tauktae brought heavy rainfall and flash floods to several states and resulted in at least 169 deaths in India.

Cyclone Gulab originated from a low-pressure area over the Bay of Bengal on September 24th and strengthened into a cyclonic storm the next day. Gulab made landfall in Andhra Pradesh on September 26th and weakened over land into a remnant low. It continued moving westward, emerging into the Arabian Sea on September 29th, where it re-strengthened into Cyclone Shaheen and impacted Oman, but in this report we restrict analysis to the rainfall over India. At least 20 deaths have been attributed to Gulab, which brought impacts from heavy rain in several states including Andhra Pradesh, Odisha and Telangana.

2.2. Forecast data

Three global ensemble forecast models that are included in this study: MOGREPS-G, NEPS-G and ECMWF ENS (see Table 1 for further details).

Model	Number of members	Resolution	Forecast length
MOGREPS-G	36 (18+18 time-lagged)	20 km	8 days
NEPS-G	23 (11+control + 11 time-lagged)	12 km	10 days
ECMWF ENS	51	18 km	15 days

Table 1: The details of the global ensemble forecast models included in this study.

2.3. Observation data

Dataset	Formulation
Integrated Multi-satellitE Retrievals for GPM (IMERG) (final run)	Available ~3 months after valid time. Can include additional microwave passes. Includes a monthly gauge correction. Native resolution ~10km, V06B. For further details on the algorithm see https://gpm.nasa.gov/sites/default/files/2020-10/IMERG_doc_201006.pdf
Integrated Multi-satellitE Retrievals for GPM (IMERG) (late run)	Available ~12 hours after valid time. Native resolution ~10km, V06B. For further details on the algorithm see https://gpm.nasa.gov/sites/default/files/2020-10/IMERG_doc_201006.pdf
IMD merged observations	Available one day after valid time. The IMD–NCMRWF merged satellite gauge (NMSG) data is available at 0.25 x 0.25 grid resolution. For additional details on the data merging technique and data quality see Mitra *et al.*, (2009; 2013)

Table 2: The details of the three observational datasets being using in this study

Two satellite-derived products (created for 24-hour periods ending at both 00UTC and 03UTC) and a merged satellite/gauge product (created for 24-hour periods ending at 03UTC) were made available as potential verification truth sources (see Table 2 for further information).

An example 24-hour accumulation period for Cyclone Nivar can be seen in Figure 2. The left-most panel contains the GPM IMERG late product. The middle panel contains the GPM IMERG final product and the right-hand panel contains the merged gauge product produced by NCMRWF and IMD. All three products are shown at their native resolutions. The most obvious differences between the leftmost two panels are the addition of a region of precipitation over Northern India and a lowering of the precipitation totals from the rainfall associated with Nivar; a widespread band of values in the range 65-115 mm day^{-1} is diluted somewhat in the GPM final product.

5

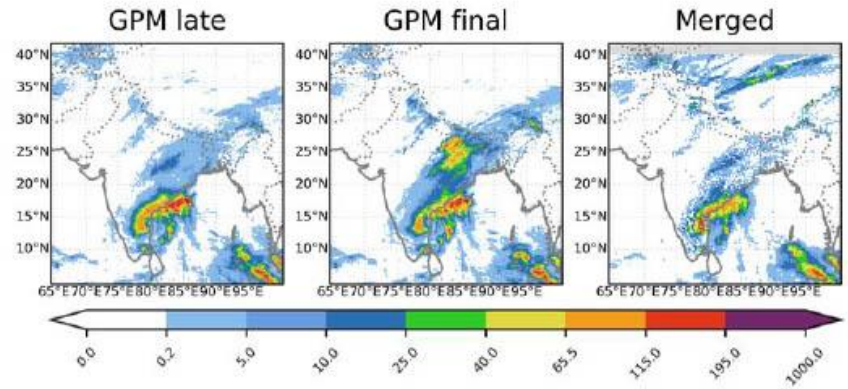


Figure 2. Example 24-hour accumulation ending at 3UTC on 27th November 2020, showing Cyclone Nivar, from a) GPM IMERG late V06B; b) GPM IMERG final V06B and c) IMD-NCMRWF merged satellite/gauge product.

We have used the GPM IMERG final product for verification in the SAL/MODE results. The GPM team recommend using the GPM IMERG final product for research applications, presumably due to the extra microwave overpass data which is incorporated and the monthly gauge data adjustment which is applied. We also wanted to compare the three models over the longest possible lead times. The IMD merged observation dataset covers the 03-03UTC period and the products which are archived from the ECMWF ensemble don't allow accumulations to be calculated over this period beyond day 3. The CRA results use the IMD merged observations as the truth source, since only the NEPS-G results are verified and therefore all lead times can be accommodated.

3. Verification methods and code

3.1. SAL

The SAL method (Wernli et al. 2008; 2009) is an object-based quality measure which quantifies the quality of a forecast in terms of 3 parameters: Structure (S), Amplitude (A) and Location (L), for gridded forecasts compared to gridded observations. The practical application of SAL involves three steps, and how each of these have been applied in this study are detailed below:

i) Specification of a domain

In this study the domain varies for each cyclone and each accumulation period, and is based on the tropical cyclone track. The gridded observation and forecast fields are masked so that the domain consists of those points within 500km of the observed tropical cyclone track data between the start and end of the accumulation period as taken from the International Best Track Archive for Climate Stewardship (IBTrACS) (Knapp et al. 2010, 2018). This represents the area of precipitation defined as being associated with the cyclone, as used in multiple studies that assign a set radius when defining TC-related rainfall (Prat and Nelson, 2013, 2016; Jiang et al., 2011; Luitel et al., 2018; Titley et al., 2021).

ii) The choice of a threshold to define rainfall objects

A rainfall object is defined when it exceeds statistically defined threshold value. Wernli et al. (2009) suggest that this be done using the formula for the threshold (R*) as follows: $R^* = \frac{1}{x} * R^{95}$ whereby the threshold is a fraction of the 95th percentile of all grid-point values in the domain which are larger than 0.1mm. They used a value of 15 for '*x*' in their study, but the code gives flexibility for this to be adjusted as required. After a sensitivity analysis (not shown) it was decided that a value of 3 was more suitable for this study, as this tended to give 24-hour

totals roughly equivalent to the heavy rainfall threshold defined by IMD.

iii) The calculations of the three SAL components

Once the rainfall objects have been identified in the gridded verification domain for the observed and forecast fields, the identified objects are then compared across the forecast and observed fields and the S, A, and L components calculated. The calculation methods for each attribute are detailed in Wernli et al. (2008; 2009), but the summary of what each attribute represents and its range of values is given in Table 3. A perfect forecast is characterized by zero values for all three SAL components.

Attribute	Description of calculation	Range of attribute scores
Structure (S)	- Compares the volumes of the rainfall objects to provide information about their size and shape	Range of S is [-2, 2] Positive values indicate that modelled objects are too large/flat, and negative values indicate that objects are too sharp/small.
Amplitude (A)	- Represents a normalized difference between the domain-averaged forecast and observation fields - Is independent of the identification of features as it depends on the total rainfall amount	Range of A is [-2, 2] Positive values indicate overestimation of total rainfall, while negative values indicate underestimation
Location (L)	- L component combines information about the distance of predicted and forecast mass centres and the error of a weighted average distance between the rainfall objects - L1 measures the normalized distance between the mass centres of the forecast and observation fields - L2 is the difference of normalized distance between the mass centres of the total rainfall fields and the individual rainfall objects over observed and forecast fields	Range of L1 is [0, 1] Range of L2 is [0, 1] L=L1+L2; therefore range of L is [0, 2]

Table 3: The details of the three SAL components (Structure (S), Amplitude (A) and Location (L))

The SAL code is available in Fortran at the Met Office and was run for each ensemble member and observation pair. The SAL components were then compared in box plots to look at the range of scores across each ensemble for the Structure, Amplitude and Location attributes at a range of lead times and accumulation periods for each of our cyclone cases. These results are summarised in section 4.1.

3.2. MODE

MODE is an object-based method designed to identify objects in gridded forecasts and compare them to those from a gridded truth, in much the same way as a meteorologist might interpret them. It was initially demonstrated with an example application from precipitation verification (Davis et al., 2006), but in practice could be used on any field for which a threshold can be used to identify an object. There are essentially four steps in the process:

i) *Input data fields are smoothed using a convolution radius.*

This replaces the value at a grid point with an average over the area within a disk whose centroid is located there. The convolution radius should be big enough to account for the fact that the realistic scales captured by a forecast are at least eight grid lengths long (Skamarock, 2004).

ii) *A binary mask indicating the extent of the object is created by thresholding the data resulting from step i).*

The choice of threshold is important, and a suitable convolution threshold should be supplied to assist the algorithm in detection – a lower merge threshold is also supplied to facilitate detecting objects which would have

been identified as a single object had the lower threshold been used. The combination of steps i) and ii) allows objects to be defined which are consistent with how a human would draw object outlines on the field and fills in gaps or holes in those objects.

iii) *Original field values are put back into the objects identified using the combination of the convolution and the binary mask.*

This ensures that variability in the intensities within the objects are captured and can be analysed as a distribution. The object outline identification process in step ii) ensures that it is not only grid points which exceed the requested threshold which are included within the objects.

iv) *Objects identified in the field(s) are then matched and/or merged depending on the user-supplied configuration.*

Merging of objects is done within a field, *e.g.*, the forecast, to ensure that individual objects can be associated as 'cluster' objects; those which would be identified as a single object had a slightly lower threshold been defined. Matching of objects is done between the forecast and truth fields, *i.e.*, in separate fields.

These steps are pictured in Figure 2 of Davis et al. (2006a). The attributes of the resulting objects (*e.g.*, area, intensity within object, angle of orientation) can then be analysed as a whole, either individually for the forecasts and observations, or as matched pairs between forecast and observation. Attributes for the objects in the forecasts or observations measure geometrical properties, *e.g.*, the area bounded by the objects, the aspect ratio of the objects, the centroid location. Also, properties of the intensity distributions contained with the objects are calculated. Attributes for the matched-pair objects look at the differences between the matched objects in the forecast and the observation. For example, centroid difference (the distance between the centroid locations of the matched object in the forecast field compared to the observation field), angle difference, difference in aspect ratios, area which overlaps in both objects, etc. Full details of all the outputs which can be produced using MODE can be found in the relevant section on the documentation webpage (https://met.readthedocs.io/en/latest/Users_Guide/mode.html).

In the experiments carried out here, the decision was taken to re-grid all the fields to the coarsest resolution model grid (MOGREPS-G, ~20km) before object identification. This was to facilitate a consistent comparison across models since differences in the results due to resolution differences are minimised. In any case, the higher resolution models should be able to better capture the processes involved; some of this benefit should be retained, and hopefully measurable, even if the forecasts are upscaled for verification.

METplus wrappers are a python infrastructure for running MET tools. Here, METplus configuration files were set-up and used to run MODE for the three ensemble forecast systems using the GPM IMERG final dataset for verifying. These were initiated using a shell script. Results from daily and five-day rainfall accumulations (starting and ending at 00UTC) are presented here.

See https://metplus.readthedocs.io/en/main_v4.0/Users_Guide/overview.html and sections of the user guide within for further details.

3.3. CRA

The CRA technique is an object/feature-based method that verifies the properties of spatial forecasts of observed and forecasted entities. This method was developed to evaluate systematic errors in the prediction of rain systems (Ebert and McBride 2000). The CRA method measures the errors in predicted rain systems and can be used to separate this total error into components due to incorrect location, amplitude, and differences in fine scale patterns. The CRA technique is described in detail in Ebert and McBride (2000), and Ebert and Gallus (2009), and only a brief overview is provided here.

In the CRA verification, an entity finder is applied to isolate distinct CRAs in the merged field of observations

and forecasts according to a minimum intensity threshold. An entity/object is anything that can be defined by a closed contour. Some examples of entities, or blobs, are contiguous rain areas, convective outlook regions, and low-pressure minima. In the case of rainfall; firstly, for every observed entity a rectangular box is fitted to the CRA which is then expanded till a certain distance (5º latitude-longitude or the maximum dimension of the CRA (whichever is smaller)) to look for a matching forecast entity (Ebert and Gallus, 2009). Secondly, in order to create match between CRAs, the forecast entities are translated horizontally over the observed entities and the match is created based on a best-fit criterion. Many studies have been conducted to ascertain the best-fit criterion e.g. the minimum squared error was analysed by Ebert and McBride (2000), and the maximum overlap between the entities was studied by Ebert *et al*., (2004). The method of minimum squared error and maximum correlation analyzed by Grams *et al*., (2006) is considered the most optimal criterion for pattern matching between forecasted and observed CRAs (Tartaglione *et al*., 2005). In addition to these criteria, to find the best fit the pattern correlation between the forecasted and observed entities has to be positive and significant at a 95% confidence interval (to check the reliability of the fit). Another check is the retaining of at least 75% of the forecast grids along with the forecast maximum while matching the patterns between the two objects. For each entity that can be identified in the forecast and the observations, CRA verification uses pattern matching methods to determine the location error, as well as errors in area, mean and maximum intensity, along with the fine-scale spatial pattern.

The original CRA code is written in IDL, which is not freely available to all users and therefore limits the uptake of the method. For this study, the development of CRA software using R (cRa) and its extension for spatial verification of ensemble model forecasts (excRa) has been undertaken at NCMRWF. The new R code is named 'cRa' since it is fully developed based on the original ideas implemented in IDL. An extension of the tool for ensemble forecasts 'excRa' has also been developed which allows CRA verification to be applied to forecasts from ensemble members. For further details of the newly developed cRa and excRa R code see Appendix 1.

In this study we apply the CRA method using the new cRa and excRa methods, to forecasts of 24-houro rainfall (03-03UTC accumulations) from NEPS-G for the four cyclones listed in Section 2.1, following the approach of Gallus (2010) who first used the CRA verification method for ensemble rainfall forecasts. The IMD merged observations are used as the verifying dataset.

3.4. Additional ensemble verification metrics

Assessment of RMSE and spread among the rainfall forecast attributes output by the various object-based methods helps to quantify the uncertainty in terms of object parameters. Spread among the ensemble members in, say, highest rainfall amount, or the spatial extent of the objects, reflects the spatial uncertainty in the forecasts. For any EPS to be perfectly calibrated and reliable the joint distribution of the ensemble members should not change when the observations are substituted for any one of the members (Wilks, 2011). In other words, the observations should be indistinguishable from the ensemble members and therefore the RMSE in the mean should match the average spread of the ensemble members (Fortin *et al*., 2014). Under dispersive (*over dispersive*) EPSs are overconfident (*underconfident*) which leads to underestimation (ov*erestimation*) of the probability of extreme events thereby leading to misses (*false alarms*) in the forecast. It is common to compare the skill given by the RMSE of the ensemble mean to the average ensemble spread, which is calculated as the square root of the average ensemble variance following Fortin *et al*., (2014) (*spread-skill relationship*) which helps in analyzing if the EPS has an optimal spread and can represent all sources of uncertainty. For a perfectly reliable ensemble system, the ensemble spread should, on average match the RMSE in the ensemble mean, at all lead times (Leutbecher, 2010). In this study we have calculated RMSE and spread in our MODE and CRA results (see sections 4.2 and 4.3).

The Continuous Ranked Probability Score (CRPS) is a recommended measure of performance of a probabilistic forecast. It is a quadratic measure of difference between the forecast cumulative distribution function (CDF) and

the empirical CDF of the observations (Zamo et al., 2018 and Hersbach, 2000). The CRPS generalizes the mean absolute error; in fact, it reduces to the mean absolute error (MAE) if the forecast is deterministic. For further information see https://www.cawcr.gov.au/projects/verification/#Methods_for_EPS. The CRPS is computed for the CRA attributes to summarise their probabilistic performance (see section 4.3).

4. Results

4.1. SAL

4.1.1. Cyclone Nivar

Box plots summarising the ensemble forecasts rainfall forecasts for Cyclone Nivar are shown in Figure 3, using GPM final as the comparison observation data. The top row shows the S, A and L components for all 24-hour rainfall forecasts during Cyclone Nivar. The structure component (3a) shows significant spread in the structure of the rainfall objects, which is similar across lead times. For MOGREPS-G and NEPS-G these are more evenly spread around the ideal zero value, whereas most ECMWF ENS members show structure scores greater than one, indicating that the rainfall objects are too large and flat. The Amplitude plot (3b) reveals a general underestimation of rainfall amounts at all but the shortest and longest lead times. Location errors (3c) gradually reduce as lead time shortens, as the confidence in the track forecasts improves. The second row shows the SAL components for the storm total rainfall (120-hour rainfall totals from 00UTC 23rd to 00UTC 28th November 2020). Most of the object structures in the members are too large/flat in all models and are generally worse in the ECMWF ENS (3d). Rainfall amounts are generally underestimated, although this is less pronounced at shorter lead times (3e). Again, location errors gradually reduce as lead time shortens (3f). However, the results change when you apply a land sea mask to this same storm total field in order to focus on the rainfall over land that is so crucial to flood impacts (bottom row). Now the ECMWF ensemble has more optimum object structures than the MOGREPS-G and NEPS-G forecasts, whose rainfall objects are shown to be too sharp and small (3g). The Amplitude component reveals an underestimation of rainfall amounts from MOGREPS-G and NEPS-G. EC ENS is better overall, but with slight overestimations at shorter lead times (3h). Location errors gradually reduce in NEPS-G and EC ENS, but several members maintain larger location errors in MOGREPS-G. Similar plots were produced (not shown) using GPM late and the IMD merged observations in order to examine the impact of observation uncertainty. Although there is some variation in the results over land, where sharper observation objects lead to an increase in S scores, the overall findings are similar.

SAL diagrams (Wernli et al., 2008) for the total storm rainfall forecasts (without land sea masks) from the three ensembles at two different lead times are shown in Figure 4. Whereas in the box plots you cannot examine the relationship between the parameters as their distributions are plotted separately, here the relationship between the S, A and L components can be analysed, with S on the X axis, A on the Y axis, and points shaded according to L. Excellent forecasts (small values of all three components) are found as pale dots in the centre of the diagram. Dashed lines indicate the median values of S and A. In the first quadrant (top right), forecasts overestimate both the amplitude and the structure components of SAL. In the third quadrant (bottom left), both components are underestimated. The high density of ECMWF ENS members in the two right quadrants in both lead times indicates that the model typically has too large and/or flat rainfall objects. MOGREPS-G and NEPS-G are close to the ideal centre line for the S component. At the longer lead time most of the members from all models are in the bottom two quadrants where the Amplitude, or overall rainfall amount, is underestimated. It is notable that no forecasts are situated in the top left-hand quadrant, indicating that it is difficult to produce a strong overestimation of rainfall amplitude with too small objects. Although these diagrams are an interesting way to compare the relationships between the components, it is harder to summarise performance efficiently across lead times, and therefore these diagrams are not included for the other storms.

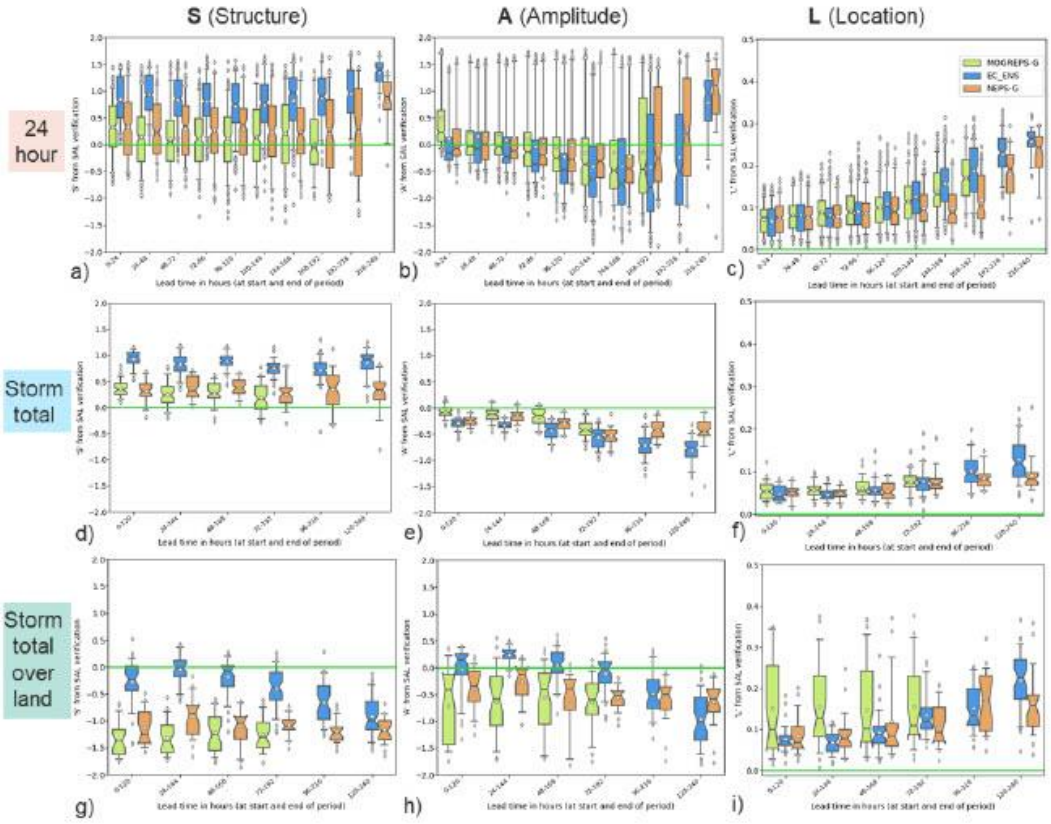


Figure 3: Box plots comparing the distribution of S, A, and L components of the SAL verification for rainfall in Cyclone Nivar for MOGREPS-G (green), ECMWF ENS (blue) and NEPS-G (orange) using GPM final data as the observation dataset: Top row (a-c) = combined for all 24-hour periods (from 00UTC 23/11/2020 to 00UTC 28/11/2020); Middle row (d-f) = storm total rainfall (defined as 00UTC 23/11/2020 to 00UTC 28/11/2020); Bottom row (g-i) = storm total over land only (also defined as 00UTC 23/11/2020 to 00UTC 28/11/2020, but with land/sea mask applied). All 00UTC forecast runs from 18/11/2020 to 27/11/2020 are included in the verification.

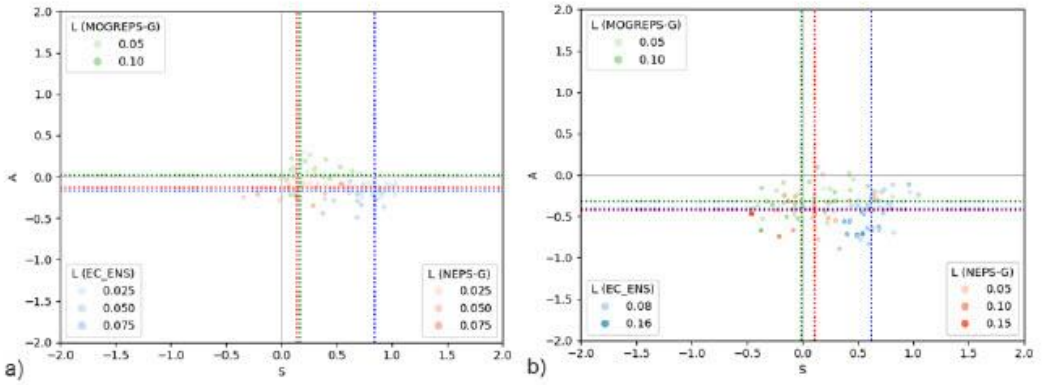


Figure 4: SAL plots plotting Structure (S) component vs Amplitude (A) component, shaded by Location (L) component for MOGREPS-G (green), ECMWF ENS (blue) and NEPS-G (red). Median values for each ensemble are shown by the dashed lines. These plots are for the storm total rainfall (defined as the 120-hour period 00UTC 23/11/2020 to 00UTC 28/11/2020), for lead times a) 0-120 hours; b) 72-192 hours.

4.1.2. Cyclone Burevi

Box plots summarising the ensemble forecasts rainfall forecasts for Cyclone Burevi are displayed in Figure 5. Many of the signals shown in the Cyclone Nivar case are again present. Most notably, the signal changes in the storm total rainfall verification once the land/sea mask is applied, with the structure of the objects being closer to the optimum zero value in MOGREPS-G and NEPS-G when rainfall over sea is included, but being better in the ECMWF ENS once the land/sea mask has been applied (5d and 5g). One difference between Nivar and Burevi is that the Amplitude attribute indicates generally better estimations of the storm total rainfall amounts in Burevi.

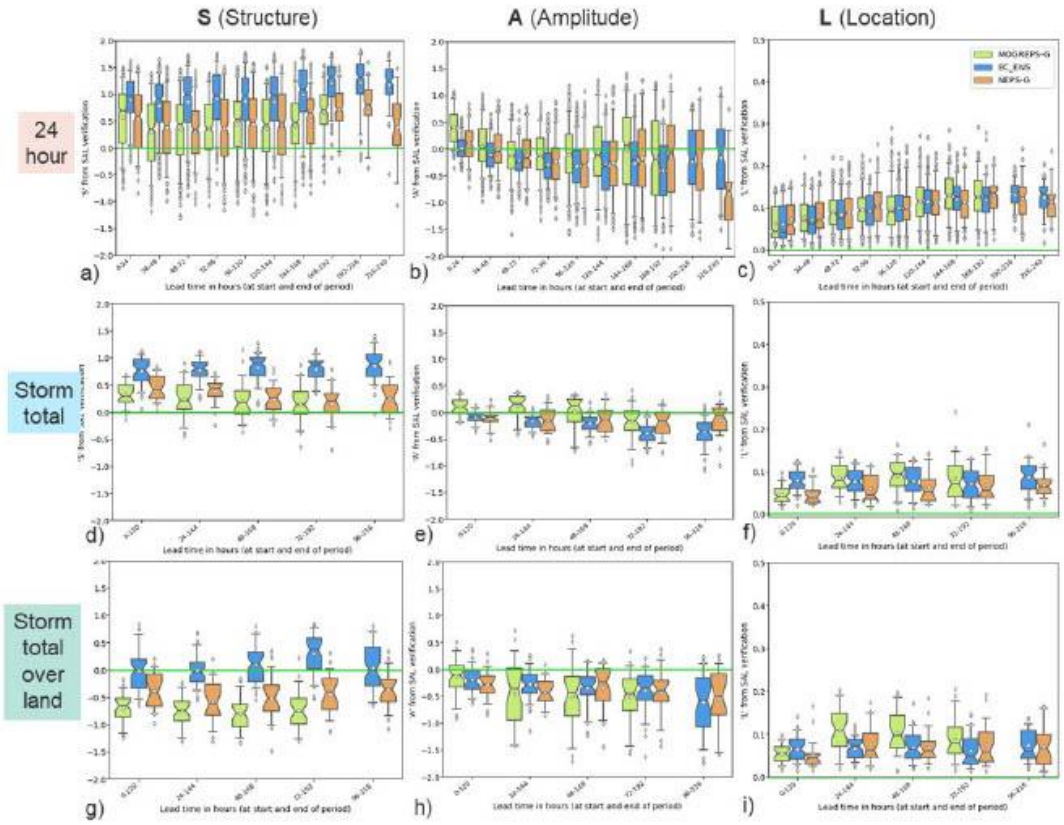


Figure 5: Box plots comparing the distribution of S, A, and L components of the SAL verification for rainfall in Cyclone Burevi for MOGREPS-G (green), ECMWF ENS (blue) and NEPS-G (orange) using GPM final data as the observation dataset: Top row (a-c) = combined for all 24-hour periods (from 00UTC 01/12/2020 to 00UTC 07/12/2020); Middle row (d-f) = storm total rainfall (defined as 00UTC 01/12/2020 to 00UTC 06/12/2020); Bottom row (g-i) = storm total over land only (also defined as 00UTC 01/12/2020 to 00UTC 06/12/2020, but with land/sea mask applied). All 00UTC forecast runs from 27/11/2020 to 06/12/2020 are included in the verification.

4.1.3. Cyclone Tauktae

The box-and-whisker plots summarising the forecasts for each SAL component in Cyclone Tauktae are displayed in Figure 6. There is notably large spread in the Structure and Amplitude components across all of the ensembles for the 24-hour rainfall forecasts (6a and 6b), which only reduces a little as lead time reduces. The structure of the storm total rainfall forecast objects are mainly too large/flat and are best for the NEPS-G forecasts

(6d). Rainfall amounts are underestimated at longer lead times but closer to zero at shorter lead times (6e). Again, once the land sea mask is applied the structure values lower and are often too sharp and small (6g), in particular in NEPS-G at short lead times and EC ENS at longer lead times. The rainfall over land is overestimated at short lead times but underestimated at longer leads (6h).

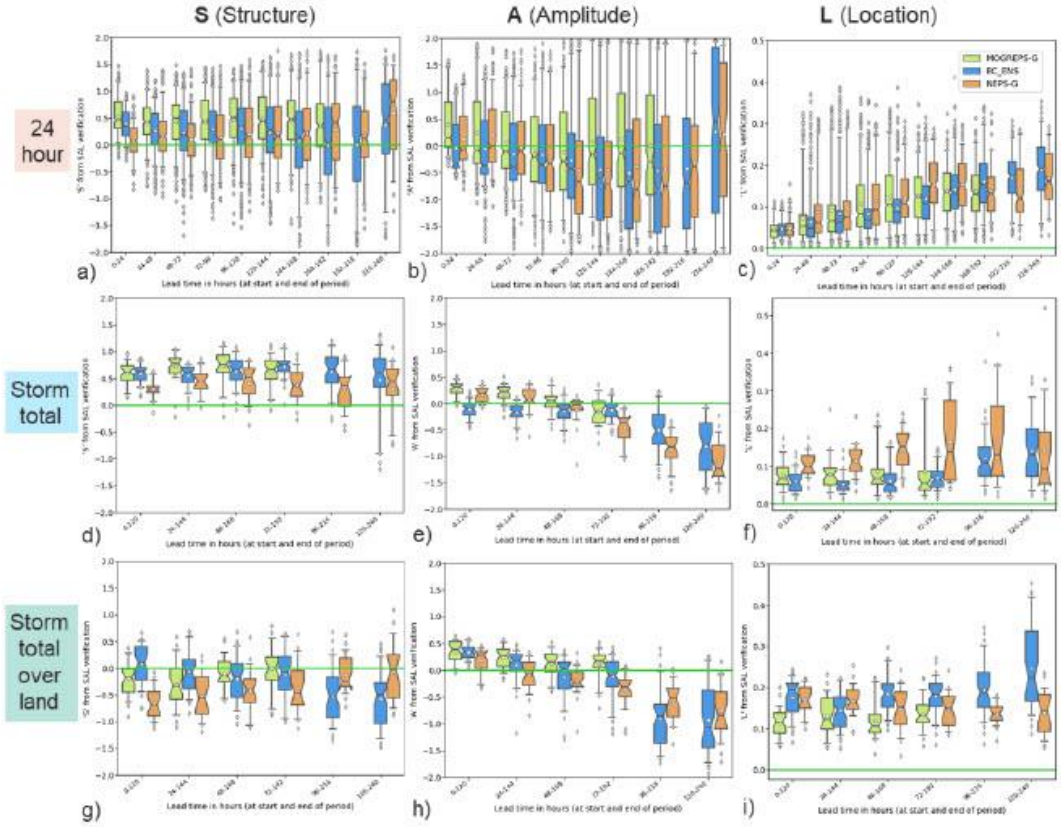


Figure 6: Box plots comparing the distribution of S, A, and L components of the SAL verification for rainfall in Cyclone Tauktae for MOGREPS-G (green), ECMWF ENS (blue) and NEPS-G (orange) using GPM final data as the observation dataset: Top row (a-c) = combined for all 24-hour periods (from 00UTC 14/05/2021 to 00UTC 21/05/2021); Middle row (d-f) = storm total rainfall (defined as 00UTC 15/05/2021 to 00UTC 20/05/2021); Bottom row (g-i) = storm total over land only (also defined as 00UTC 15/05/2021 to 00UTC 20/05/2021, but with land/sea mask applied). All 00UTC forecast runs from 10/05/2021 to 19/05/2021 are included in the verification.

4.1.4. Cyclone Gulab

The box-and-whisker plots for Cyclone Gulab show that there is little bias and a large spread in the structure of the 24-hour rainfall forecasts and this is similar across models (Figure 7a). There is a general underestimation of the rainfall amounts which improves at shorter lead times (7b). The Location component has lower spread and is closer to zero than in the other three cases (7c). There is a contrast between the storm total structure results before and after the application of the land/sea, being too large/flat overall, but too sharp/small when only rainfall over land is examined (7d and 7g). However, the overall estimation of rainfall amounts is similar in both (7e and 7h). As with the 24-hour plots, the Location scores are better for Gulab compared to the other cyclones.

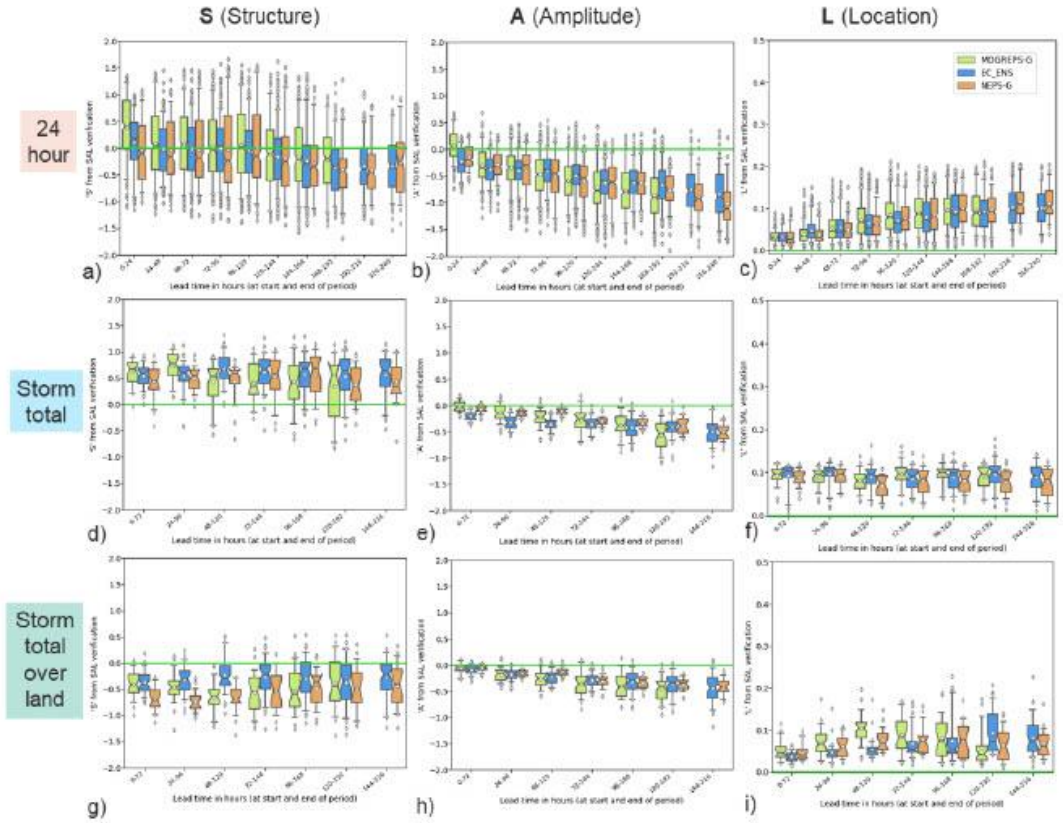


Figure 7: Box plots comparing the distribution of S, A, and L components of the SAL verification for rainfall in Cyclone Gulab for MOGREPS-G (green), ECMWF ENS (blue) and NEPS-G (orange) using GPM final data as the observation dataset: Top row (a-c) = combined for all 24-hour periods (from 00UTC 24/09/2021 to 00UTC 30/09/2021); Middle row (d-f) = storm total rainfall (defined as 00UTC 26/09/2021 to 00UTC 29/09/2021); Bottom row (g-i) = storm total over land only (also defined as 00UTC 26/09/2021 to 00UTC 29/09/2021, but with land/sea mask applied). All 00UTC forecast runs from 20/09/2021 to 28/09/2021 are included in the verification.

4.1.5. Summary

Does the SAL method provide a useful way to analyse ensemble rainfall forecasts in tropical cyclone cases?

The box-and-whisker plots provide a comprehensive and efficient way of summarising the Structure, Amplitude and Location components from SAL across the three global ensemble forecast models at a range of lead times and types of rainfall accumulation. SAL diagrams give the additional dimension of being able to look at the relationship between the parameters.

How do scores vary with lead time (the proximity to zero and spread between members)?

The S, A and L component scores generally improve (become closer to zero) as lead time reduces, but some cases e.g., storm total over land in Nivar/EC ENS buck this trend with overestimation of A at short leads. The change with lead time also varies over cases, with for example less of a change with lead time for the location component in the Gulab case, where the tropical cyclone track forecasts are more consistent.

How much do these results vary from case to case and are there systematic biases?

The results are similar for the two southern storms (for example MOGREPS-G and NEPS-G having better structure overall but EC ENS being much better for the object structures after a land/sea mask is applied). In the other two storms the results are generally more similar across models.

How different do the results look when only verify rainfall over land?

Some large changes are seen in the results when a land/sea mask is applied to the storm total rainfall forecasts. For example, ECMWF ENS structure forecasts are much better for rainfall over land in Nivar/Burevi than they are without the land/sea mask. To investigate the causes of these results further a postage stamp rainfall plot for an example ECMWF ENS and MOGREPS-G member with scores close to the ensemble median is shown in Figure 8. In the top row you can see that the rainfall forecast for the EC ENS member is too flat (S>0) and doesn't reach the same intensities as shown in GPM. This also results in an underestimation of the total rainfall (A<0), whereas the MOGREPS-G member has very good S and A scores (both close to zero). However, once the land sea mask is applied, the failure of the MOGREPS-G forecasts to bring the rainfall inland beyond the coastal strip means the object is too small and sharp (S<0) and the rainfall total underestimated (A<0), whereas the ECMWF ENS object has S and A values much closer to zero once the land/sea mask is applied.

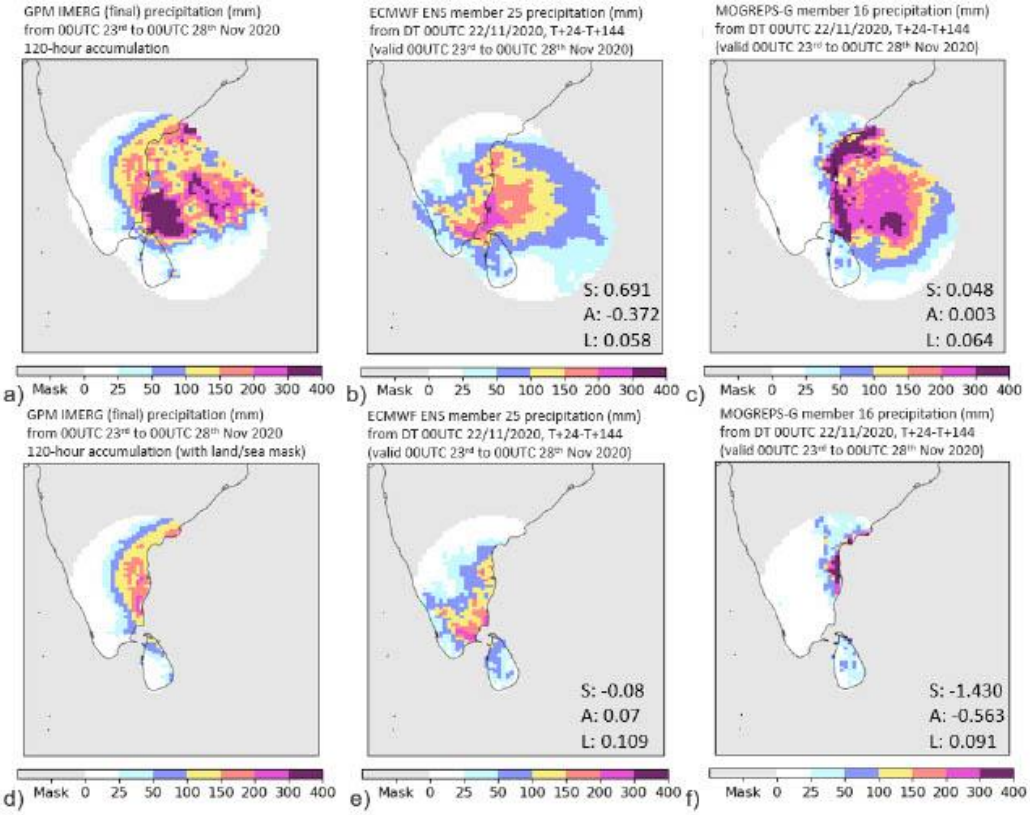


Figure 8: Postage stamps from an example member from ECMWF ENS (middle) and MOGREPS-G (right) for the 120-hour storm total rainfall for Cyclone Nivar from 00UTC 23rd to 00UTC 28th November 2020, compared with the corresponding GPM final data (left). The top row shows the data prior to applying a land sea mask and the bottom row shows the data after applying the land sea mask.

4.2. MODE

All ensemble members were passed through the MODE tool and verified against both GPM late and GPM final products for accumulations ending at 00UTC. Accumulations ending at 03UTC were also verified against the merged product. The use of the different observation product didn't substantially change the narrative, and hence only the 00UTC results from GPM final are included here. No tropical cyclone mask was used in the MODE experiments. A land mask was used on the storm total accumulations to see if differences in behaviour over land could be detected.

Here, a threshold of 150mm was selected for storm total accumulations to identify objects. MODE contains a setting which defines a threshold value to ensure the frequency bias between the forecast and observation fields is close to one. This is effectively a bias-removal tool. It was opted to use this setting here for the storm total accumulations because an initial inspection of the raw data identified some large discrepancies between forecast and observed values. For the 24-hour accumulations, a threshold of 65mm was used, taken from Table 2 of Barde et al. (2020) to align with the "heavy rain" category and therefore assist with identification of rainfall objects from tropical cyclones.

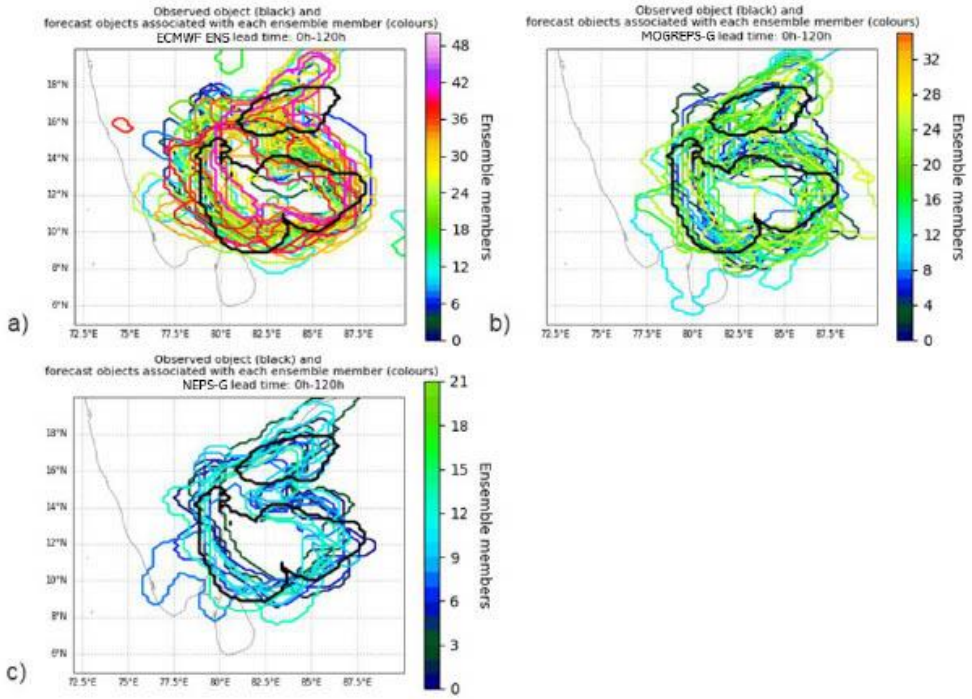


Figure 9: Example of the objects identified in the T+0–T+120h forecast 120-hour accumulations from the 00UTC forecast run on 23th November 2020 (valid from 00UTC 23rd to 00UTC 28th November 2020) from each member of each ensemble model: a) ECMWF ENS, b) MOGREPS-G and c) NEPS-G. The black object is the observed object.

An example of the objects we found with Cyclone Nivar is illustrated in Figure 9. Each ensemble member is a different colour, with the member colours consistent across the ensemble models (ECMWF ENS has more members, hence uses more colours). The black outline is for the observed object. NEPS-G objects seem to better correspond, visually, with the size and orientation of the observed object. ECMWF ENS objects subjectively look to be rather too large, as are the MOGREPS-G objects, but they have different orientations.

4.2.1. Cyclone Nivar

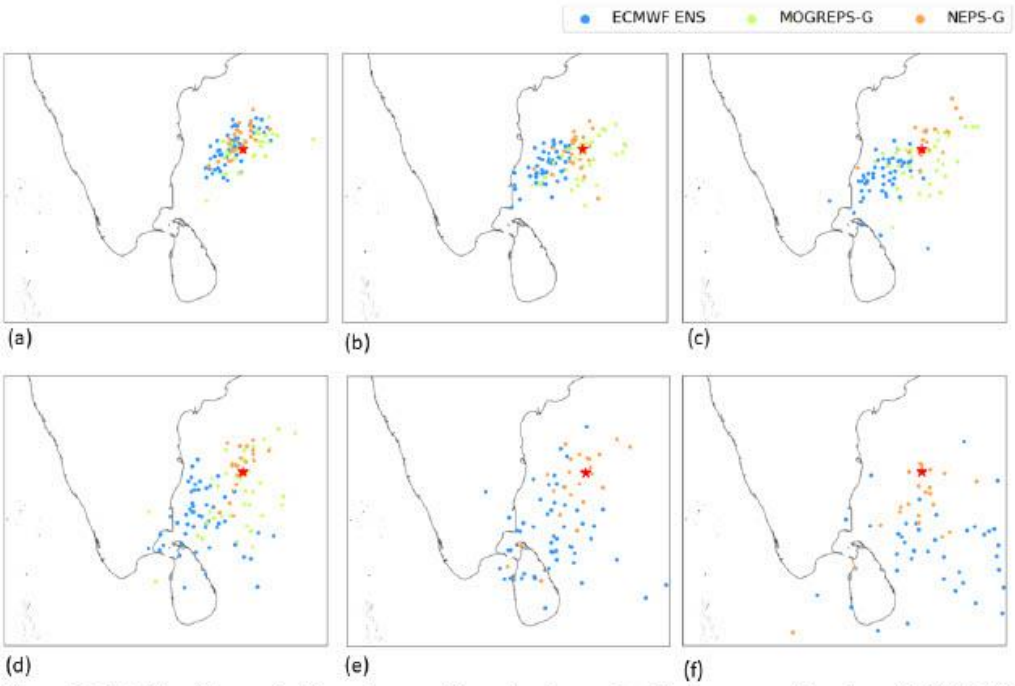


Figure 10: Centroid positions marked for each ensemble member from each of the three ensemble systems (ECMWF ENS, blue; MOGREPS-G, green and NEPS-G, orange) for 120-hour accumulations (23rd – 28th November). No land/sea mask is applied. Each panel contains centroid positions from different lead time forecasts (a) T+0–T+120h (b) T+24–T+144h; (c) T+48 – T+168h; (d) T+72 – T+192h; (e) T+96 – T+216h and (f) T+120 – T+240h

Figure 10 shows the evolution of the centroid position of the storm total forecast objects with lead time, with the observed location marked with a star. For accumulations from the lead time range 0–120 hours (Figure 10a), all three ensembles have centroids clustered around the observed centroid location. By lead time range 72–192 hours (d), the ECMWF ensemble centroids appear to be mostly to the south and west of the observed centroid and further west than the other two ensembles. NEPS-G centroids appear to have a more limited spatial spread than either ECMWF ENS or MOGREPS-G. At the longest lead time of 120–240 hours (f), there appear to be large differences in how ensemble members have evolved the TC; the centroids are mostly to the south of the observed value, but again NEPS-G has a smaller spread in positions.

Box plots showing the distributions of object attributes with forecast lead time from each of the three models for Cyclone Nivar are displayed in Figure 11. Cluster objects are used to generate the distributions, which means that unmatched objects (false alarms and misses) are also included. Here, the sum of the intensity within the objects is used as a proxy for the volume of rain from the tropical cyclone (left-hand column). The 90th centile value of the intensity is used as a proxy for the maximum rain (middle column). The final column contains the area exceeding the threshold within the object; comparing the areas that exceed the threshold within the objects to the total areas of the objects (not shown) can indicate differences in the rainfall patterns in the forecast and observed fields.

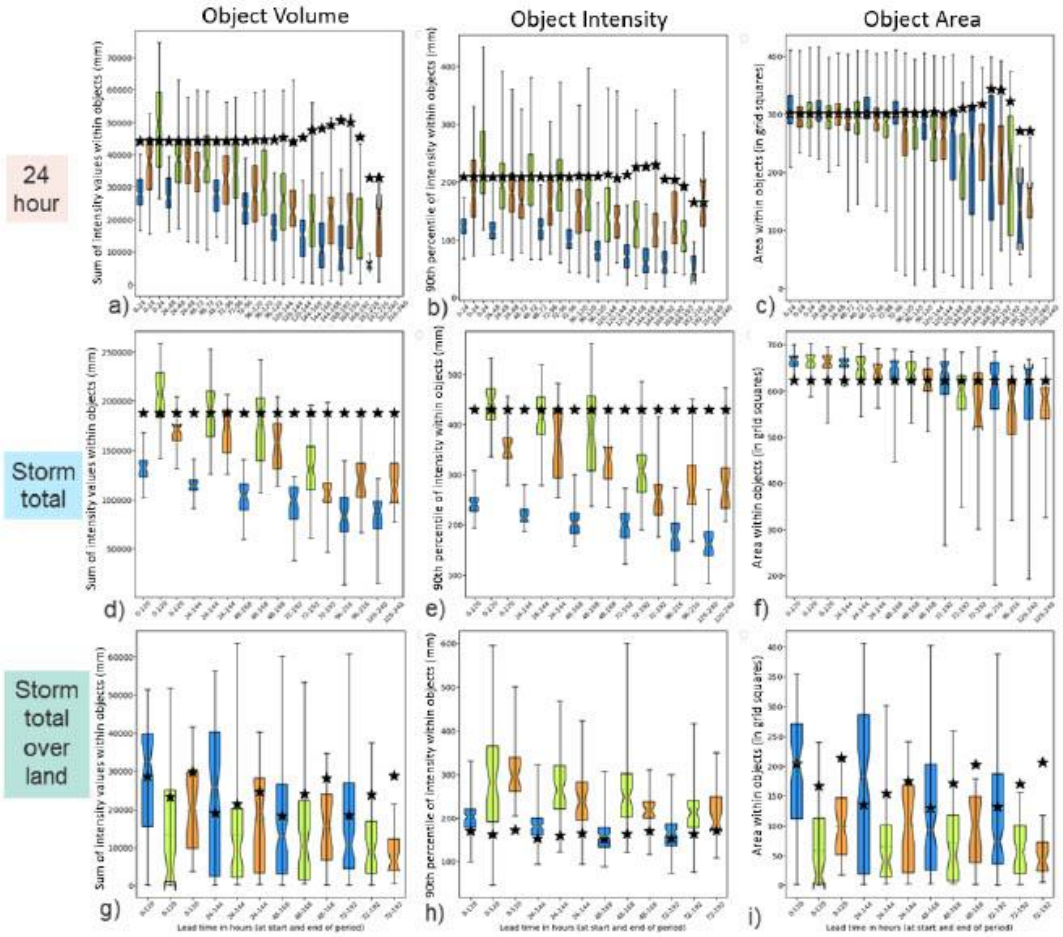


Figure 11: Cluster object distributions in Cyclone Nivar from MOGREPS-G (green), ECMWF ensemble (blue) and NEPS-G (orange). Mean value from observed objects is plotted as a black star. Results from 24-hour accumulations for all valid times are on the top row. Storm total (120-hour) accumulations are in the middle row and storm total (120-hour) accumulations only over land are in the bottom row. The 120-hour accumulations are valid for the period between 23rd and 28th November 2020. a), d) and g) Sum of intensity values within objects; b), e) and h) values for 90th percentile of intensity within objects; c), f), i) area within objects, in grid squares, exceeding the set threshold.

Rainfall volumes appear to be generally underestimated compared to those observed, for both daily and storm-total accumulations (Figure 11a, 11d, and 11g). Daily totals are under-forecast (a) with a large drop-off with lead time. MOGREPS-G appears to over-forecast the rainfall volume at short lead times. For the storm total, ECMWF ENS has far too little rainfall in the objects and there isn't a wide variation in amounts from the ensemble members (d). However, ECMWF ENS better captures the rainfall amounts over land than the two UM-based ensemble models (g), which have too little rain in the storm at all lead times supporting the conclusions drawn in section 4.1. Interestingly, the area within the objects exceeding the threshold appear to match well with that observed at short lead times for the daily accumulations (out to day four) (c). From a storm total perspective, short lead times have too large an area with these large accumulations (all three models display this tendency) (f). Over land, the picture changes and the UM-based ensembles have too small an area exceeding the threshold (i).

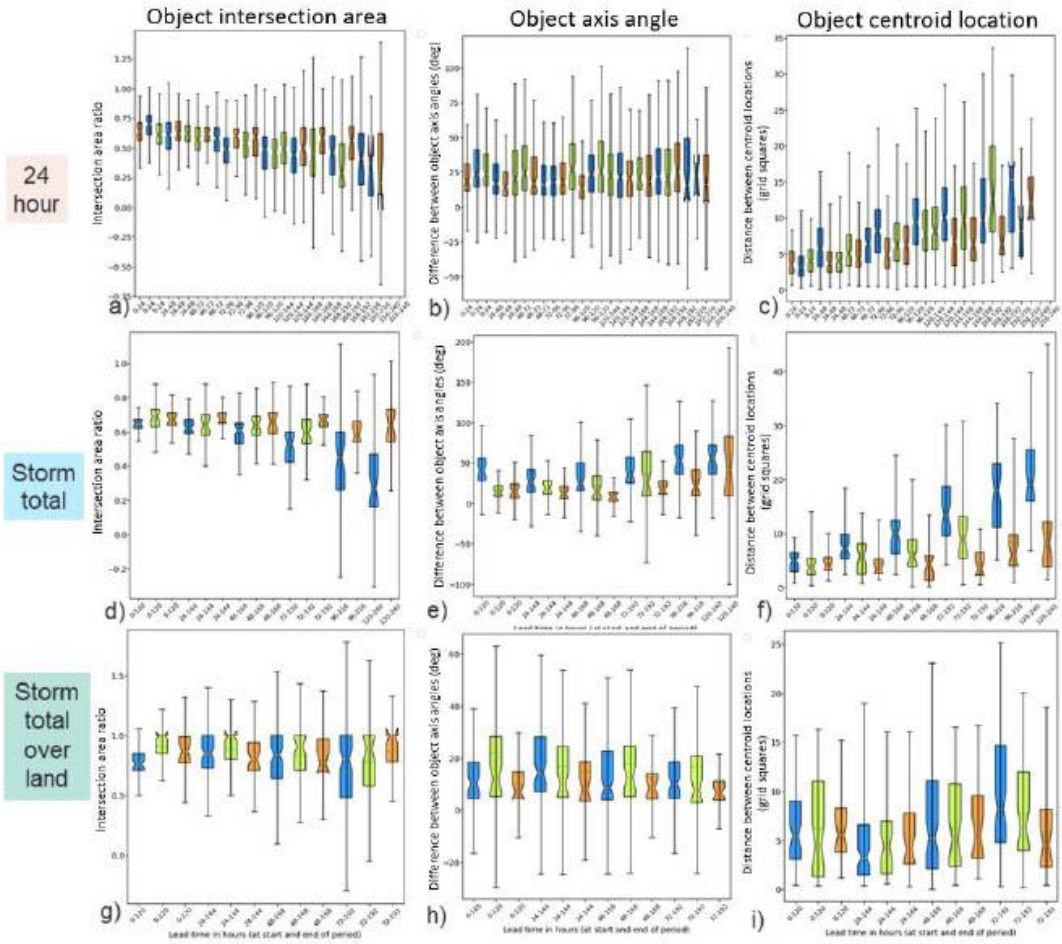


Figure 12: Attributes from the paired-cluster objects for Cyclone Nivar with lead time for ECMWF ENS (blue), MOGREPS-G (green) and NEPS-G (orange). Accumulations end at 00UTC. a), d), g) The ratio of the intersection area to the lesser of the forecast or observed area (ideal value is 1); b), e), h) the difference between the axis angles of two objects; c), f), i) the distance between the forecast and observed centroid locations in grid squares. Top row considers daily accumulations, middle row is for storm total (120h) accumulations and the bottom row is for storm total accumulations but only over land.

Another way of looking at the results from MODE is to assess the paired objects (matched between forecast and observation). Figure 12 shows an example of the results for Nivar. Where an object is identified in both the forecast and observations, they are within ~100km at the early lead times (right-hand column). There's a larger spread of centroid differences in MOGREPS-G than the other two models at early lead times, but ECMWF ENS appears to have an increased spread in values by the later lead times. The angle difference attribute (middle column) highlights whether the forecast and observed objects are oriented in a similar direction within the field. Angle values of up to 90 degrees are possible, so the maximum difference could be 180 degrees. Storm-total objects only over land appear to be well-matched in terms of orientation. Analysing the paired object attributes is only useful when objects can be identified in both the forecast and the observation field. This type of plot is only being included for Nivar as a different example of how to present the matched-pair object attributes which come out of the software.

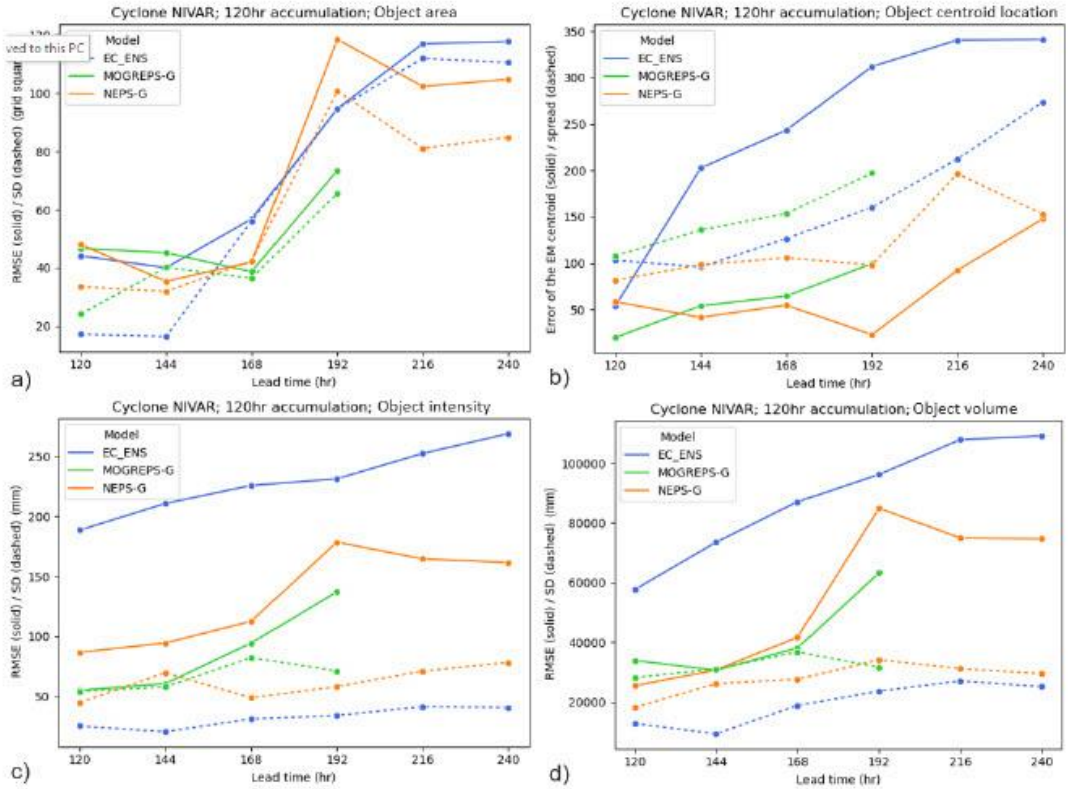


Figure 13: Spread and error plots for the MODE object verification for the 120-hour accumulations (23rd – 28th November, no land/sea mask applied), for each of the MODE attributes: a) Area; b) Centroid location; c) 90th percentile intensity within the objects; d) Sum of the intensity values in the objects. For a, c and d, the RMSE and standard deviation are shown; for b, it is the error of the ensemble mean centroid and the spread in the centroids from each ensemble members (measured by their distance from the ensemble mean.

The spread and error in the attributes, and how this varies with the lead time of the forecast, is shown in Figure 13. As expected, both the error and the spread increase with lead time. The models are under-dispersed (spread less than error) for the intensity and volume measures (c and d), and slightly under-dispersed for area (a). For location the results are more complex, being under-dispersed for ECMWF ENS but over-dispersed for MOGREPS-G and NEPS-G (b).

4.2.2. Cyclone Burevi

The object attribute distributions for Cyclone Burevi are displayed in Figure 14. Again, the storm total rain volumes and intensities are too low, especially for ECMWF ENS (d). The area within objects exceeding the threshold for the storm totals appears consistent with the observed values; the daily accumulations appear to have a larger interquartile range for all three models than the previous storm (c). The observed value sits towards the lower end of the range of forecast values. For Burevi, no objects are identified at lead times of 120–240 hours at the selected threshold when the land sea mask is used.

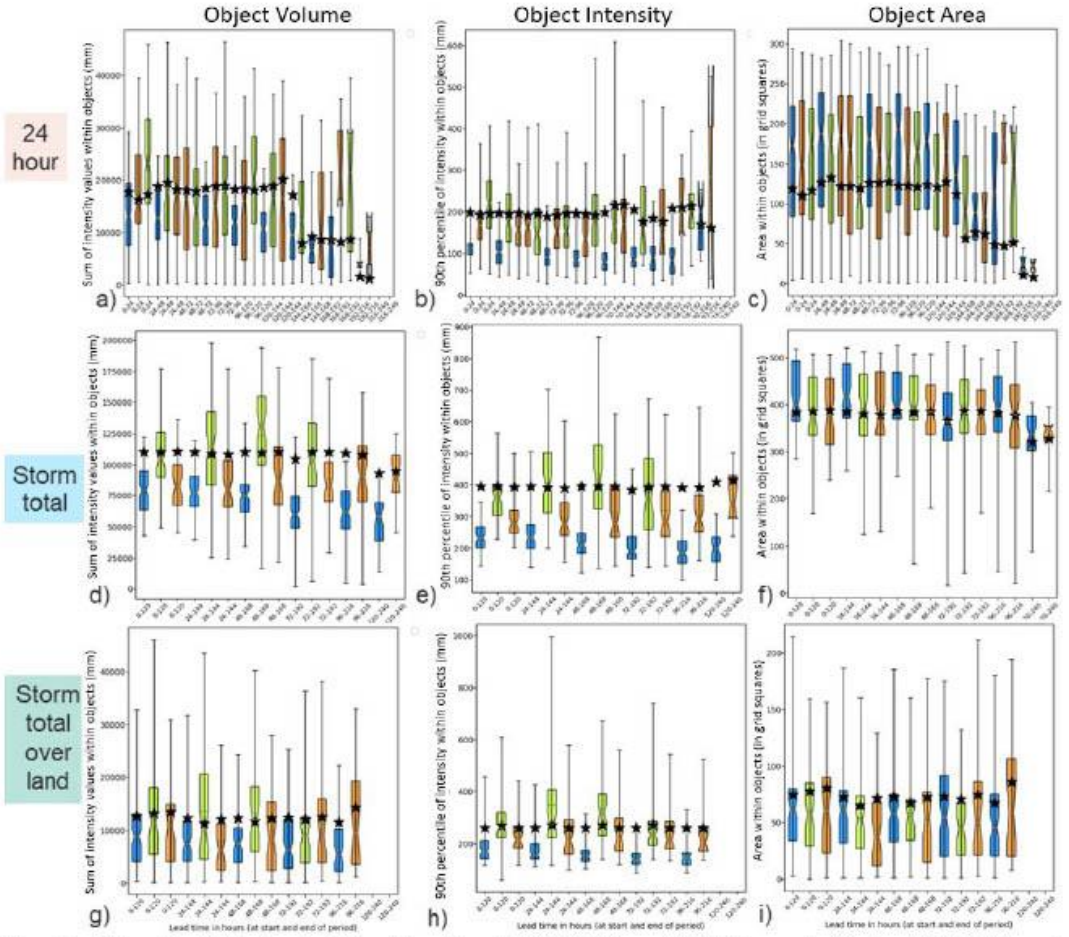


Figure 14: Cluster object distributions in Cyclone Burevi from MOGREPS-G (green), ECMWF ensemble (blue) and NEPS-G (orange). Mean value from observed objects is plotted as a black star. Results from 24-hour accumulations for all valid times are on the top row. Storm total (120-hour) accumulations are in the middle row and storm total (120-hour) accumulations only over land are in the bottom row. a), d) and g) Sum of intensity values within objects; b), e) and h) values for 90th percentile of intensity within objects; c), f), i) area within objects, in grid squares, exceeding the set threshold.

The object attributes can be used along with the centroid locations to look for any systematic behaviour of ensemble members with location (Figure 15). Objects which have an attribute value within 10% of the observed value are categorised 'Near obs' and coloured green. The centroids are spread out spatially far more than seen for Cyclone Nivar indicating less predictability for this storm. As seen with Nivar, the centroid locations are generally to the south and west of the observed centroid location, especially for ECMWF ENS. ECMWF ENS has more members whose centroids are over land – to the west of the cluster.

ECMWF ENS objects appear on the small side (Figure 15a). The ensemble members which have objects smaller than observed mostly appear to the north and west of the observed centroid. MOGREPS-G has a few members whose areas are too large (b). The intensity of the objects is too low in ECMWF ENS (d); there are some members with objects within 10% of the observed value for both MOGREPS-G and NEPS-G. MOGREPS-G has a few members in the vicinity of the observed centroid which are too intense (e). This figure is included here to showcase the capability but won't be shown for the other cases.

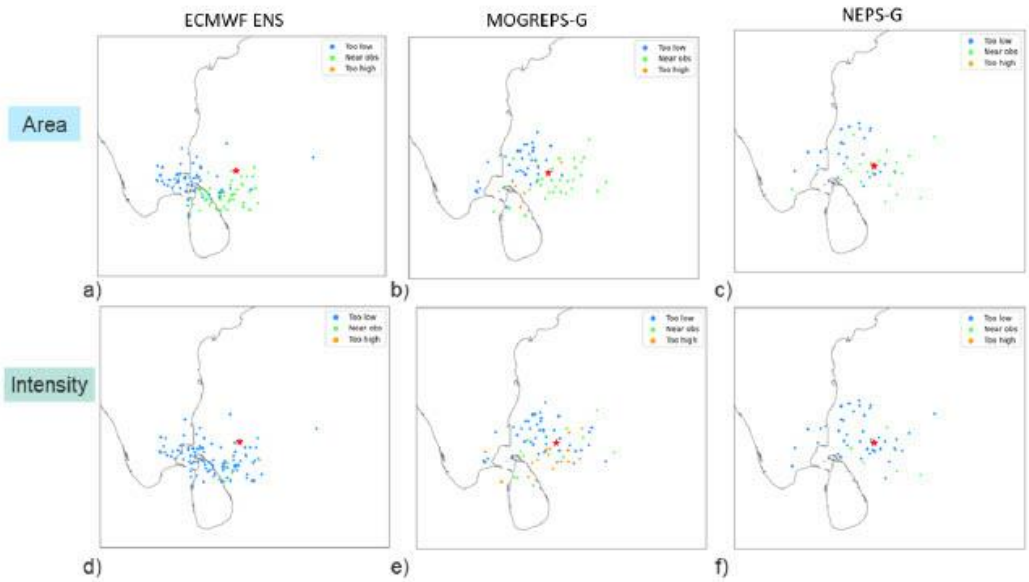


Figure 15: Centroid positions coloured according to object attribute, from 120-hour accumulations (0-120 hours lead time) for Cyclone Burevi from ECMWF ENS (left), MOGREPS-G (centre), NEPS-G (right). a), b), c) centroid location coloured according to object area. d), e), f) centroid location coloured according to rainfall volume (sum of intensity within objects)

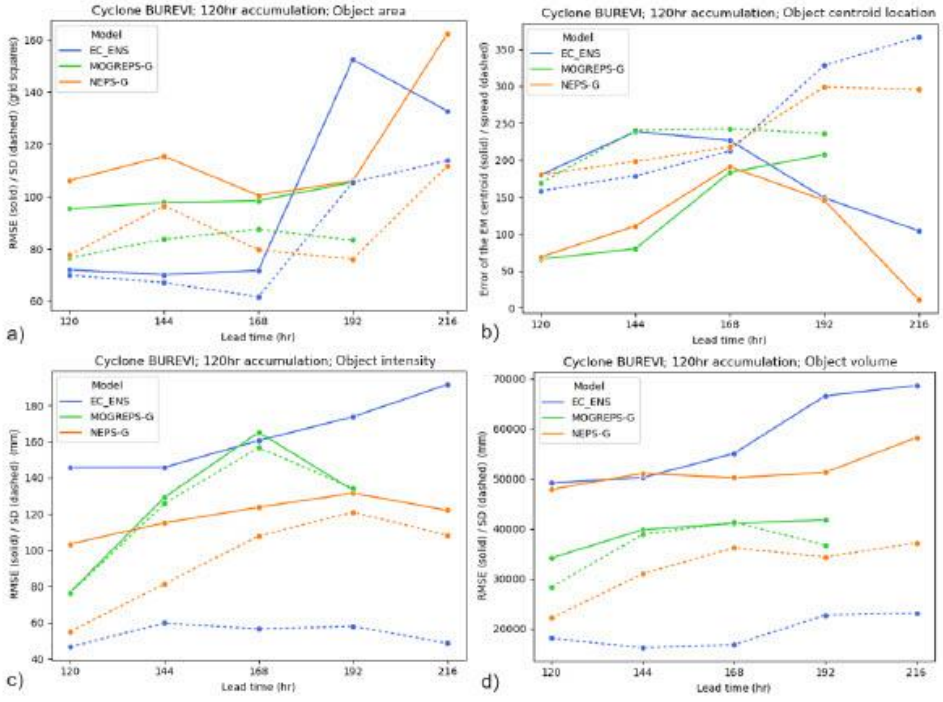


Figure 16: Spread and error plots for the MODE object verification for the 120-hour accumulations (X-X, no land/sea mask applied), for each of the MODE attributes: a) Area; b) Centroid location; c) 90th percentile intensity within the objects; d) Sum of the intensity values in the objects. For a, c and d, the RMSE and standard deviation are shown; for b, it is the error of the ensemble mean centroid and the spread in the centroids from each ensemble members (measured by their distance from the ensemble mean).

The spread and error in the object attributes in the Burevi case is shown in Figure 16. The models are under-dispersed (spread less than error) for the intensity and volume (c and d) and the area (a), with MOGREPS-G having the best match. For centroid location the results are more complex, but over-dispersed for MOGREPS-G and NEPS-G (b).

4.2.3. Cyclone Tauktae

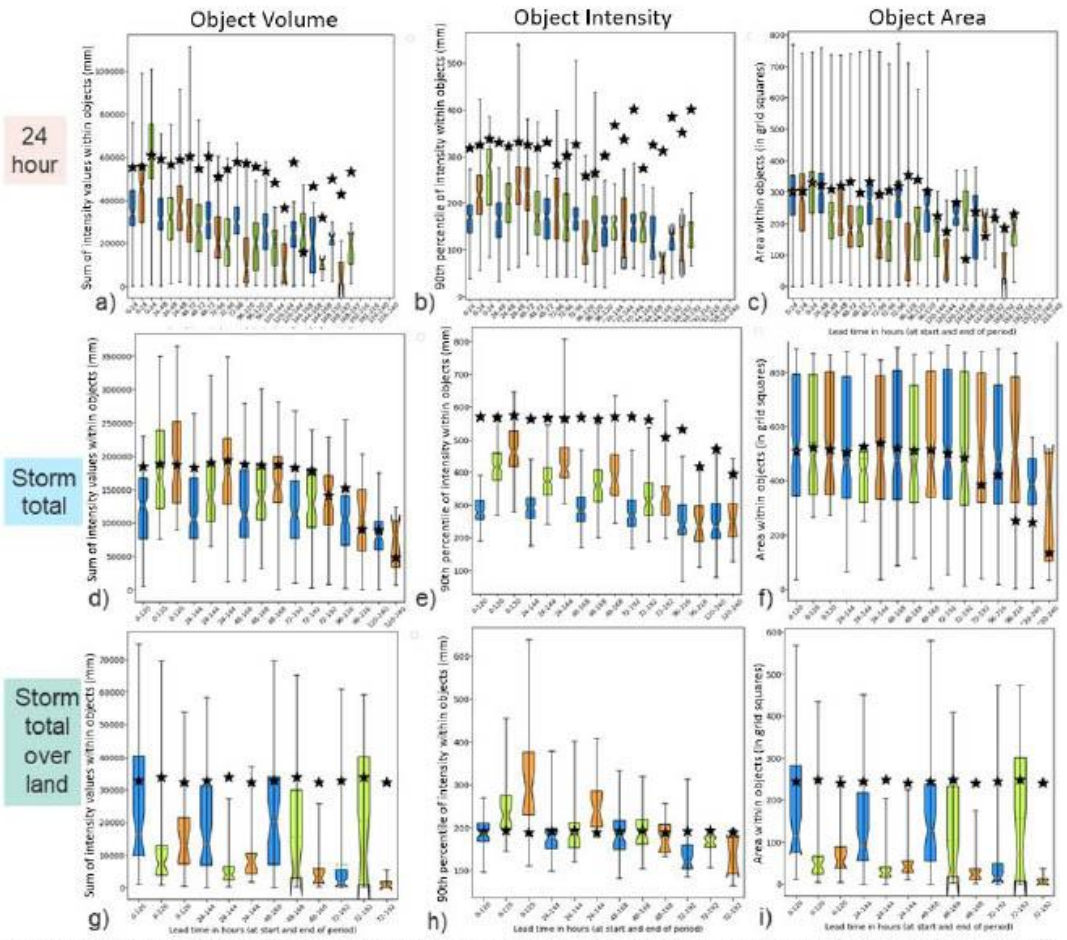


Figure 17: Cluster object distributions in Cyclone Tauktae from MOGREPS-G (green), ECMWF ensemble (blue) and NEPS-G (orange). Mean value from observed objects is plotted as a black star. Results from 24-hour accumulations for all valid times are on the top row. Storm total (120-hour) accumulations are in the middle row and storm total (120-hour) accumulations only over land are in the bottom row. a), d) and g) Sum of intensity values within objects; b), e) and h) values for 90th percentile of intensity within objects; c), f), i) area within objects, in grid squares, exceeding the set threshold.

Cyclone Tauktae has similar variability in the box plots as in Cyclone Burevi. Object centroids are spread over a wide area, even at the shorter lead times (not shown). NEPS-G has larger storm total rainfall intensity over land at the shortest lead time range (Figure 17h); larger than either observed or the other two models. This storm has less rainfall over land than the other cases. No objects are identified at the longest two forecast ranges for the 24-hour rainfall objects; the storm total GPM observation shows two distinct areas of rainfall in the Arabian Sea, one of which is Tauktae. Inspection of the output from each member shows large variability in the rainfall evolution in this region.

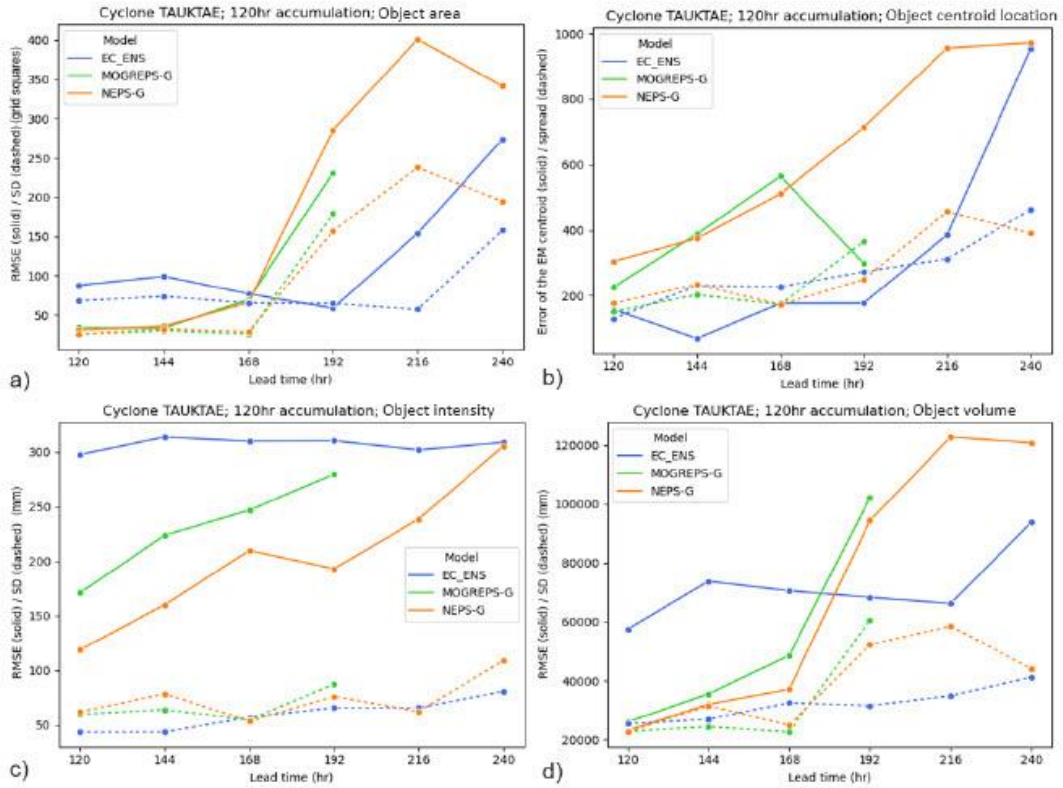


Figure 18: Spread and error plots for the MODE object verification for the 120-hour accumulations (X-X, no land/sea mask applied), for each of the MODE attributes: a) Area; b) Centroid location; c) 90th percentile intensity within the objects; d) Sum of the intensity values in the objects. For a, c and d, the RMSE and standard deviation are shown; for b, it is the error of the ensemble mean centroid and the spread in the centroids from each ensemble members (measured by their distance from the ensemble mean).

The spread and error in the object attributes for Tauktae is shown in Figure 18. The models are under-dispersed (spread less than error) for the intensity and volume measures (c and d) and the area (a) for all models, with this being particularly pronounced for the 90th percentile intensity attribute. For centroid location the model is under dispersed for NEPS-G and MOGREPS-G, but better matched for ECMWF ENS (b).

4.2.4. Cyclone Gulab

For Cyclone Gulab, the forecast 90th percentile of intensity within the objects for the 24-hour and storm total accumulations are lower than observed (Figure 19b and 19e). MOGREPS-G and NEPS-G distributions span a higher range of values, with larger means, than ECMWF ENS at most lead times. However, the storm total over land 90th percentile intensities better match the observed values. The storm total areas are too small over land (Figure 19i), but the mean area values correspond well with those observed before the land/sea mask is applied. It should be noted that the interquartile ranges for each model are large for the area attribute, suggesting different sized objects from each ensemble member.

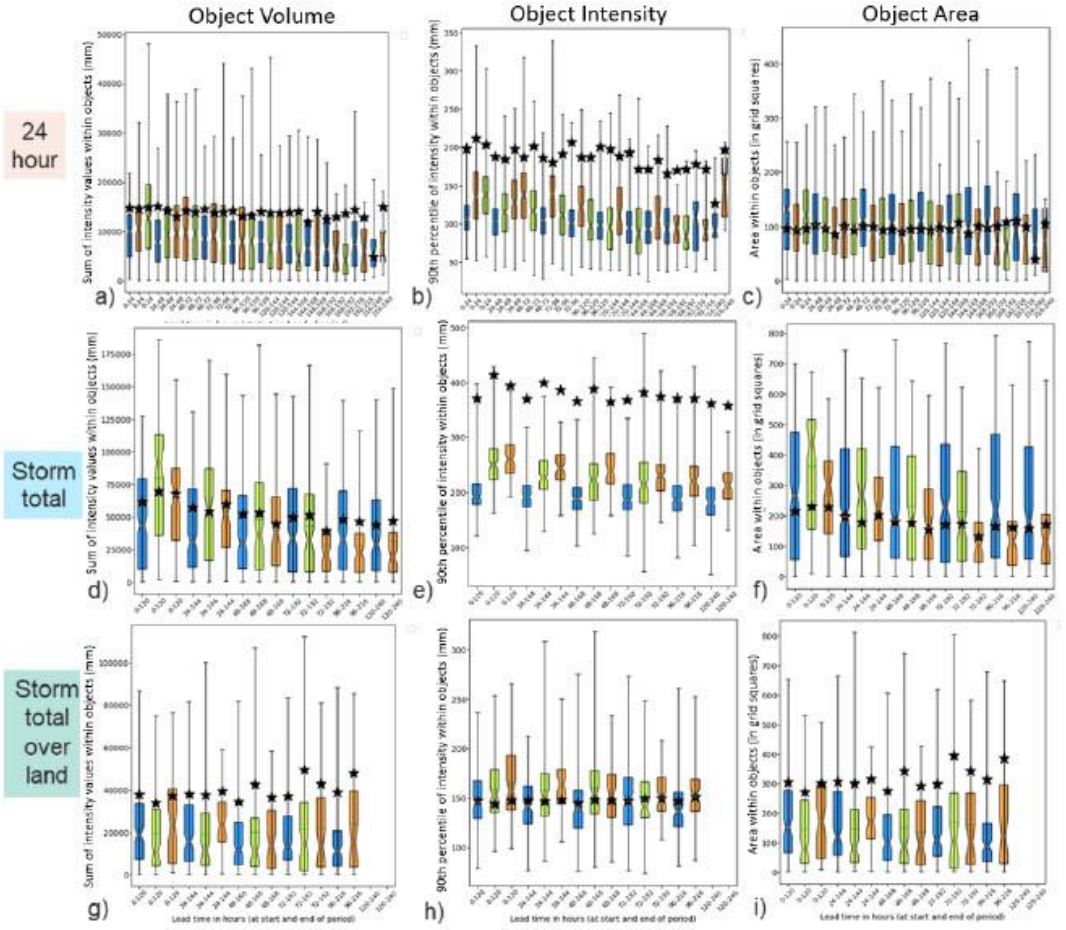


Figure 19: Cluster object distributions in Cyclone Gulab from MOGREPS-G (green), ECMWF ensemble (blue) and NEPS-G (orange). Mean value from observed objects is plotted as a black star. Results from 24-hour accumulations for all valid times are on the top row. Storm total (120-hour) accumulations are in the middle row and storm total (120-hour) accumulations only over land are in the bottom row. a), d) and g) Sum of intensity values within objects; b), e) and h) values for 90th percentile of intensity within objects; c), f), i) area within objects, in grid squares, exceeding the set threshold.

The spread and error in the object attributes in the Cyclone Gulab case is shown in Figure 20. The models are generally under-dispersed (spread less than error) for the intensity and volume measures (c and d) and the area (a) for all models. For centroid location the error scores are quite jumpy between lead times and more constant for the spread (b).

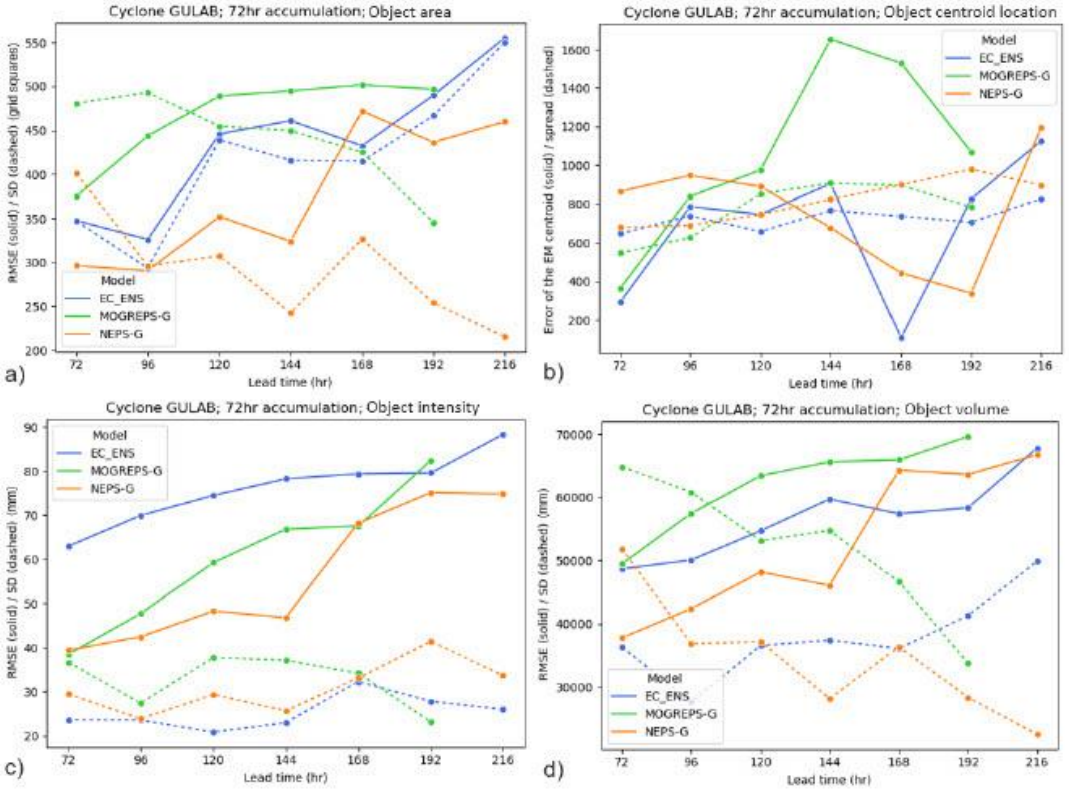


Figure 20: Spread and error plots for the MODE object verification for the 72-hour accumulations (X-X, no land/sea mask applied), for each of the MODE attributes: a) Area; b) Centroid location; c) 90th percentile intensity within the objects; d) Sum of the intensity values in the objects. For a, c and d, the RMSE and standard deviation are shown; for b, it is the error of the ensemble mean centroid and the spread in the centroids from each ensemble members (measured by their distance from the ensemble mean).

4.2.5. Summary

Does MODE provide a useful method for analysing ensemble forecasts of rainfall from tropical cyclones?

Yes, however care must be taken to investigate suitable settings to identify objects of interest. Attributes examined here were used as proxies for rainfall volume, intensity and areal coverage. There are other useful attributes which can be studied, for example orientation and aspect ratio. It is important to ensure a suitable threshold/convolution radius pair is chosen. Too low a threshold (or convolution radius) increases the number of objects identified. It is also important to choose a merge threshold which ensures objects which would be merged at the lower threshold do in fact link areas which would be identified as one object by a meteorologist's eye. MODE does provide a 'quilt plot' setting to experiment with values which should be used before running the tool proper. This allows you to investigate how the number of objects changes with convolution radius/threshold combinations.

Can changes in rainfall behaviour over land, or other systematic issues be seen using this method?

Yes, the box plots from the ensembles clearly show systematic errors in the attribute under consideration. Object size and shape can also be visually investigated. Plotting object centroids coloured according to an object attribute is a useful way to diagnose location-dependent errors.

All four cases indicated that storm-total rain volume, and the intensity within the objects, were too low. Object

areas which exceed the threshold were well-matched in general for daily accumulations but look too large at early lead times for storm-total amounts. 90th percentile intensities appear to be too low for all three ensembles, for both storm total rainfall and daily accumulations. However, early lead times in the UM-based ensembles for Cyclone Nivar have a larger than observed 90th percentile for the 24-hour object intensity. ECMWF appears to better capture rainfall intensities over land. Only in Cyclone Burevi is the storm total 90th percentile less than observed. For the other three cases, ECMWF ENS matches well the observed storm total 90th percentile.

If a forecast has a sizeable bias compared to the observation, then it can be difficult to choose a suitable threshold where a sensible number of objects are identified. MODE provides a solution to this issue: an *'==FBIAS1'* option exists which can be used to replace an absolute forecast threshold, say, when a particular value is chosen for the observed threshold. MODE determines the threshold at which the average value of the frequency bias across the field equals one.

Issues which should be considered when preparing your analysis

At the shorter accumulation lengths, it's very difficult to know which objects pair together, or perhaps would be merged at a lower threshold. In this case, it's very easy to get the situation where more than one object is identified in the observation grid. Masking the parts of the field where rainfall is not of interest helps to narrow down finding objects to only those where the rainfall generated is of interest. This occurs at the risk of excluding some parts of the rain object if the forecast is particularly bad in terms of location (which can be the case at long lead times).

4.3. CRA using 'excRa'

Spatial verification of ensemble rainfall forecasts obtained from NEPS-G is presented using the newly developed 'excRa' discussed in section 3.3. It was not possible to compare the results for MOGREPS-G and ECMWF ENS during the timescale of this study, but this could be undertaken in the future. The approach used in this study closely follows Gallus (2010) who used the CRA verification method for ensemble rainfall forecasts. The CRA spatial verification method allows decomposition of total error in the forecast into errors due to displacement, volume and pattern errors. This can help in analyzing the individual contribution from each of these components to the total error. The verification can be carried out for different rainfall thresholds like, 10, 20, 40 and 80 mm/day, but for brevity results are presented for just a 40 mm/day threshold applied to the four tropical cyclone cases discussed in section 2.1. The different attributes like maximum rainfall intensity, volume and area are verified using standard ensemble verification metrics. A detailed example of the CRA verification process using excRa, for Day-1 forecast rainfall valid on 27th Nov 2020, is shown in the Appendix.

4.3.1. Cyclone Nivar

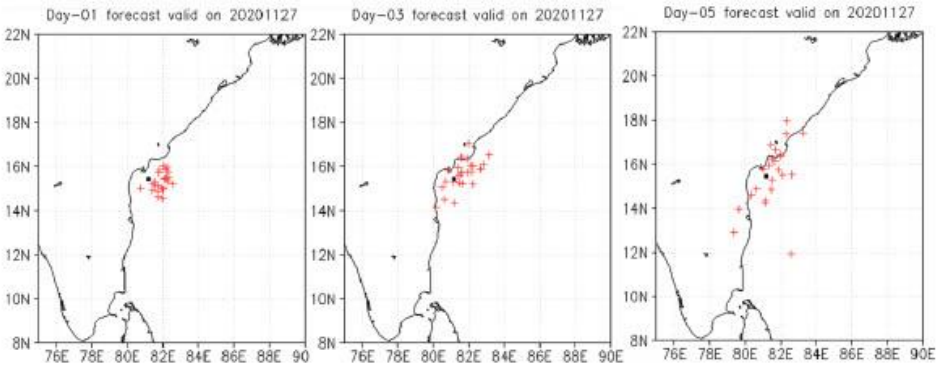


Figure 21. Centre of gravity (COG) in the 40mm/day CRA object in Observed (black) and ensemble member forecasts valid on 27th Nov 2020 for the case of Bay of Bengal tropical cyclone Nivar.

In the CRA verification, the first step after isolating the objects based on the user-defined rainfall threshold is to compute and compare the observed and forecast object parameters: object area, rain rate, rain maximum, and center of gravity (COG). A comparison of observed and forecast COG is presented in Figure 21 in Day-1, Day-3, and Day-5 forecasts. In Day-1, the COGs from ensemble members are closely clustered close to the COG from the observed object. In Day-3 and Day-5 forecasts, the spread among the members is higher, but the observed COG is located well within the extent of COGs of ensemble members.

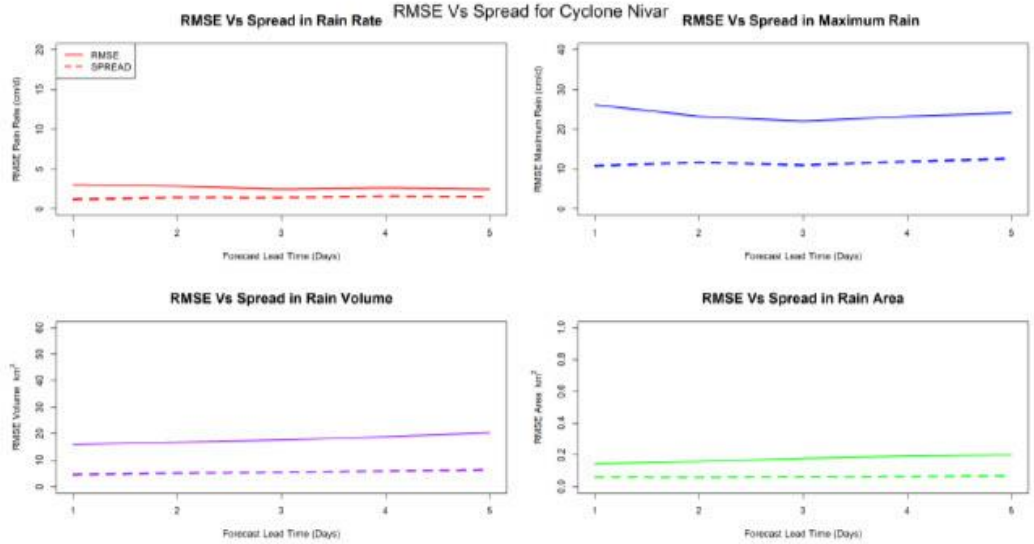


Figure 22. Spread Vs Bias in the 40 mm/day object parameters (I) Rain rate (ii) Maximum Rain (iii) Rain Volume (iv) Rain area for the case of Bay of Bengal tropical cyclone Nivar during 22-27 Nov 2020.

The RMSE Vs Spread in the attributes for Cyclone Nivar is shown in Figure 22. It is seen from the figure that the EPS system shows under dispersion (spread is smaller than the RMSE) for all lead times and all attributes. The maximum difference between the RMSE and spread is seen for rainfall maximum followed by volume. Also, it is seen that the maximum rainfall shows the highest error (highest RMSE) and the uncertainty among the ensemble members as compared to the other attributes. Rainfall area shows the lowest RMSE and Spread among the EPS members.

The percentage contribution from errors in displacement, volume and pattern to the total mean squared error (MSE) obtained for Cyclone Nivar is displayed in Figure 23. The maximum contribution to the total MSE is from the pattern error. The contribution from the volume error shows the lowest median values which range from (13% in Day-1 to about 20% in the Day-5 forecasts). The displacement error shows median values which are slightly higher than the volume error. But the difference between the upper and lower quartiles of the displacement values is much smaller than in the pattern and volume. This shows that the position spread among the members is not too large in the case of Cyclone Nivar.

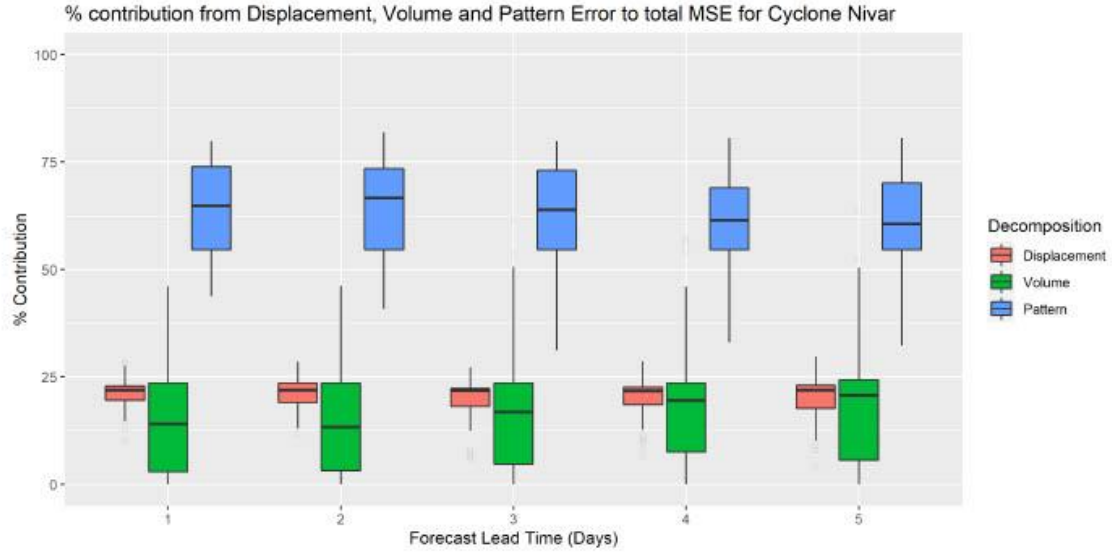


Figure 23. Box and whisker plot comparing the percentage contribution from errors in displacement, volume and pattern to the total mean squared error (MSE) in forecast 24 hr accumulated rainfall across the ensemble for Cyclone Nivar during 22-27 Nov 2020 at a range of lead times.

4.3.2. Cyclone Burevi

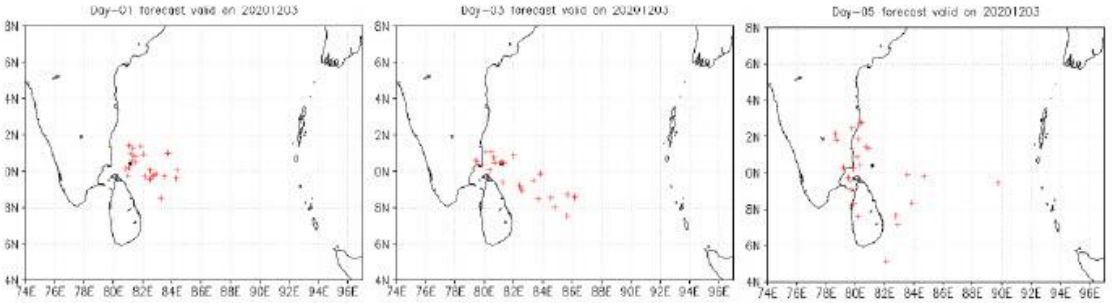


Figure 24. Centre of gravity (COG) in the 40mm/day CRA object in Observed (black) and ensemble member forecasts valid on 3rd Dec 2020 for the case of Bay of Bengal tropical cyclone Burevi.

Similar to Figure 21, a comparison of observed and forecast COG is presented in Figure 24 for Day-1, Day-3, and Day-5 forecasts. In Day-1, the COGs from ensemble members are closely clustered close to the COG from the observed object. As can be seen from the first panels in Figure 21 and 24, the members seem to lag behind the observed COG in Day-1. In the Day-3 and Day-5 forecasts, the spread among the members is higher but the observed COG is located well within the extent of COGs of ensemble members.

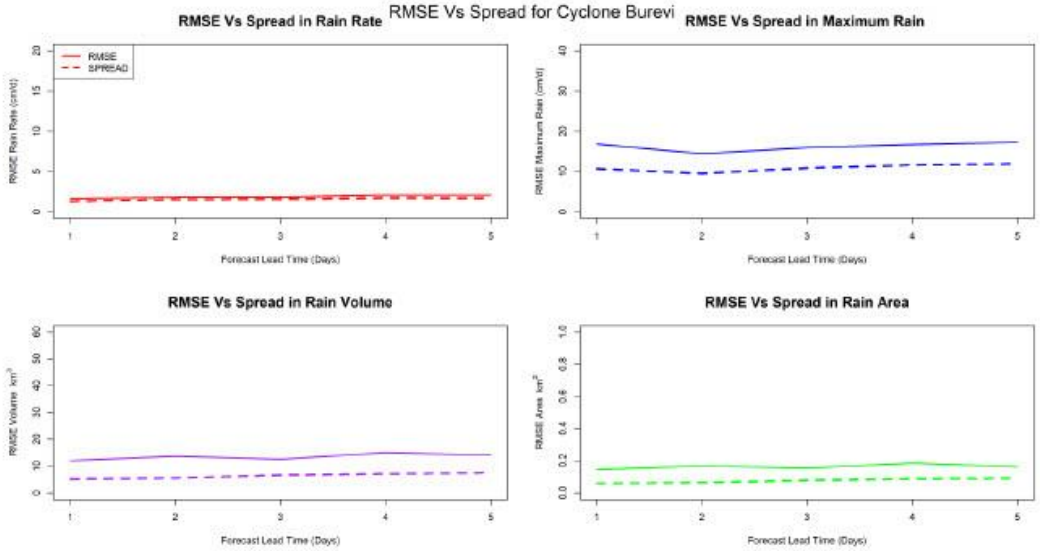


Figure 25. Spread Vs Bias in the 40 mm/day object parameters (I) Rain rate (ii) Maximum Rain (iii) Rain Volume (iv) Rain area for the case of Bay of Bengal tropical cyclone Nivar during 22-27 Nov 2020.

The comparison of RMSE and Spread for rainfall attributes in the case of Cyclone Burevi are presented in Figure 25. For both rain rate and area the match between the RMSE and Spread is the highest. For maximum rainfall and volume the match between RMSE and Spread is almost similar. However, both RMSE and Spread are seen to be higher for maximum rainfall. The RMSE is seen to be the least for rainfall area as seen in the case of Cyclone Nivar. The EPS system is seen to be under dispersive for all attributes.

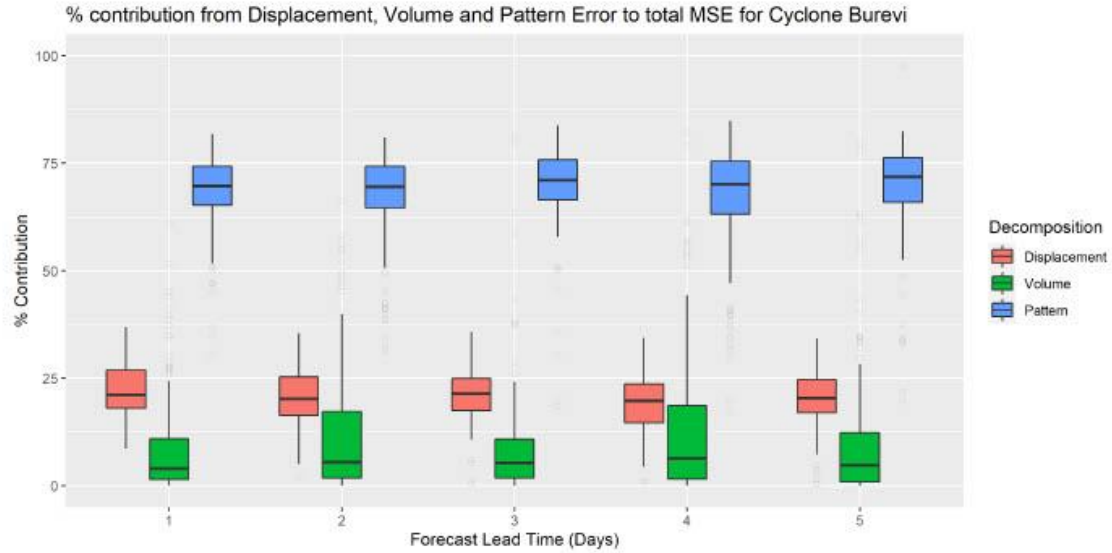


Figure 26. Box and whisker plot comparing the percentage contribution from errors in displacement, volume and pattern to the total mean squared error (MSE) in forecast 24 hr rainfall across the ensemble for Cyclone Burevi during 29 Nov–04 Dec 2020 at a range of lead times.

The percentage contribution from displacement, volume and pattern to the total MSE is shown on Figure 26. As in the previous case the maximum contribution to the total error is from the pattern error; the volume contributes the least followed by displacement error. In this case the median values for contribution from volume vary from 5% in Day-1 forecast to approximately 7% in Day-5. The percentage contribution from displacement is very similar for both Nivar and Burevi cyclones, but there is more spread in the displacement values between members in the Burevi case.

4.3.3. Cyclone Tauktae

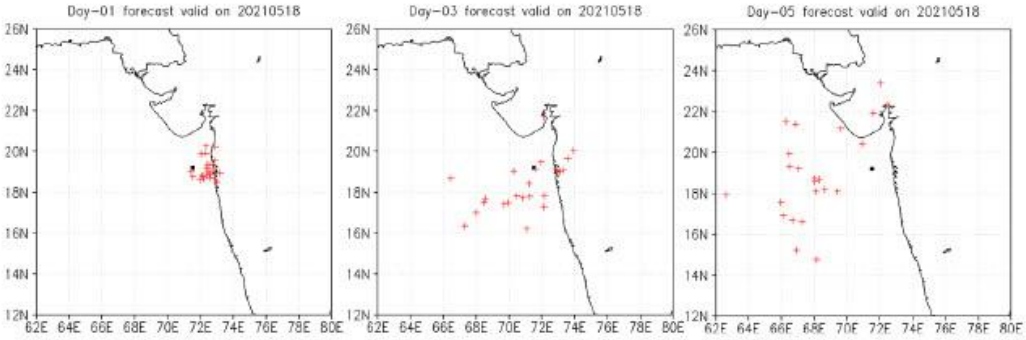


Figure 27. Centre of gravity (COG) in the 40mm/day CRA object in Observed (black) and ensemble member forecasts valid on 18th May 2021 for the case of Arabian Sea tropical cyclone Tauktae.

For the case of 'Tauktae', a comparison of observed and forecast COG is presented in Figure 27 from Day-1, Day-3, and Day-5 forecasts. In Day-1, the COGs from ensemble members are closely clustered close to the COG from the observed object. Again the members seem to lag behind the observed COG in Day-1. In the Day-3 and Day-5 forecasts, the spread among the members is higher, particularly in a SW to NE direction. Compared to the two earlier cases discussed in Figure 21 and 24, the Day-3 and Day-5 forecasts in Tauktae show larger spread in COGs, with the observed COG being closer to the edge of the ensemble COG spread.

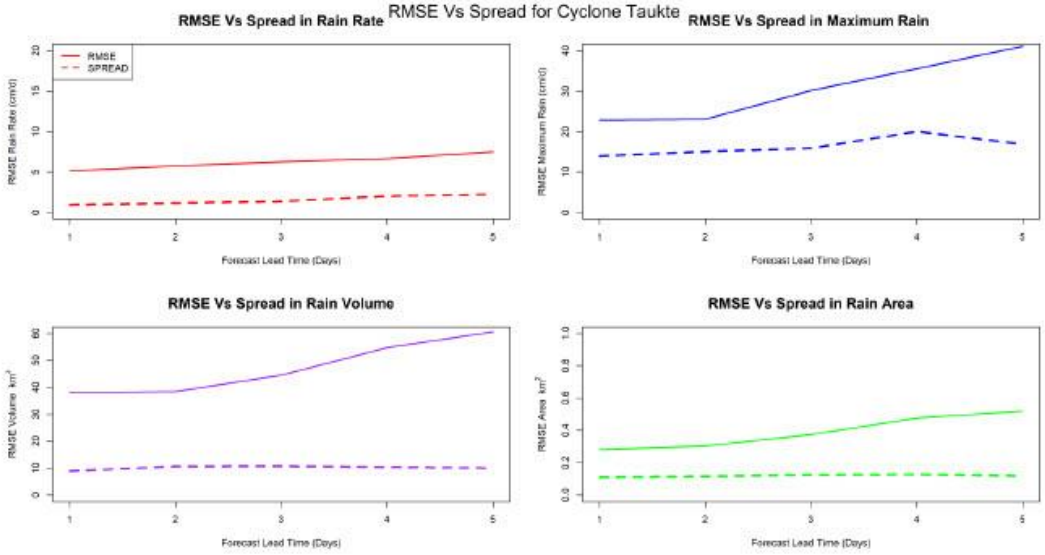


Figure 28 Spread Vs Bias in the 40 mm/day object parameters (I) Rain rate (ii) Maximum Rain (iii) Rain Volume (iv) Rain area for the case of Arabian Sea tropical cyclone Tauktae during 14-19 May 2021.

For Cyclone Tauktae the comparison between RMSE and Spread for rainfall attributes like rain rate, maximum, volume and area of rainfall is shown in Figure 28. The highest difference between the RMSE and Spread is seen for rainfall volume and maximum rainfall, with this difference increasing with lead time. The highest values of RMSE are seen in the rainfall volume where the day-1 RMSE is approximately 40 km³. The least RMSE is seen in the case of rainfall area which also shows the highest match between RMSE and Spread as compared to the other attributes.

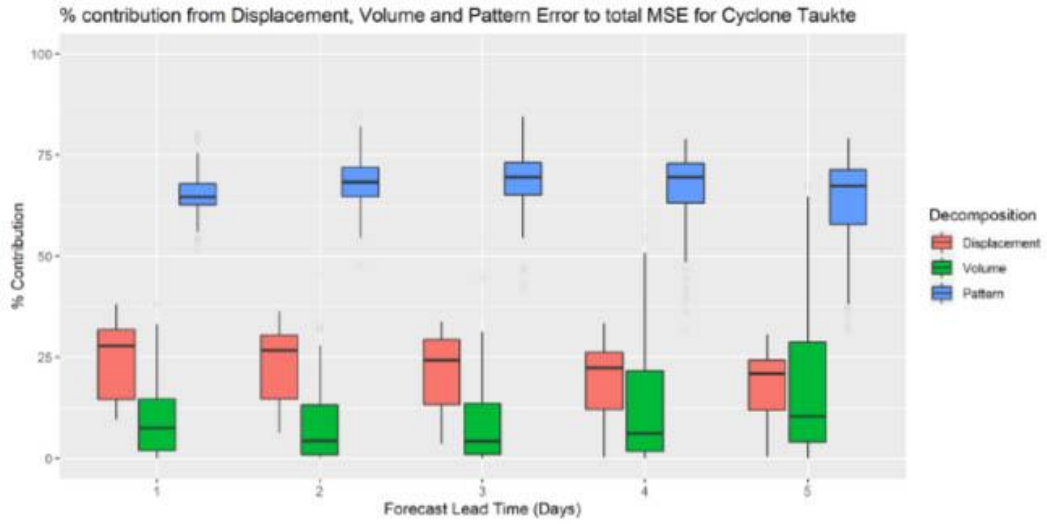


Figure 29. Box and whisker plot comparing the percentage contribution from errors in displacement, volume and pattern to the total mean squared error (MSE) in forecast 24 hr rainfall across the ensemble for Cyclone Tauktae during 14-19 May 2021 at a range of lead times.

The percentage contribution from the components (displacement, volume and pattern) to the total MSE for Cyclone Tauktae is shown in Figure 29. As in the previous case the pattern error is seen to have the maximum contribution followed by displacement and volume shows the least contribution. The contribution from volume error is seen to be increasing with lead time. However, the contribution from displacement error decreases with lead time.

4.3.4. Cyclone Gulab

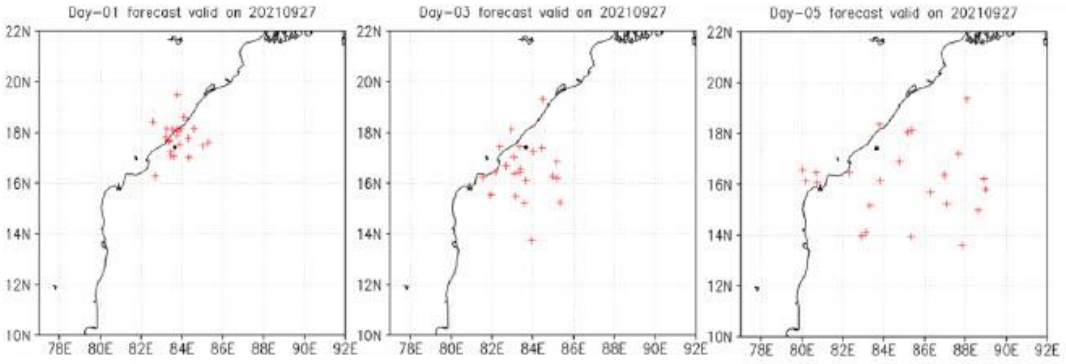


Figure 30. Centre of gravity (COG) in the 40mm/day CRA object in Observed (black) and ensemble member forecasts valid on 27th Sep 2021 for the case of Bay of Bengal tropical cyclone Gulab.

For the case of 'Gulab', a comparison of observed and forecast COG is presented in Figure 30 for Day-1, Day-3, and Day-5 forecasts. In Day-1, the COGs from ensemble members are closely clustered close to the COG from the observed object. In Day-3 and Day-5 forecasts, the spread among the members is higher, but the observed COG remains within the extent of COGs of ensemble members.

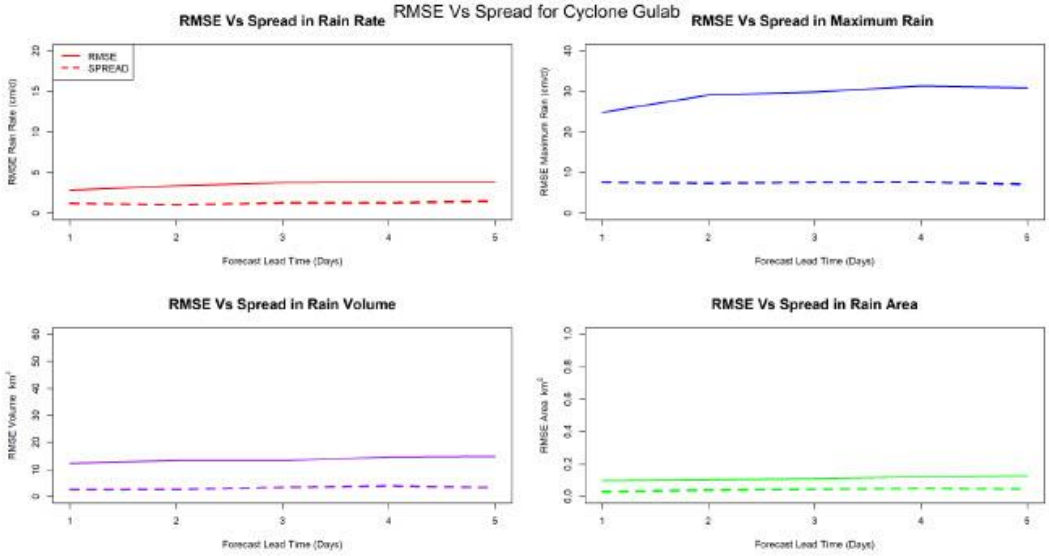


Figure 31. Spread Vs Bias in the 40 mm/day object parameters (I) Rain rate (ii) Maximum Rain (iii) Rain Volume (iv) Rain area for the case of Bay of Bengal tropical cyclone Gulab during 25-28 Sep 2021.

The error/spread comparison for each of the rainfall attributes in the case of Cyclone Gulab is plotted in Figure 31. The highest difference between RMSE and Spread is seen for rainfall maximum which also shows the highest values of RMSE and the Spread among the ensemble members. The rainfall volume shows the second highest RMSE values which is followed by rain rate. Rainfall area shows the least RMSE and the highest match between RMSE and Spread.

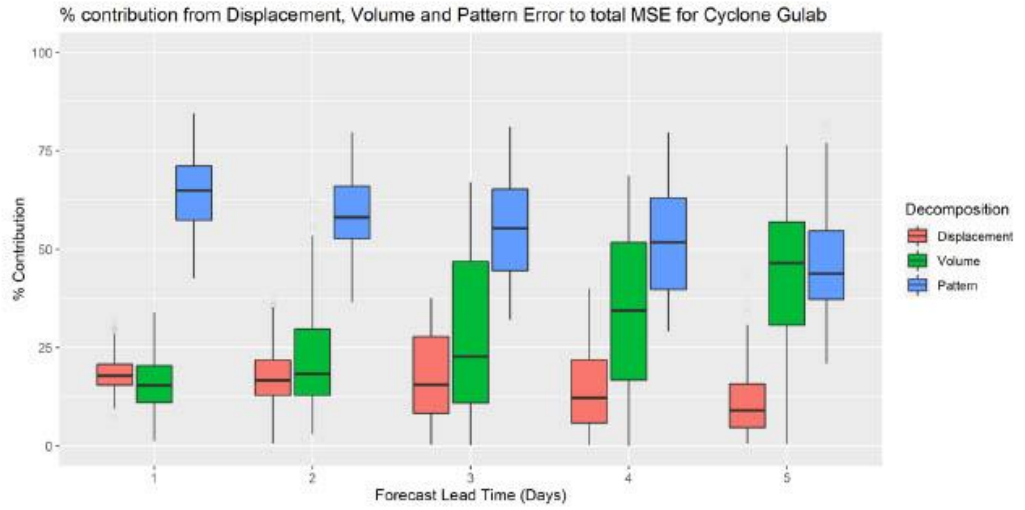


Figure 32. Box and whisker plot comparing the percentage contribution from errors in displacement, volume and pattern to the total mean squared error (MSE) in forecast 24 hr rainfall across the ensemble for Cyclone Gulab during 25-28 Sep 2021 at a range of lead times.

For Cyclone Gulab, the percentage contribution from the displacement, pattern and volume is shown in Figure 32. In this case for smaller lead times (Day-1 to Day-4) the maximum contribution to the total MSE is seen from the pattern error. However, for the Day-5 forecasts the contribution from both volume and pattern is comparable. Also, the contribution from volume error is higher than displacement (except for Day-1) for this cyclone case. The contribution from volume error is seen to be increasing with lead time, whereas from displacement and pattern the percentage contribution is seen to be decreasing with lead time.

4.3.5. Summary

The decomposition of the MSE rainfall errors using CRA verification for each of the four cyclone cases consistently show a large contribution (~75%) to total error from pattern error; contribution from displacement and volume errors are lower (~ 25%). The RMSE Vs Spread assessed for the four object parameters indicate forecasts are under-dispersed, particularly for the maximum rain and rain volume.

The CRPS generalizes the mean absolute error; in fact, it reduces to the mean absolute error (MAE) if the forecast is deterministic. The CRPS was calculated for rainfall attributes within the 40 mm/day object for the 4 cyclone cases considered in this study. Figure 33 (a-d) shows day 1 to 5 CRPS and the MAE for the rainfall attributes (rain rate, maximum, volume and area) for the 4 cyclone cases Gulab, Tauktae, Nivar and Burevi. The MAE is calculated for the ensemble mean and it is seen from the panels that the MAE in the ensemble mean is always higher than the CRPS. Probabilistic forecasts involving Rainfall area and Rain rate followed by Maximum Rainfall exhibit the lowest CRPS and MAE among all the attributes. This implies that the accuracy of the model in predicting these variables is relatively higher. Rainfall volume shows the highest MAE and CRPS values which shows that the model is less skillful in predicting the volume of rainfall for these cyclone cases. As can be expected, both the CRPS and the MAE values are seen to be increasing with time, which indicate a decay in the skill with lead time.

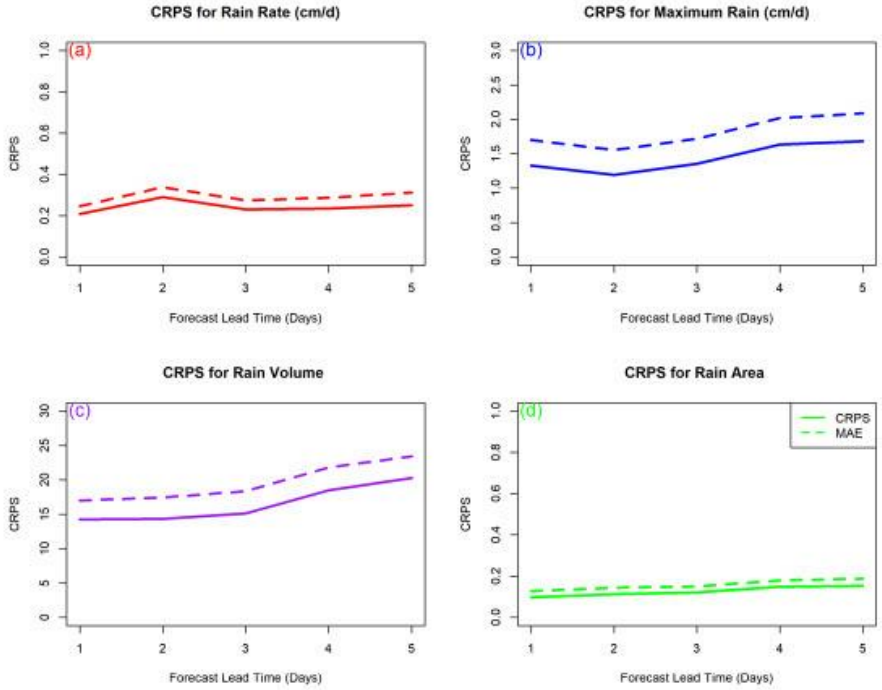


Figure 33. CRPS in the 40 mm/day object parameters (a) Rain rate (b) Maximum Rain (c) Rain Volume (d) Rain area for all the four cases of TCs 'Nivar', 'Burevi', 'Tauktae' and 'Gulab' during 2020 and 2021.

5. Summary, discussion, and recommendations for future work

This study has applied three object-based verification methods to evaluate the forecasts of TC-related rainfall in four recent landfalling cyclones in India. The code to run SAL and MODE were already available (with MODE run via METplus), although new wrapper scripts and python code for plotting the output were developed for this study. The CRA code was previously only available in IDL and has been re-written in R at NCMRWF as part of this study (see documentation of cRa and excRa in Appendix 1). This forms a valuable new resource for precipitation verification more generally in addition to its use for TC rainfall verification.

The three methods have provided useful complementary information on the ability of the ensemble models to forecast TC rainfall. Although box-and-whisker plots are created for each method they show different things in each: In SAL, the box plot displays the scores after comparing with observations (with zero being optimum); for MODE it is the object attribute distribution with the observed values overlain as stars; and for CRA it is the percentage contribution of each component to the total error. The rainfall thresholds and accumulation periods chosen in CRA and MODE are also different. Therefore, they cannot be directly compared, and each method also has a unique approach to defining the objects and calculating the object attributes. A summary of the findings for each of the four cyclones across the three methods is as follows:

- For **Cyclone Nivar**, all methods show that rainfall volumes were generally underestimated for both daily and storm-total accumulations, apart from at the shortest lead times. SAL shows ECMWF ENS has object structures that are too large and flat prior to applying a land sea mask. However, over land the volumes and structure were better estimated by ECMWF ENS than by either of the UM-based ensembles. The MODE and CRA error and spread plots indicated that the models were under-dispersed for the intensity and volume-related attributes, but for rainfall object area the two were better matched, and for centroid location it varied from model to model. The pattern error was the biggest portion of the CRA error.
- For **Cyclone Burevi**, there is less predictability at longer lead times and more spatial variability in the object centroids. There is a low bias in intensity, especially from ECMWF members. MODE shows the models are under-dispersed for the intensity attributes and the area, with MOGREPS-G having the best match. CRA reveals the NEPS-G forecast to be under-dispersive for all attributes. For centroid location the results are more complex, but over-dispersed for MOGREPS-G and NEPS-G. Again, the signal changed in the storm total rainfall verification once the land/sea mask is applied, with forecast structures closer to observed in MOGREPS-G and NEPS-G when rainfall over sea is included, but better in the ECMWF ENS once the land/sea mask has been applied. The maximum contribution to the total error is from the pattern error; the volume contributes the least followed by displacement error.
- For **Cyclone Tauktae**, many ensemble members fail to capture the cyclone evolution at longer leads, so the rainfall is also less predictable. MODE reveals wide ranges on the area exceeding the threshold within the objects, indicating the objects are all quite different. SAL also shows a notably large spread in the Structure and Amplitude components for the 24-hour rainfall objects, which only reduces a little at shorter lead times. There is a low bias in intensity, especially from ECMWF ENS, and all models have a marked drop-off in storm-total rainfall with lead time, being especially noticeable for NEPS-G. Rainfall over land is overestimated at short lead times but underestimated at longer lead times. SAL structure scores show forecast objects mainly too large/flat overall, and best for NEPS-G, but the forecast structures become too sharp and small when only rain over land is included. MODE and CRA show the models to be under-dispersed for the intensity measures and the area for all models, with this difference increasing with lead time. For centroid location the model is under dispersed for the UM-based models, but better matched for ECMWF ENS. CRA shows the pattern error is the largest contribution to overall error, with the contribution from volume error increasing with lead time.
- For **Cyclone Gulab**, the storm-total intensities are under-predicted across models, with MODE 90th percentile values about half of observed. Storm total rainfall amounts correspond better between forecasts and observations before applying a land/sea mask, but the storm total rainfall over land is noticeably lower

compared to observations; the area exceeding the threshold is also too small in all three models. SAL reveals little bias and a large spread in the structure of the 24-hour rainfall forecasts, a general underestimation of the rainfall amounts that improves at shorter lead times, and lower spread in the location component which is closer to zero than in the other three cases. The contrast between the storm total structure results before and after the application of the land/sea mask occur again, being too large/flat overall, but too sharp/small when only rainfall over land is examined, but the overall estimation of rainfall amounts is similar in both. The MODE and CRA error/spread plots show the models are generally under-dispersed for the intensity, volume and area for all models, and particularly for intensity. The CRA error analysis is interesting in this case, with the contribution from volume error being higher than from displacement and increasing with lead time.

Across the cases the rainfall intensities and volumes are generally too low for all three models for both 24-hour and storm total rainfall. The rainfall object structures of ECMWF ENS are generally larger and flatter than the UM-based ensembles, which are smaller and sharper, and over the ocean are close to the observations. However, ECMWF ENS appears to better capture rainfall over land, which is crucial to forecasting flooding impacts. The problem of the UM-based ensembles underestimating rainfall over land and not bringing rainfall far enough inland in India has been observed previously at NCMRWF and is worthy of further investigation. The error/spread relationship is consistently better for the rainfall object area than for the other attributes, with volume and intensity being generally under-dispersed.

We find that each method has its own particular strengths. The 'S' component of the SAL method is useful in looking specifically at the structure of the rainfall objects to assess whether the objects are too large and flat or too small and sharp. This highlighted key differences in the character of the rainfall forecasts between the ensembles, with the UM-based ensembles often having smaller/sharper objects than the ECMWF ENS forecasts. The MODE method returns the objects to allow a visual inspection (as in Figure 9), which is very useful to give a visual context to the verification results. The object centroid maps are useful to compare the locations, and colouring by another attribute allows for a visual examination of the relationships between the attributes (Figure 10). The unique aspect of the CRA method is the splitting of the error into the three different components, so that you can easily see the main causes of the error and how this differs with lead times and model. The box-and-whisker plots for the attributes from all three methods provided an effective way of summarising the forecasts across multiple lead times, models and accumulation periods. Analysing the spread and error of the object attributes in the MODE and CRA methods provided a good and relatively simple way to quantify ensemble reliability at predicting each of the attributes of the rainfall objects. Postage stamp plots were created for all storms, models and lead times which although not included in this report for space reasons (other than in Figure 8), were very useful in investigating in subjective detail the reason behind the objective results.

One common problem when applying all three methods was how to ensure that only the heavy rainfall from a TC was included in the results. Using a fixed storm domain approach, as in the MODE verification, can lead to some non-TC origin rainfall being included in the comparison, so it is important to restrict the area the objects are analysed over to a box closely following the tropical cyclone track to minimise. Using an observed TC track mask, as in the SAL verification, minimises the interference from non-TC rainfall objects, but larger track errors at longer lead times can mean that part of the forecast rainfall objects fall outside of the unmasked region. Masking by forecast track would not be possible as that would discount rainfall from tropical depressions which didn't meet the tracking threshold but should still be considered. Perhaps the optimum way to do this in the future would be to mask by observed TC track but increase the distance threshold used in this masking to take account of forecast track error.

There are many possible options on how this work could be further developed in the future:

- The newly developed cRa and excRa code to apply the CRA method could be implemented on JASMIN and/or at the Met Office. This would enable a comparison between NEPS-G, MOGREPS-G and ECMWF ENS data in a similar way as has been carried out for the other two methods.

- A more detailed study on a particular chosen storm, using consistent thresholds and accumulation periods, could allow for a more detailed intercomparison of the MODE/CRA attributes.
- One way of continuing the momentum of this work would be to implement object-based rainfall verification in near real time on TC cases. NCMRWF already plan to run the new CRA software in near real time for Indian cyclones using NEPS-G. If the new CRA code can be implemented at the Met Office it could potentially be also run on MOGREPS-G and ECMWF ENS rainfall forecasts, which could build up a useful case archive.
- The SAL verification could be expanded to calculate the S, A and L component scores for member-member pairs to give a measure of ensemble spread for each of the attributes and allow us to compare whether there is more difference between the ensemble and the observations than there is within the ensemble itself (following the example of McCabe *et al.* (2020) for ensemble forecasts over the UK).
- More analysis into systematic errors arising from the rainfall object shape and orientation from MODE could be done. The ability to analyse the ensemble information as probabilities is available and could be explored. Investigation into differences in object characteristics over sea or land could be completed. MODE Time-Domain could be used to investigate whether there are issues in the storm evolution from the ensemble members, for example storm duration or rainfall object centroid positions or problems with the evolution from a particular initialisation.
- If sufficient TC cases were run to have a bigger dataset, the ability of the ensemble to predict the probability of the MODE and CRA object attributes exceeding a given threshold could be assessed using other useful ensemble metrics such as the reliability of the object attributes, the ROC area, or the Brier Skill Score. This would follow the example given in Dube *et al.* (2022) for forecasts of monsoon rainfall over India using CRA.
- Additional spatial verification methods could be applied to ensemble TC rainfall to see if they can add useful information to the object-based methods that we have considered. For example, the dispersive Fraction Skill Score (dFSS) and ensemble Fraction Skill Score (eFSS) could be utilised to measure the spread and error of each ensemble for tropical cyclone rainfall.
- It would also be interesting to explore if these methods could be further tailored, or other methods applied, to look beyond the overall TC rainfall object to delve more into the detailed ability of the models to predict specific aspects of the TC rainfall, for example the heavy rain around the core and the rainfall associated with outer rain bands.
- This study had a particular focus on landfalling Indian cyclones. It would be interesting to look at applying these methods to ensemble forecasts of TC rainfall in other basins, to see whether the findings are similar or are specific to the north Indian ocean basin.

References

Barde, V., Nageswararao, M.M., Mohanty, U.C., Panda, R.K., Ramadas, M. (2020): Characteristics of southwest summer monsoon rainfall events over East India. Theor Appl Climatol, 141, 1511–1528, https://doi.org/10.1007/s00704-020-03251-y

Chen, Y., Ebert, E.E., Davidson, N.E., and Walsh, K.J.E. (2018): Application of contiguous rain area (CRA) methods to tropical cyclone rainfall forecast verification. Earth and Space Science, 5, 736–752, https://doi.org/10.1029/2018EA000412

Davis, C.A., Brown B., and Bullock R. (2006): Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. Mon. Wea. Rev., 134, 1772–1784, https://doi.org/10.1175/MWR3145.1

Davis, C.A., Brown B., Bullock R., and Halley Gotway J. (2009): The Method for Object-Based Diagnostic Evaluation (MODE) applied to numerical forecasts from the 2005 NSSL/SPC Spring Program. Wea. Forecasting, 24, 1252–1267, https://doi.org/10.1175/2009WAF2222241.1

Dube, A., Karunasagar, S., Ashrit, R. and Mitra, A.K. (2022): Spatial verification of ensemble rainfall forecasts over India, Atmospheric Research, 273, https://doi.org/10.1016/j.atmosres.2022.106169

Ebert, E., and McBride, J. (2000): Verification of precipitation in weather systems: Determination of systematic errors. Journal of Hydrology, 239(1–4), 179–202, https://doi.org/10.1016/S0022-1694(00)00343-7

Ebert, E., L. J. Wilson, B. G. Brown, P. Nurmi, H. E. Brooks, J. Bally, and M. Jaeneke, 2004: Verification of nowcasts from the WWRP Sydney 2000 Forecast Demonstration Project. Wea. Forecasting, 19, 73–96

Ebert, E. E., and Gallus, W. A. (2009): Toward better understanding of the contiguous rain area (CRA) method for spatial forecast verification. Weather and Forecasting, 24(5), 1401–1415. https://doi.org/10.1175/2009WAF2222252.1

Fortin, V., Abaza, M., Anctil, F., & Turcotte, R. (2014): Why Should Ensemble Spread Match the RMSE of the Ensemble Mean? Journal of Hydrometeorology, 15(4), 1708–1713. https://doi.org/10.1175/JHM-D-14-0008.1

Gallus, W. A., Jr. (2010): Application of Object-Based Verification Techniques to Ensemble Precipitation Forecasts, Weather and Forecasting, 25(1), 144–158. https://doi.org/10.1175/2009WAF2222274.1

Grams, J. S., W. A. Gallus, L. S. Wharton, S. Koch, A. Loughe, and E. E. Ebert, 2006: The use of a modified Ebert–McBride technique to evaluate mesoscale model QPF as a function of convective system morphology during IHOP 2002. Wea. Forecasting, 21, 288–306.

Hersbach, H. (2000): Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather and Forecasting, 15(5), 559-570.

Ji, L., Zhi, X., Simmer, C., Zhu, S., and Ji, Y. (2020): Multimodel Ensemble Forecasts of Precipitation Based on an Object-Based Diagnostic Evaluation, Monthly Weather Review, 148(6), 2591-2606. https://doi.org/10.1175/MWR-D-19-0266.1

Jiang, H., Liu, C. and Zipser, E.J. (2011): A TRMM-based tropical cyclone cloud and precipitation feature database. J. Appl. Meteor. Climatol., 50, 1255–1274, https://doi.org/10.1175/2011JAMC2662.1.

Knapp, K.R., Kruk, M.C. Levinson, D.H., Diamond, H.J. and Neumann, C.J. (2010): The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone best track data. Bulletin of the American Meteorological Society, 91, 363-376. https://doi.org/10.1175/2009BAMS2755.1

Knapp, K.R., Diamond, H.J., Kossin, J.P., Kruk, M.C. and Schreck, C.J. (2018): International Best Track Archive for Climate Stewardship (IBTrACS) Project, Version 4. NOAA National Centers for Environmental Information. https://doi.org/10.25921/82ty-9e16 (accessed 9th August 2022).

Leutbecher, M. (2010): Diagnosis of ensemble forecasting systems. Seminar on Diagnosis of Forecasting and Data Assimilation Systems, 7 - 10 September 2009. Accessible at:

https://www.ecmwf.int/sites/default/files/elibrary/2010/10725-diagnosis-ensemble-forecasting-systems.pdf

Luitel, B., Villarini, G. and Vecchi, G.A. (2018): Verification of the skill of numerical weather prediction models in forecasting rainfall from U.S. landfalling tropical cyclones. J. Hydrol., 556, 1026–1037, https://doi.org/10.1016/j.jhydrol.2016.09.019.

McCabe, A., Porson, A., Roberts, N., Mittermaier, M., Flack, D., Webster, S., Willington, S., Walters, D. and Bush, M. (2020): Evaluation of the different sources of uncertainty in MOGREPS-UK. Met Office internal report.

Marchok, T., Rogers, R. and Tuleya, R. (2007): Validation schemes for tropical cyclone quantitative precipitation forecasts: evaluation of operational models for U.S. landfalling cases. Weather and Forecasting, 22, 726–746, https://doi.org/10.1175/WAF1024.1

Mitra, A.K., Bohra, A.K., Rajeevan, M.N., and Krishnamurti, T.N., (2009): Daily Indian precipitation analysis formed from a merge of rain-gauge data with the TRMM TMPA satellite-derived rainfall estimates. J. Meteorol. Soc. Jpn.,87(A), pp. 265–279. https://doi.org/10.2151/jmsj.87A.265

Mitra, A.K., Momin, I.M., Rajagopal, E.N., Basu, S., Rajeevan, M.N., and Krishnamurti, T.N. (2013): Gridded daily Indian monsoon rainfall for 14 seasons: merged TRMM and IMD gauge analyzed values. J. Earth Syst. Sci.,122(5) (2013), pp. 1173–1182. https://doi.org/10.1007/s12040-013-0338-3

Prat, O.P., and Nelson, B.R. (2013): Mapping the world's tropical cyclone rainfall contribution over land using the TRMM Multi-satellite Precipitation Analysis. Water Resour. Res., 49, 7236–7254, https://doi.org/10.1002/wrcr.20527.

Prat, O.P., and Nelson, B.R. (2016): On the link between tropical cyclones and daily rainfall extremes derived from global satellite observations. J. Climate, 29, 6127–6135, https://doi.org/10.1175/JCLI-D-16-0289.1.

Radanovics, S., Vidal, J., and Sauquet, E. (2018): Spatial Verification of Ensemble Precipitation: An Ensemble Version of SAL, Weather and Forecasting, 33(4), 1001-1020, https://doi.org/10.1175/WAF-D-17-0162.1

Skamarock, W. C. (2004). Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra, Monthly Weather Review, 132(12), 3019-3032. https://doi.org/10.1175/MWR2830.1

Tartaglione, N., S. Mariani, C. Accadia, A. Speranza, and M. Casaioli, 2005: Comparison of rain gauge observations with modeled precipitation over Cyprus using contiguous rain area analysis. Atmos. Chem. Phys., 5, 2147–2154.

Titley, H.A., Cloke, H.L., Harrigan, S., Pappenberger, F., Prudhomme, C., Robbins, J.C., Stephens, E. M., and Zsótér, E. (2021). Key Factors Influencing the Severity of Fluvial Flood Hazard from Tropical Cyclones, *Journal of Hydrometeorology*, *22*(7), 1801-1817, https://doi.org/10.1175/JHM-D-20-0250.1.

Wernli, H., Paulat, M., Hagen, M., and Frei, C. (2008): SAL - A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts. Monthly Weather Review, 136, 4470–4487, https://doi.org/10.1175/2008MWR2415.1

Wernli, H., Hofmann, C., and Zimmer, M. (2009): Spatial Forecast Verification Methods Intercomparison Project: Application of the SAL Technique, Weather Forecasting, 24, 1472–1484, https://doi.org/10.1175/2009WAF2222271.1

Wilks, D.S. (2011): Statistical Methods in the Atmospheric Sciences. 3rd Edition, Academic Press, Oxford.

Yu, Z., Chen, Y.J., Ebert, B., Davidson, N.E., Xiao, Y., Yu, H., and Duan, Y. (2020): Benchmark rainfall verification of landfall tropical cyclone forecasts by operational ACCESS-TC over China. Meteorol Appl. 2020; 27:e1842. https://doi.org/10.1002/met.1842

Zamo, M. and Naveau, P. (2018): Estimation of the continuous ranked probability score with limited information and applications to ensemble weather forecasts. Mathematical Geosciences, 50(2), 209-234.

Appendix 1: Development of 'cRa and excRa'

The development of CRA software using R (cRa) and its extension for spatial verification of ensemble model forecasts (excRa) has been undertaken at NCMRWF. The original CRA software developed by Beth Ebert at CAWCR (https://cawcr.gov.au/projects/verification/CRA/CRA_verification.html) is used to carry out spatial verification of any observation-forecast pair. The code is very well documented and helpful for advanced users who wish to make changes to make the verification relevant for their regions e.g. defining geographical domains, rainfall thresholds etc. The original CRA code is written in IDL (https://www.l3harrisgeospatial.com/Software-Technology/IDL), which is not freely available to all users and therefore limits the uptake of the method.

A new CRA code using R has been developed and tested at NCMRWF. R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. The new R code is named 'cRa' since it is fully developed based on the original ideas implemented in IDL. The 'cRa' code makes extensive use of the 'raster' package of R. The 'raster package provides classes and functions to manipulate geographic (spatial) data in 'raster' format. Raster data divides the world into cells (rectangles; pixels) of equal size (in units of the coordinate reference system). Such data are also referred to as 'grid' data. The package is particularly useful when using very large datasets that cannot be loaded into memory. The package has several functionalities of basic computations, raster arithmetic, raster algebra, spatial contextual computations, model prediction etc., (https://www.rdocumentation.org/packages/raster/versions/1.3-11/topics/raster-package). Additionally, for graphics and visualization 'plotly' have been used.

An extension of the tool for ensemble forecasts 'excRa' has also been developed which allows CRA verification to be applied to forecasts from ensemble members. The code 'excRa' can be executed serially or in parallel on a multi-core system. In addition to the 'raster' package used in 'cRa', 'excRa' uses 'foreach', 'doParallel' and doMC packages. The `foreach` package provides a new looping construct for executing R code repeatedly. The main reason for using the `foreach` package is that it supports *parallel execution*, that is, it can execute those repeated operations on multiple processors/cores, or on multiple nodes of a cluster. If each operation takes over a minute, and you want to execute it hundreds of times, the overall runtime can take hours. But using `foreach`, that operation can be executed in parallel on hundreds of processors on a cluster, reducing the execution time back down to minutes. Much of parallel computing comes to doing three steps (i) splitting the problem into pieces, (ii) executing the pieces in parallel, and (iii) combining the results back together. Using the `foreach` package, the iterators help you to split the problem into pieces, the `%dopar%` function executes the pieces in parallel, and the specified `.combine` function puts the results back together. The doMC package provides a parallel backend for the foreach `%dopar%` function using the multicore functionality of the parallel package. The doParallel package is a "parallel backend" for the foreach package. It provides a mechanism needed to execute foreach loops in parallel. The foreach package must be used in conjunction with a package such as doParallel in order to execute code in parallel. The user must register a parallel backend to use, otherwise foreach will execute tasks sequentially, even when the `%dopar%` operator is used.

As an example, the case of Day-1 forecast rainfall valid on 27th Nov 2020 is verified using excRa in Figure A1 & A2. Observed rainfall used for verification is the 0.25° x 0.25° resolution NMSG data discussed in section 2.3 and Table 2. The observed and Day-1 forecast rainfall over the eastern cost of India is shown in Figure A1.

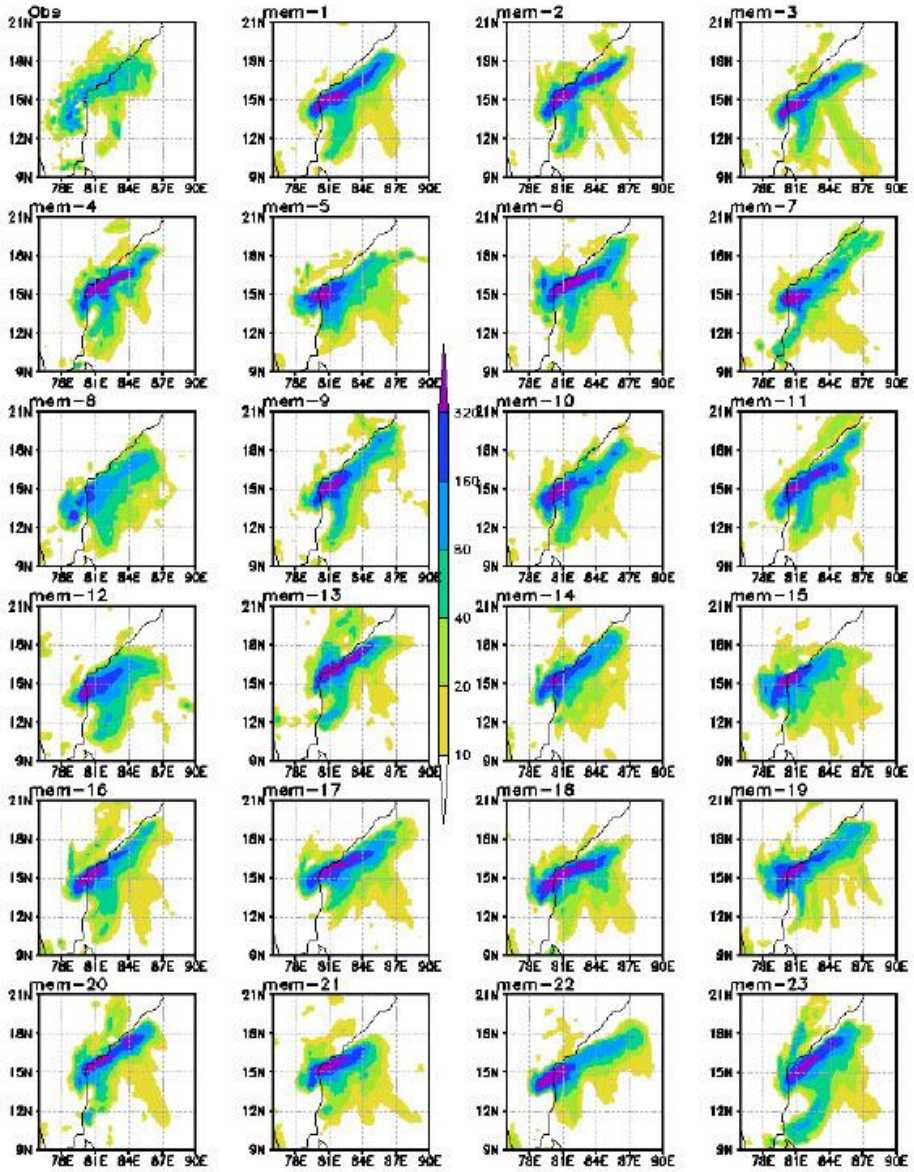


Figure A1 Observed and Day-1 forecast rainfall valid on 27th Nov 2020 near the eastern coast of India.

The observed rainfall of 27th Nov 2020 is partly distributed over land and sea as Nivar cyclone tracked close to the coast. The NEPS-G members also capture this feature with some variability in spatial extent among members. All members consistently show much higher peak rainfall amounts. The differences in spatial coverage among the members is remarkable. It is also evident that the forecast rainfall patterns are displaced slightly to the east relative to the observations.

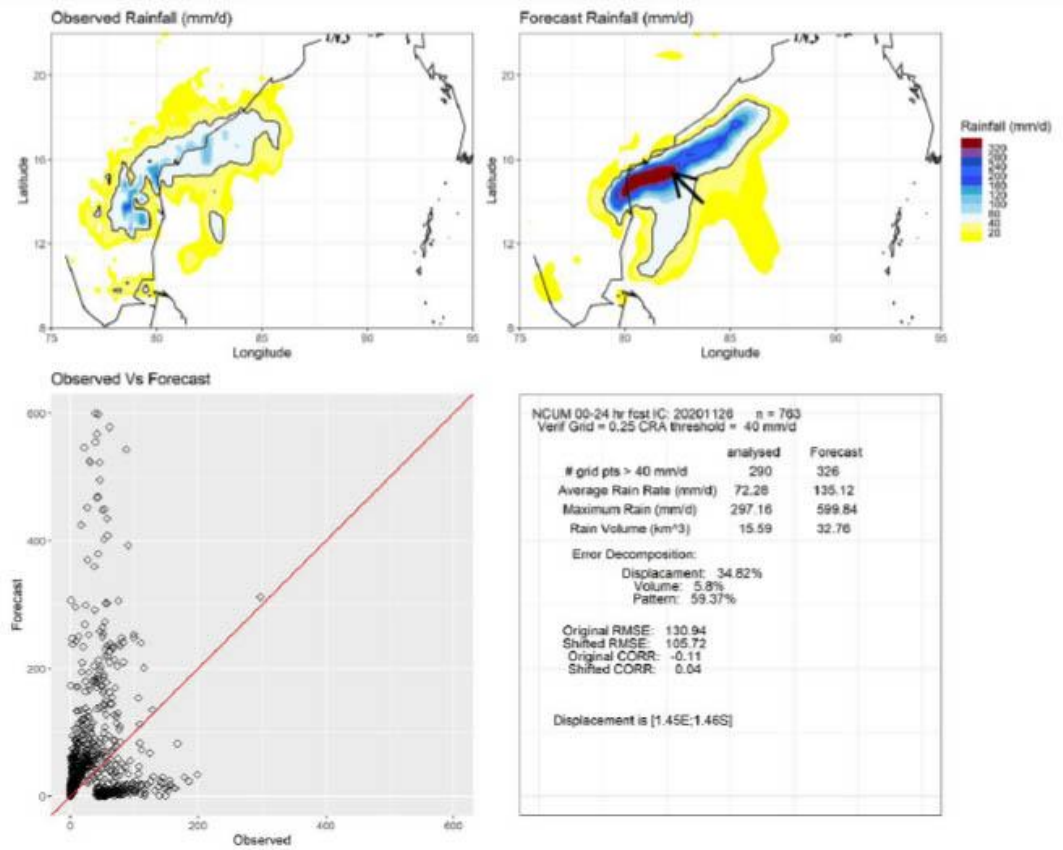


Figure A2 CRA verification applied to NEPS-G (CTL) member. Observed and Day-1 forecast rainfall valid on 27th Nov 2020 near the eastern coast of India. CRA results correspond to 40 mm/day object.

The CRA verification of Day-1 forecast rainfall in control member is shown in Figure A2. The top panels show the spatial distribution of the observed and forecast rainfall outlining the 40 mm/day CRA threshold. Forecast rainfall shows higher rainfall amounts. The displacement of the 40 mm/day CRA object is indicated by arrow. The CRA stats are shown in the bottom. The scatter plot shows observed and forecast rainfall at different points clearly indicating overestimation by a huge margin. Other stats shown include the decomposition of total error in terms of displacement, patter and volume. Contribution from pattern error is highest (~60%) followed by displacement (~35%) and volume (~6%).

A8: Published Article: Emergency flood bulletins for Cyclones Idai and Kenneth: A critical evaluation of the use of global flood forecasts for international humanitarian preparedness and response

This appendix contains the published version of the following paper (summarised in Section 7.5.3):

Emerton, R., Cloke, H., Ficchi, A., Hawker, L., de Wit, S., Speight, L., Prudhomme, C., Rundell, P., West, R., Neal, J., Cuna, J., Harrigan, S., Titley, H., Magnusson, L., Pappenberger, F., Klingaman, N. and Stephens, E. (2020) Emergency flood bulletins for Cyclones Idai and Kenneth: A critical evaluation of the use of global flood forecasts for international humanitarian preparedness and response. *International Journal of Disaster Risk Reduction*, 50: 101811. doi:10.1016/j.ijdrr.2020.101811.

International Journal of Disaster Risk Reduction 50 (2020) 101811

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

journal homepage: http://www.elsevier.com/locate/ijdrr

Emergency flood bulletins for Cyclones Idai and Kenneth: A critical evaluation of the use of global flood forecasts for international humanitarian preparedness and response

Rebecca Emerton [a,b,*], Hannah Cloke [c,d,e], Andrea Ficchi [c,b], Laurence Hawker [f], Sara de Wit [g], Linda Speight [c], Christel Prudhomme [b,h,i], Philip Rundell [j], Rosalind West [j], Jeffrey Neal [f,k], Joaquim Cuna [l], Shaun Harrigan [b], Helen Titley [m,c], Linus Magnusson [b], Florian Pappenberger [b], Nicholas Klingaman [a], Elisabeth Stephens [c]

[a] National Centre for Atmospheric Science, University of Reading, UK
[b] European Centre for Medium-Range Weather Forecasts, Reading, UK
[c] University of Reading, UK
[d] Uppsala University, Sweden
[e] Centre of Natural Hazards and Disaster Science, Sweden
[f] University of Bristol, UK
[g] University of Oxford, UK
[h] University of Loughborough, UK
[i] Centre for Ecology and Hydrology, Wallingford, UK
[j] Department for International Development, UK
[k] Fathom, Bristol, UK
[l] Technical University of Mozambique, Maputo, Mozambique
[m] Met Office, Exeter, UK

ARTICLE INFO

Keywords:
Tropical cyclone
Flood
Forecasts
Bulletins
Early action

ABSTRACT

Humanitarian disasters such as Typhoon Haiyan (SE Asia, 2013) and the Horn of Africa drought (2011–2012) are examples of natural hazards that were predicted, but where forecasts were not sufficiently acted upon, leading to considerable loss of life. These events, alongside international adoption of the Sendai Framework for Disaster Risk Reduction, have motivated efforts to enable early action from early warnings. Through initiatives such as Forecast-based Financing (FbF) and the Science for Humanitarian Emergencies and Resilience (SHEAR) programme, progress is being made towards the use of science and forecasts to support international humanitarian organisations and governments in taking early action and improving disaster resilience. However, many challenges remain in using forecasts systematically for preparedness and response. The research community in place through SHEAR enabled the UK government's Department for International Development to task a collaborative group of scientists to produce probabilistic real-time flood forecast and risk bulletins, aimed at humanitarian decision-makers, for Cyclones Idai and Kenneth, which impacted Mozambique in 2019.

The process of bulletin creation during Idai and Kenneth is reviewed and critically evaluated, including evaluation of the forecast information alongside evidence for how useful the bulletins were. In this context, this work seeks to navigate the "murky landscape" of national and international mandates, capacities, and collaborations for forecasting, early warning and anticipatory action, with the ultimate aim of finding out what can be done better in the future. Lessons learnt and future recommendations are discussed to enable better collaboration between producers and users of forecast information.

* Corresponding author. European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK.
E-mail address: rebecca.emerton@ecmwf.int (R. Emerton).

https://doi.org/10.1016/j.ijdrr.2020.101811
Received 20 March 2020; Received in revised form 17 July 2020; Accepted 10 August 2020
Available online 8 September 2020
2212-4209/© 2020 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

In early 2019, two tropical cyclones (TCs) made landfall in Mozambique with devastating impacts. Cyclone Idai made landfall in central Mozambique in March and Cyclone Kenneth in northern Mozambique in April. Both were classified as intense TCs, with Kenneth the strongest cyclone to impact Mozambique in modern history (based on records from 1980 onwards); Idai resulted in more than 600 fatalities and left at least 1.85 million people in need of humanitarian assistance [1] in Mozambique alone, with further fatalities and impacts in Zimbabwe and Malawi, while Kenneth caused 45 fatalities and displaced thousands [2] in Mozambique.

Usually the first thing that comes to mind when we hear about TCs is the destructive winds. However, in many cases the water can be much more dangerous, as waves and storm surges flood the coasts and heavy rainfall causes riverine flooding further inland [3]. The impact of the rainfall has a longer timescale and can obstruct humanitarian aid during the weeks and months after a cyclone. It is therefore essential for humanitarian and civil protection agencies to have the right information on upcoming rainfall and flood risks. Since 1980, 18 tropical systems have impacted Mozambique, affecting between 11,000 (Cyclone Hudah, April 2000) and ~1.85 million (Cyclone Idai, March 2019) people, and resulting in a total of more than 2000 fatalities. The most severe of these were Cyclones Idai and Eline. Cyclone Eline made landfall on 22nd February 2000, shortly after severe flooding in January 2000, and was followed just a few days later by Cyclone Gloria, which made landfall on 8th March. This combination of events affected ~650,000 people and resulted in ~750 fatalities [4]. While TC landfalls do not occur in Mozambique every year, cyclones with the intensity of Eline, Idai and Kenneth are not unprecedented in the region.

While the Sendai Framework for Disaster Risk Reduction (SFDRR [5]) recognizes member states' primary responsibility to prevent and reduce disaster risk in their own countries, it also articulates the need for strengthening of international cooperation and global partnership to allow high-risk countries to implement DRR programmes with the overall goal to build resilience. It is not the case that national authorities simply have either 'capacity' or 'no capacity' for using forecasts, providing warnings and taking action. It is much more of a 'murky landscape' demanding "multi-level governance systems" [6] and a complex series of multisectoral, inclusive and accessible collaborations [5]. In addition to governments, humanitarian and development agencies and other relevant stakeholders need to collaborate to prepare for and respond to these types of events and are increasingly looking towards using scientific forecasts to anticipate the impacts and act early. The basic rationale for using forecasts is to reshape humanitarian assistance through innovation that improves efficiency and prevents human suffering and losses [7].

Through initiatives such as Forecast-based Financing (FbF) and the UK's Science for Humanitarian Emergencies and Resilience (SHEAR) research programme,[1] progress is being made towards the use of science and forecasts in taking early action ahead of a disaster. For example, the Red Cross took action based on forecasts of flooding in Uganda and Peru in 2016 [8]. However, many challenges remain for international organisations to use forecasts systematically to respond ahead of disasters. These barriers involve technical, communication and infrastructural issues [9], but also relate to different institutional practices, expectations, values and mandates, which further influences how success and evidence is perceived and measured [10]. Moreover, who produces knowledge and where it will be implemented touch upon deeper questions that revolve around history, epistemic politics and geographic divides [11–13] that need to be taken into account in the long-term goal towards DRR and building resilience. While the mandate for providing warnings lies with the national authorities, and triggers for early humanitarian action must be based on these mandated forecasts, international organisations can provide key supporting information. In the case of Mozambique, a WMO mission following Idai found that significant gaps and weaknesses exist in terms of accuracy of the (flood) warnings, but also in terms of overall emergency preparedness, response and coordination. This includes a limited understanding of risk at institutional and individual levels, which might be due to the low frequency nature of tropical cyclones [14].

On 19th March 2019, 5 days following the landfall of Cyclone Idai, the President of Mozambique declared a state of emergency, requesting international assistance [15]. The research community in place through the SHEAR research programme enabled the UK government's Department for International Development (DFID)[2] to task this team of authors, a collaborative group of scientists and model developers, to produce real-time flood forecast bulletins in order to support humanitarian decision-making during the flooding that followed Idai's landfall.

Less than 6 weeks after Cyclone Idai, when forecasts indicated a second TC would impact Mozambique, the same team were able to provide these emergency flood bulletins ahead of Cyclone Kenneth's landfall, after a request for reactivation from the United Nations Office for the Coordination of Humanitarian Affairs (UN OCHA). The bulletins were also shared by DFID with UN OCHA, INGC and humanitarian organisations in real-time, and the team shared the information with research partners at the Red Cross in Mozambique. The bulletins were not disseminated to the public. We used fluvial flood forecasts from the Copernicus Emergency Management Service's Global Flood Awareness System (GloFAS), based on atmospheric forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF), and then undertook detailed flood inundation estimation and impact risk assessment for population exposure estimates, providing daily bulletins for ~2 weeks at a time for each cyclone. An example from the front page of one of these bulletins is shown in Fig. 1, and a full bulletin from Cyclone Kenneth is provided in the Appendix. Fig. 2 details the daily timeline of the bulletin creation.

Emergency briefings and bulletins are a way of communicating natural hazard forecast information to decision-makers and stakeholders such as civil protection and humanitarian actors. They can be part of an online decision support system (e.g. Ref. [16,17]) or stand-alone documents that can be emailed or downloaded [18] and can also feed into synthesis situation reports such as those produced by UN OCHA. How and when the forecast information is communicated is of critical importance [19,20] and such bulletins must be able to rapidly convey the upcoming danger, as well as the uncertainty in the forecasts, through images and clear textual guidance [21,22].

The series of events that led to the request for these emergency flood bulletins suggests that, on an international scale, there are not yet adequate systems in place to make the best use of scientific forecasts of natural hazards. In addition, the rapidly increasing interest from humanitarian and development partners in using forecast information for real-time decision-making before (the impact of) a natural hazard event occurs, requires not only a critical assessment of whether the forecasts achieve an acceptable level of skill and accuracy for the intended purpose, but also of the 'how' and 'when' of the information provided by emergency bulletins, and what the needs of the users are in this process [23–25]. This paper contributes to this discussion by critically evaluating this process of real-time bulletin creation for these two events, and makes an assessment of how the bulletins were used and how they could be improved in the future. In doing so, this work seeks to navigate the

[1] Science for Humanitarian Emergencies and Resilience (SHEAR) is an international research programme jointly funded by the UK's Department for International Development (DFID), Natural Environment Research Council (NERC) and Economic & Social Research Council (ESRC) [www.shear.org.uk].

[2] In September 2020, the UK government's Department for International Development merged with the Foreign and Commonwealth Office and is now the Foreign, Commonwealth and Development Office (FCDO).

University of Reading · Copernicus Europe's eyes on Earth · ECMWF · University of Bristol

Flood hazard & impact Emergency Report - Cyclone Kenneth

Event start:	Expected 25 April 2019	Report date (first):	24 April 2019
Area	Mozambique, Tanzania	Update #1:	26 April 2019

Key points

Cyclone Kenneth made landfall in northern Mozambique approximately 90km north of Pemba (near the town of Olumbua) at ~18:00 local time on 25 April (source: WMO). It has significantly weakened since landfall, and is no longer "hurricane" strength, but will bring significant rainfall and flood hazard over the coming days, particularly in the Cabo Delgado province.

Meteorological forecast

- The cyclone has "stalled" over the Cabo Delgado province of northern Mozambique, after making landfall on 25 April, and is forecast to remain mostly over this region for at least the next 2 days.
- The stalling of the cyclone means that significant amounts of rain will fall over a more localised region, rather than a larger region further inland. This is likely to cause significant and impactful flooding in Cabo Delgado.
- ECMWF forecasts show a signal of likely high rainfall also for Southern Tanzania over the next 10 days. Thus, the evolution of flooding likelihood for this larger region should be monitored over the next 5-10 days
- Predicted rainfall totals are reduced slightly since forecasts issued on 24 April, but are still forecast to be in excess of 500mm (over the 5-day period 26-30 April) over the Cabo Delgado province (ECMWF).

Flood forecast

- GloFAS forecasts (issued 26 April) indicate that river flows will increase beyond the severe alert threshold (20-year return period) with moderate to high probability, on the Messalo, Montepuez and Megaruma rivers in the Cabo Delgado province.
- Exceedance of (at least) the high alert threshold (5-year return period) is estimated from 26 (Messalo), 27 (Montepuez) and 29 (Megaruma) April onwards. The flood peak for these rivers is currently estimated to occur around 30 April.
- GloFAS flood probabilities have decreased to the north and west of Cabo Delgado province, compared to previous forecasts issued on 24 and 25 April, due to the stalling of the cyclone.

Flood Impact

- Exposure is for river flow risk only (i.e. excludes urban, storm surge and windstorm risk). Out of the estimated ~700k people at risk, over 65k people across Cabo Delgado and Nampula are exposed to river flooding of at least 10% probability, and more than 14k people are exposed in areas with at least 50% chance of river flooding
- In Cabo Delgado province the districts most at risk are Pemba (~10k people) and Mecufi (>5k people). Significant exposure (1k-4k each) is forecast in Macomia, Mueda, Muidumbe, Ancuabe, Quissanga, Montepuez, Mocimboa da Praia, Chiúre, Meluco and Balama.
- In Nampula province the districts most at risk are Mossuril, Monapo and Memba (4k-6k people each). Significant exposure (1k-4k people each) is forecast in Nacala Velha, Muecate, Meconta, Namapa.

Note: GloFAS is designed to simulate large scale hydrological systems, so predictions for smaller watercourses should be evaluated with caution. GloFAS also does not simulate dam release or dam breaks which could be a major problem in the affected region. Estimates of exposure only account for river flooding over the next 30 days.



Figure 1 Mozambique provinces (left) and inset showing the main rivers and towns of Cabo Delgado, expected to be the most likely area to be impacted by flooding from Cyclone Kenneth, (right)

Flood Hazard Emergency Report - Cyclone Kenneth 1 *www.globalfloods.eu*

Fig. 1. Example front page from an emergency flood bulletin produced by the Universities of Reading and Bristol, and ECMWF, for DFID and the Mozambique Red Cross on 26th April 2019 for Cyclone Kenneth, detailing the key points of each aspect of the forecast including an overview of the meteorology, flood hazard and flood risk/impact. The full document is provided in the Appendix. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

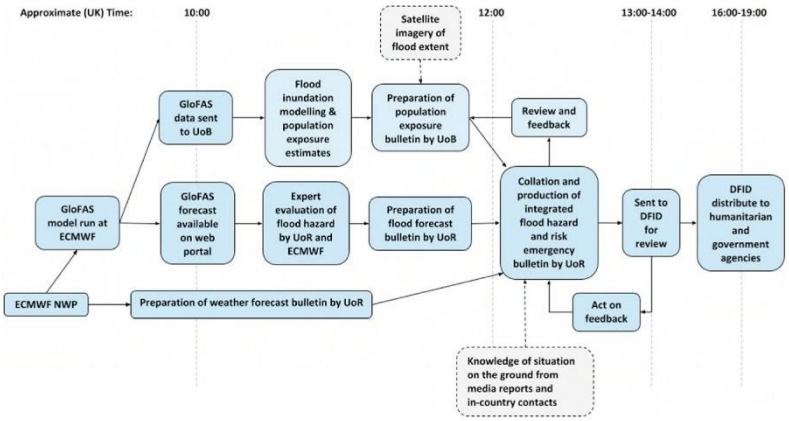


Fig. 2. Timeline of the daily emergency flood bulletin creation. Abbreviations: GloFAS: Global Flood Awareness System, ECMWF: European Centre for Medium-Range Weather Forecasts, UoR: University of Reading, UoB: University of Bristol, NWP: Numerical Weather Prediction model, DFID: UK government's Department for International Development.

"murky landscape" of national and international mandates, and capacities and collaborations for forecasting, early warning and anticipatory action, with the ultimate aim of discussing what can be done better in the future, particularly to enable increased collaboration between producers and users of forecast information.

The following sections provide a hydro-meteorological overview of the two cyclones and their impacts, an overview of the forecasts and warnings available from the national authorities in Mozambique, a description of the forecasts and models used to produce the bulletins and an evaluation of the forecasts of the two cyclones, followed by a critical discussion on the use of and response to the flood bulletins alongside lessons learnt and recommendations for the provision of such information for future events.

2. Hydro-meteorological summary of Cyclones Idai and Kenneth

The 2018–2019 south-west Indian Ocean (SWIO) cyclone season saw the largest number of intense TCs recorded in one season (based on records from 1980 onwards) in this ocean basin; of the 18 tropical systems, 11 were classified as intense TCs with wind speeds exceeding 165 km/h. In the SWIO, the cyclone season typically runs from September through to April, with the majority of systems occurring between

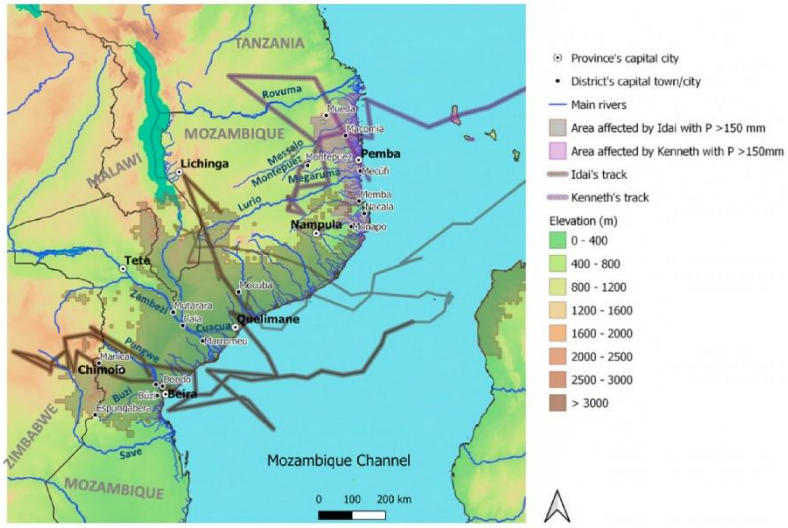


Fig. 3. Map of Mozambique, highlighting the regions affected by Cyclones Idai (grey shading) and Kenneth (purple shading), approximated by indicating the area that received > 150 mm of rainfall during each cyclone. The main rivers and cities are also highlighted, and the tracks of Idai (grey) and Kenneth (purple) are shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

December and March. In the 2018–2019 season, the first system to impact Mozambique was tropical storm Desmond, which made landfall ~200 km north of Beira (see Fig. 3) on 19th January 2019. While the storm was short-lived and much weaker than Cyclones Idai and Kenneth, with maximum 10-min sustained wind speeds of 65 km/h [26], it brought significant rainfall and some flooding to the region that would later be impacted by Idai.

The precursor of Cyclone Idai originated in the Mozambique Channel (Fig. 4a) and first affected Mozambique as a tropical depression (with wind speeds $\leq$ 62 km/h) on 4th March 2019. The rainfall from the first landfall led to significant flooding across central Mozambique and southern Malawi from 5th March onwards, particularly in the Zambezi River and its tributaries. Upstream within the affected area, the flood peak on the Zambezi occurred on 8th March [27]. Further downstream, the flooding from this first landfall peaked more than four days later, at Mutarara on the 12th, Caia on the 14th and Marromeu on the 16th March (see Fig. 3a). Water levels in some locations, including Tete and Marromeu, reached up to 1.2 m above the flood alert levels [27].

On 9th March, the tropical depression moved back over the Mozambique Channel, where it rapidly intensified. Idai was declared an intense TC on 12th March, with maximum 10-min sustained wind speeds of 195 km/h [26], before moving back towards the Mozambique coastline. Cyclone Idai made landfall near Beira on 15th March, with 10-min sustained wind speeds of 165 km/h and a storm surge of ~4.5 m [28], which, combined with intense rainfall, led to further extensive flooding.

After landfall, Cyclone Idai quickly weakened, but continued to move slowly inland, resulting in continuous rainfall for several days that led to widespread and devastating flooding in central Mozambique, especially on the Pungwe and Buzi rivers. The national hydrological bulletins reported that river levels started to rise in the Pungwe and Buzi rivers on 15th March. However, due to a breakdown of communication systems caused by the cyclone, there are no recorded observations of the flood peak. Some discontinuous observations for the Pungwe river at Mafambisse (45 km upstream of Beira) show two clear characteristics of the event: (i) a fast, extreme increase in river levels between 14th and 19th March, from 4.63 m to 9.3 m, exceeding the flood alert level by more than 3 m, and (ii) a slow flood recession from 20th March to 6th April at a rate of around 10 cm per day.

Beyond the hydro-meteorological hazards, flooding from TCs can lead to outbreaks of disease, and a cholera outbreak was declared in Mozambique on 27th March. This outbreak affected more than 6700 people in the flood-affected Sofala Province [29].

Less than 6 weeks later, another tropical disturbance began to organise to the northeast of Madagascar on 21st April and move westward towards Mozambique (Fig. 4b). This system became a tropical depression and later a tropical storm on 23rd April, at which point it was named Kenneth. Kenneth continued to rapidly intensify and was declared an intense TC on 24th April with maximum 10-min sustained wind speeds of 215 km/h [26], before weakening slightly shortly before making landfall on the evening of 25th April in northern Mozambique, near Pemba (Fig. 3b). In the period from 1950 onwards, just 12 TCs have reached intense TC status in the SWIO during the month of April, Kenneth being the latest and strongest of these.

The rainfall from Cyclone Kenneth led to flooding that began on 26th April in the Megaruma river, with a significant rise in river levels from 28th April in all major rivers in the region, including the Megaruma, Messalo, Montepuez, Lurio, Meluli, Monapo and Ligonha rivers (Fig. 3b). Water levels remained above the flood alert levels until 2nd May [27]. This severe flooding across the Cabo Delgado province of northern Mozambique during the days following Cyclone Kenneth's landfall resulted in an estimated 45 deaths and the destruction of at least 2500 homes [1], alongside the loss of a significant number of crops, fishing boats and fishing equipment [30].

3. Forecasts, data & bulletin creation

This section provides an overview of the forecast and warning information available from national authorities in Mozambique, followed by a discussion of the forecast models and data used to produce the flood bulletins, alongside additional data and methods used for the forecast evaluation undertaken as part of this study. We primarily made use of ensemble forecast products, which provide a range of possible forecast outcomes taking into account the various uncertainties associated with hydro-meteorological forecasting, and allowing the provision of probabilistic forecast information [31]. Sections 3.2 to 3.4 describe the chain of forecasts used to produce the bulletins in real-time during the two cyclones, from the meteorological forecasts that were discussed in the bulletins, and also as input to the flood forecasting system, through to the population exposure estimates, which themselves make use of the flood forecast data and additional flood inundation modelling. In the bulletins, forecast information was provided through a combination of maps and figures directly from the forecasts and forecast data, alongside expert interpretation of the data to provide a written summary of each aspect of the forecasts. The terminology used within these written summaries made reference to the forecast uncertainty and probabilities based on the ensemble forecasts. For this study, we have further evaluated the forecast accuracy through a retrospective analysis using the raw data from the real-time forecasts that were used to produce the bulletins. The bulletins were recommended for use by decision-makers alongside forecasts from the national authorities, and were not publicly disseminated.

3.1. Forecasts and warnings from national authorities

The institutions mandated to issue warnings for meteorological and hydrological hazards are the National Institute of Meteorology (INAM) and the National Directorate of Water Resources Management (DNGRH) in collaboration with regional operational water administrations (ARAs). The INGC (National Institute of Disaster Management) is responsible for coordinating the response to warnings issued by INAM and DNGRH. The disaster management structure in Mozambique is shown in Fig. 5.

INAM issue TC warnings detailing the severity of the storm (ranging from a warning for 'heavy rain, severe thunderstorm and strong wind' through to 'intense tropical cyclone'), the target area (regions likely to be impacted), an alert colour code (indicating the number of hours before a TC makes landfall; blue 24–48 h, yellow $<$ 24 h, red $<$ 6 h), and any available observed data for wind speeds and precipitation. These warnings are updated at least daily during an event.

For TC forecasting and warnings, INAM make use of the TC forecasts provided by the Regional Specialised Meteorological Centre (RSMC). RSMCs have the WMO-mandated responsibility to monitor and name TCs in their region and provide forecasts to national hydromet services. In the SWIO, the RSMC is Météo France La Réunion, who provide daily updates on the meteorological situation and potential for cyclogenesis, and issue technical bulletins and graphical warning products every 6 h during a TC. The technical bulletins contain detailed information on the location, size and intensity of the tropical system, in text format designed for the use of operational forecasters at the national authorities. Graphical warnings products are issued through the Météo France website (www.meteofrance.re/cyclone/). These provide maps of the predicted track of the centre of the tropical system over the next 5 days, including a cone of uncertainty or 'potential track area' based on forecasts from a range of models, alongside an indication of the expected intensity of the storm. The TC forecasts provided by the RSMC do not currently provide information on rainfall or flooding; INAM's operational forecasters use a variety of rainfall forecast products produced by global forecasting centres, to prepare rainfall forecasts based on their expert analysis.

During the two TCs, DNGRH also issued warnings for flooding, based on observations of river levels, whether the river levels exhibited a rising trend, and qualitative assessment of forecasts and observations of a

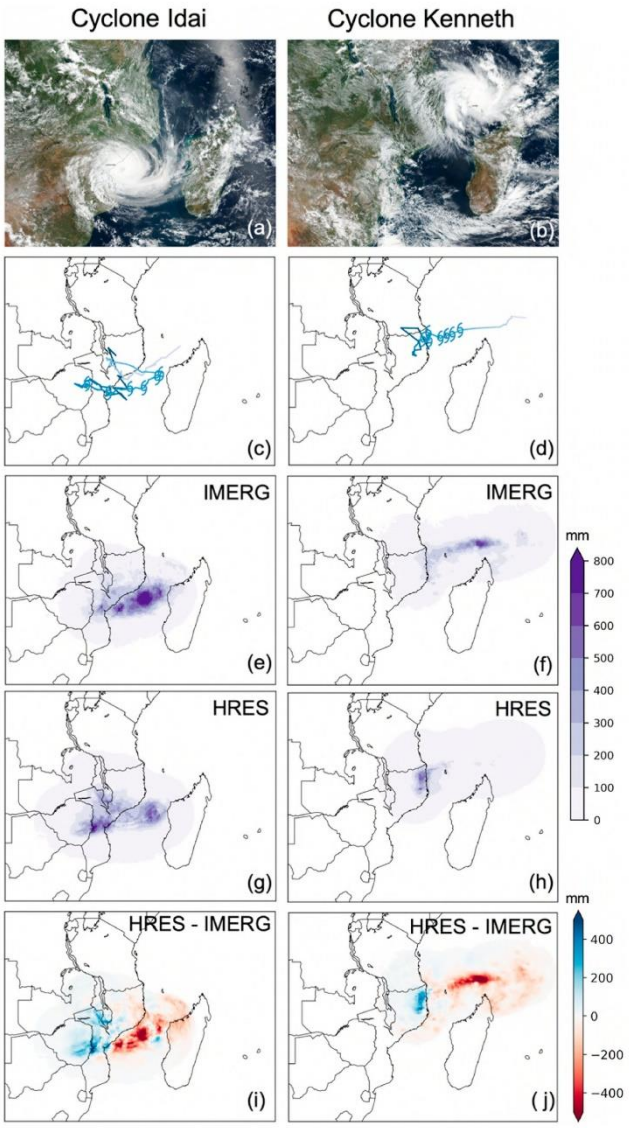


Fig. 4. Observed tracks and rainfall analysis for Cyclones Idai and Kenneth. The top panels show **satellite images** (NASA Worldview, 2019) of (a) Cyclone Idai, taken on 14th March 2019 and (b) Cyclone Kenneth, taken on 24th April 2019, followed by the **tracks** of (c) Cyclone Idai and (d) Cyclone Kenneth from genesis to dissipation, identified in the ECMWF operational analysis data using the methodology described in section 3.1. Tracks progress from light to dark shading, and cyclone symbols depict the portion of the track when the storms were classified as tropical cyclones. **Total observed rainfall** (mm) is shown for (e) Cyclone Idai, from 1 to 24 March 2019, and (f) Cyclone Kenneth, from 21st to 28th April 2019, using the IMERG satellite precipitation data (see section 3.1). Also shown is the **total forecast rainfall (mm) from the ECMWF HRES forecasts at 1 day lead time**, for (g) Cyclone Idai and (h) Cyclone Kenneth. This is the sum of all 24-h rainfall accumulations from forecasts produced 1 day ahead (for example, a forecast produced at 00UTC on 12th March for the 24-h total rainfall accumulation on 13th March) for the duration of each storm. Finally, the **mean error of the total rainfall forecast** (mm) of the ECMWF HRES forecasts at 1 day lead time is shown for (i) Cyclone Idai and (j) Cyclone Kenneth. Red indicates too little rainfall, and blue indicates too much rainfall in the forecasts. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

tropical cyclone and heavy rain. The warnings provided for Cyclone Idai, after landfall, also noted the possibility of water release from a dam in the region which could increase the risk of flooding. This knowledge of the local context, and incorporating upstream observations of river levels into warnings is key information that it would not be possible to provide using a global flood forecasting system such as GloFAS. A WMO mission report ([14], p27-29) provides further details surrounding the warnings from both INAM and DNGRH, and the forecasting capacity of both institutions.

While INAM, DNGRH and INGC are continually working towards improving the forecasts and warnings they provide, including through various research and operational collaborations (e.g. Ref. [14,33–35]) at the time of Idai and Kenneth, there was limited capacity to provide real-time forecasts of flood hazard and risk information for anticipatory action [14]. As such, the flood bulletins for Cyclones Idai and Kenneth sought to provide complementary information on the hazards and risk associated with the cyclones based on real-time global scale hydro-meteorological forecast models. The warnings issued by the RSMC

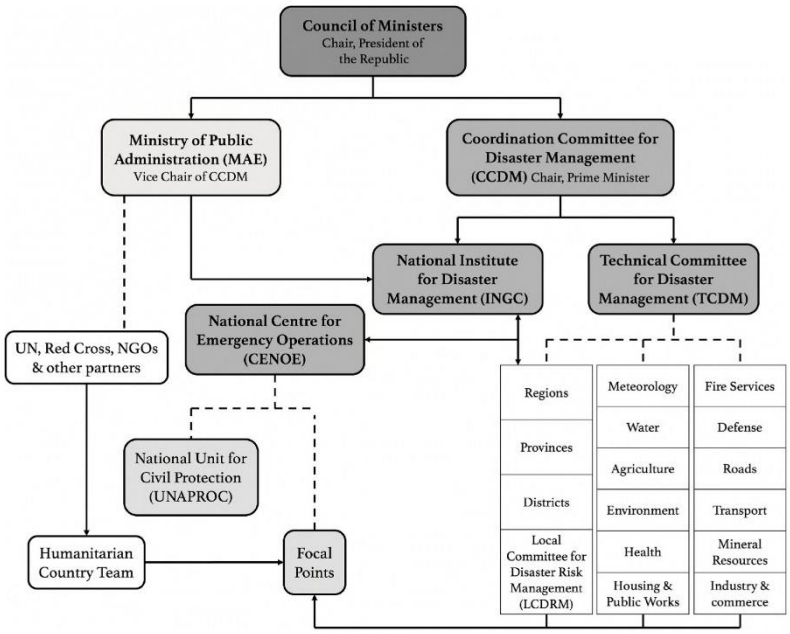


Fig. 5. Disaster risk management structure in Mozambique. Adapted from INGC [14,32] (Presented at a FATHUM project meeting in Maputo, September 2019, hosted by Universidade Tecnica de Mocambique (UDM) in collaboration with the Universities of Reading, Oxford and Bristol).

and used by INAM were considered during creation of the flood bulletins, for comparison with the ENS forecasts (see section 3.2) and to ensure consistency of the information provided. The information provided by DNGRH regarding the potential for release of water from a dam was also brought to the team's attention by our Red Cross research partners, and was cited in the flood bulletins.

3.2. ECMWF meteorological forecasts

3.2.1. Flood bulletin creation

For the bulletins, we made use of probabilistic meteorological forecasts from ECMWF's Ensemble Prediction System (ENS). The ENS is part of the ECMWF Integrated Forecasting System (IFS, cycle 45r1) providing twice-daily forecasts out to 15 days ahead, with 51 ensemble members at ~18 km horizontal resolution. The ENS graphical forecast products were used to provide contextual information on the predicted track (path) of the cyclones, alongside the amount and spatial extent of rainfall expected from the cyclones. ENS forecasts are also used as input to the flood forecasts; more information is provided in section 3.2. ECMWF also produce a high (9 km) resolution deterministic forecast (HRES), which was used as supplementary information in the bulletins to provide rainfall maps. A recent study by Titley et al. [36] found that, based on analysis of three ensemble forecasting systems from the UK Met Office, ECMWF and the National Centre for Atmospheric Prediction (NCEP), ECMWF provided the most accurate TC forecasts in the SWIO, although a multi-model ensemble can provide improved skill. Forecast skill was also found to be worse in the SWIO than other ocean basins, for the UK Met Office and ECMWF.

ECMWF's TC track forecast products become publicly available (via www.ecmwf.int) once the system is declared a TC by the Regional Specialised Meteorological Centre (RSMC) responsible for the distribution of warnings in the region.

3.2.2. Forecast analysis

In this study, we identify the TC tracks in the ENS and HRES forecast data using the tracking scheme of Hodges [37–39]. This method, described in detail by Hodges and Klingaman [40], locates vorticity maxima matching a set of criteria identifying them as TCs. The predicted TC tracks are then verified against the observed tracks, obtained from the International Best Track Archive for Climate Stewardship (IBTrACS [41]), which combines TC track data from weather centres worldwide, providing a dataset of historical tracks. Operationally, the ECMWF TC track forecasts make use of a different tracking scheme [42,43] than we use here. The tracking scheme used in this study is also currently being used to produce a long-term evaluation of TC forecast skill in the SWIO, in collaboration with the Red Cross, to provide information that can be used towards forecast-based early action for cyclones in south-east Africa. We use it here for consistency, and to allow for further comparison of the forecasts of these storms with a long-term analysis, as it is important not to make an assessment of the overall skill of the forecasting systems based on the forecasts of an individual event.

We further assess the accuracy of the rainfall forecasts for the two cyclones. Following the method of Peatman et al. [44] and Guo et al. [45], we produce composites of the rainfall associated with each TC, whereby rainfall within 5° of a track point is attributed to the cyclone. This is done for both the HRES and ENS precipitation forecast data using the forecast tracks, and for NASA's Integrated Multi-SatellitE Retrievals for Global Precipitation Measurement (IMERG V05B [46]) gridded satellite precipitation data (0.1° resolution) using the observed tracks, in order to verify the forecasts. Precipitation products based on satellite data provide valuable and consistent information, particularly in data-sparse regions, but it is important to note that while previous studies have found IMERG to satisfactorily represent the spatiotemporal distribution of TC rainfall, it has also been found to over-represent high-intensity rainfall, and in some cases, under-estimate coastal rainfall over land [47–49].

3.3. GloFAS flood hazard forecasts

3.3.1. Flood bulletin creation

The flood forecasts used were those of the Global Flood Awareness System (GloFAS, v2.0, www.globalfloods.eu), an early warning component of the European Commission Copernicus Emergency Management Service (emergency.copernicus.eu). The system couples ECMWF's ENS forecasts of surface and sub-surface runoff [50] with a hydrological river routing model (Lisflood [51]), to produce ensemble (probabilistic) forecasts of river flow for the global river network, at 0.1° (~10 km) resolution with 51 ensemble members. The initial conditions for the GloFAS model are generated by the state-of-the-art GloFAS-ERA5 river flow reanalysis [52,53]. GloFAS provides daily forecasts of flooding in major rivers around the globe, out to 30 days ahead [54], but does not currently provide forecasts for coastal flooding, which can be a significant concern during tropical cyclones. Due to this limitation, when available, we pointed to storm surge forecast information from other sources, such as the European Emergency Response Coordination Centre, and the RSMC, in the bulletins.

While GloFAS v2.0 uses an updated version of Lisflood that has been calibrated using river flow observations at 1287 stations worldwide [55], the model is not yet calibrated in the region affected by Idai and Kenneth, as no observed river flow data were available at the time the model was calibrated.

Each new GloFAS forecast is compared against flood thresholds at every grid point, providing a probability of exceeding three different flood severity thresholds. These thresholds are calculated from the GloFAS-ERA5 reanalysis for various return periods [54]; the medium, high and severe alert thresholds correspond to the 2-year, 5-year and 20-year return periods (50%, 20% and 5% annual exceedance probabilities[3] (AEPs)), respectively. This approach limits the influence of systematic biases, which are expected in regions where the model remains uncalibrated. The GloFAS user guide [53] suggests that decision-makers focus on the hydrological variability, trends, timing and relative magnitude of the flood hydrographs, rather than the exact predicted magnitude of the river flow. This is a key aspect of the GloFAS user interface, and of the interpretation of GloFAS forecasts for use in the emergency bulletins, but it should be noted that this is not simple to carry through to the inundation and exposure estimates, which must make use of GloFAS river flow forecasts and thresholds in order to provide estimates of populations exposed to flooding.

3.3.2. Forecast analysis

To evaluate the GloFAS forecasts for Cyclones Idai and Kenneth, we extract and assess the predicted timing of the flood peak and recession, and the probabilities of exceeding critical flood alert thresholds. These characteristics are the key aspects of the forecast information used for decision-making purposes. We compare these aspects of the flood forecast with observations of flood peaks and timings in the affected region, provided by DNGRH through their hydrological bulletins.

3.4. Flood risk and impact estimation

3.4.1. Flood bulletin creation

Population exposure due to flooding was estimated by combining GloFAS forecast probabilities of exceeding the flood alert thresholds, with flood inundation and population information. GloFAS' ensemble river flowforecasts were first downscaled to the ~90 m resolution of the flood inundation information, using inverse-distance-weighting. The exposure is calculated as the population exposed to a particular return period flood inundation, multiplied by the probability of exceeding a return period threshold according to GloFAS. The population is described by the High Resolution Settlement Layer (HRSL [56]) dataset, and the return period flood inundation is a binary yes/no (1/0 where wet = 1 and dry = 0) at each grid point of the global flood inundation model. The GloFAS probability of exceedance is calculated using the percentage of ensemble members that exceed the given return period threshold.

To estimate the flood inundation, a global flood inundation model framework [57] was used to delineate flood inundation zones across the region at ~90 m resolution. Return periods ranging from 5- to 1000-years (20%–0.1% AEP) were calculated in order to provide a range of possible scenarios based on the forecasts. The model estimates riverine flooding for all basins with an upstream area >50km^{2}using a sub-grid hydrodynamic model within the LISFLOOD-FP code [58]; there is no coastal flooding component. A regionalised flood frequency analysis conducted at the global scale [59] provides model boundary conditions by linking river discharge and rainfall measurements in gauged catchments to ungauged catchments, based on catchment characteristics and climatological indicators. The modelling framework therefore allows for estimation of riverine flooding at a global scale, including data-sparse regions.

Leyk et al. [60] describe the various available gridded population datasets available and their differences. For the bulletins, we used the HRSL [56] dataset, based on data availability and the work of Smith et al. [61], who demonstrated that the method used by HRSL more accurately placed populations just outside of the most hazardous areas, resulting in a better estimate of exposure, especially in rural areas. To estimate population exposed to flooding during Cyclones Idai and Kenneth, the population data (~30 m) were aggregated to the resolution of the flood inundation data (~90 m). In order to provide the total population exposure per administrative unit, zonal statistics were used. Although GloFAS forecasts do not explicitly provide the probability of exceeding return periods greater than the severe (20-year/5% AEP) alert level, many ensemble members indicated that flooding may substantially exceed the severe alert level on some rivers. As such, we additionally calculated exposure to a range of more extreme flood return periodsflood hazard, in order to report a range of exposure estimates. Exposure information was provided in the bulletins through tables and maps (see Appendix, Table 2 and Figs. 7–9).

4. Forecast analysis

4.1. Cyclone Idai

Retrospective analysis of the raw ENS probabilistic forecast data indicates that the forecasting system first began to consistently pick up the potential development of a tropical system in the Mozambique Channel, from 26th February onwards. From 1st March, the GloFAS flood forecasts indicated a 10–20% probability (based on the forecast ensemble) of severe flooding (exceeding the 20-year return period/5% AEP) across the region affected by Idai's precursor, in southern Malawi (the Shire River basin) and central Mozambique (the Zambezi River basin, including the Zambezi and Cuacua Rivers). At this point, the flood peaks associated with this first landfall were predicted to occur on 9–10th March across the affected river network, which is consistent with the flood timing later reported by the national hydrological bulletins [27]. From 4th March onwards, probabilities of severe flooding increased, exceeding 80% in rivers across the affected region from 5th March, such as along the Cuacua river (see Fig. 6a). The expected exposure also rapidly increased on 4th March (see Fig. 8), with a peak on 6th March of ~200,000 based on the 20-year return period (5% AEP), and a maximum exposure estimate of ~450,000 people (based on the

[3] Annual exceedance probabilities (AEPs) are provided alongside return periods throughout. While return periods currently represent commonly-used terminology in hydrological applications, they can be misleading when communicating potential risk to scientists, decision-makers and non-specialists from a variety of backgrounds. For example, it may unintentionally imply that if a 5-year return period flood occurs, it will not be observed again for 5 years, when in fact there is a 20% chance of a flood of that magnitude occurring in any given year (20% AEP).

1000-year return period/0.1% AEP).

From 6th March, 9 days ahead of Cyclone Idai's landfall near Beira on 15th March, the ENS track forecasts indicated a high probability (~70%) of the system 'looping around' over the Channel and making landfall as a TC in central Mozambique, although the precise landfall location remained uncertain. An example of the forecast and the ensemble spread (i.e. forecast uncertainty) is shown in Fig. 7b and d, for the forecast produced on 10th March, and the forecast progression throughout the storm's lifecycle is shown in Animation 1 in the supplementary material. This coincides with GloFAS forecasts beginning to indicate the possibility of a second flood event, in the Pungwe and Buzi River basins, with an expected peak ~18th-20th March in the two river basins, which is consistent with the available observations in the Pungwe river. Fig. 6a shows the evolution of the probability of exceeding the severe flood alert threshold for the two main rivers affected by the flooding from Idai, the Pungwe and Buzi Rivers, and for two of the main rivers affected by the first flooding event from Idai's precursor (Zambezi and Cuacua Rivers). The evolution of the GloFAS forecast probabilities, across the region, is shown in Animation 2 in the supplementary material.

From 10th March, coinciding with the intensification of the storm and its upgrade to TC status by the RSMC, the ENS forecasts for the landfall location became much more confident, alongside forecasts of severe rainfall, extreme winds and flooding in the region around Beira. Fig. 7a shows the track location errors (i.e. the distance between the forecast location of the cyclone's centre and the observed location) in the ECMWF ENS and HRES forecasts. At 3 days ahead, the average track location error was ~200 km, and at 1 day ahead, the errors were ~75 km. This is comparable to the average ECMWF forecast track location errors for TCs in the SWIO, based on the forecasts of 35 recent TCs (2014–2018; not shown).

The location of the storm in the forecasts is key for both the precipitation forecasts and the GloFAS flood forecasts. It is also important to consider that track forecasts indicate the predicted location of the centre of the TC, but the winds and rain associated with the storm can extend for hundreds of kilometres around this point (see Fig. 4a–d). This was a consideration after the storm made landfall, when track forecasts produced on 13th to 15th March indicated that Idai was likely to continue moving further west before dissipating. However, the cyclone stalled over central Mozambique rather than moving further west, resulting in sustained periods of heavy rainfall over the same region; this stalling was picked up in the track forecasts with approximately 1 day's lead time, on 16th March, and this resulted in uncertainty in the flood forecasts.

This is shown in Fig. 6a, by a drop in the probability of severe flooding, from ~40% to ~20% during the 13 - 15th March period when forecasts were indicating the cyclone was likely to move further to the west. When the stalling was picked up in the track forecasts, the probabilities of severe flooding increased rapidly, and remained consistently high throughout the affected river network (particularly the Pungwe, Buzi and Save Rivers) after Idai made landfall.

Evaluation of the HRES rainfall forecasts using IMERG satellite rainfall data (Fig. 4e–j) indicates that, over land and at short lead times, the ECMWF HRES forecasts for Cyclone Idai typically over-predicted the rainfall totals across much of central Mozambique, and under-predicted the rainfall in northern Mozambique and over the Channel. At 0 days lead time (i.e. a forecast produced at 00UTC for the total rainfall over the following 24 h) errors over land are equivalent to <30 mm per day, or <400 mm over the duration of the storm. Taking all 1-day-ahead forecasts for the duration of the storm (shown in Fig. 4 for the HRES), rainfall was over-predicted by up to 300 mm in central Mozambique, and up to 400 mm in western Mozambique. In contrast, with increasing lead time beyond 1 day ahead, the forecasts show an under-estimation over much of the affected area. At 2 days ahead, we see an under-prediction in central Mozambique of up to 300 mm, and an over-prediction of up to 300 mm in western Mozambique. Results for the ensemble mean ENS forecast (based on the ensemble mean rainfall associated with the ensemble mean track, not shown) indicate a similar over-prediction in the west, but an under-prediction at all lead times

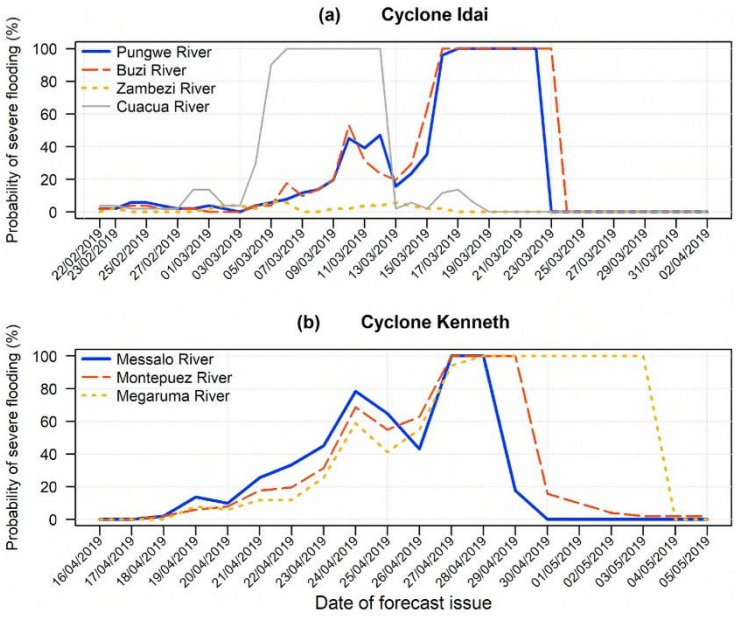


Fig. 6. GloFAS maximum probability of exceeding the severe flood alert threshold (20-year return period/5% AEP) during the 30-day forecast horizon, for major rivers affected by (a) Cyclone Idai and (b) Cyclone Kenneth, for forecasts issued daily ahead of and during each cyclone. The rivers and locations (see Fig. 3) shown are (a) Pungwe at Mafambisse (15 km northwest of Dondo), Buzi at Buzi, Zambezi at Tete, Cuacua at Campo (Mopeia district, 50 km west of Quelimane), and (b) Messalo at Narere (60 km north of Macomia), Montepuez at Quissanga district (45 km southeast of Macomia) and Megaruma at Chiúre district (12 km south of Mecúfi).

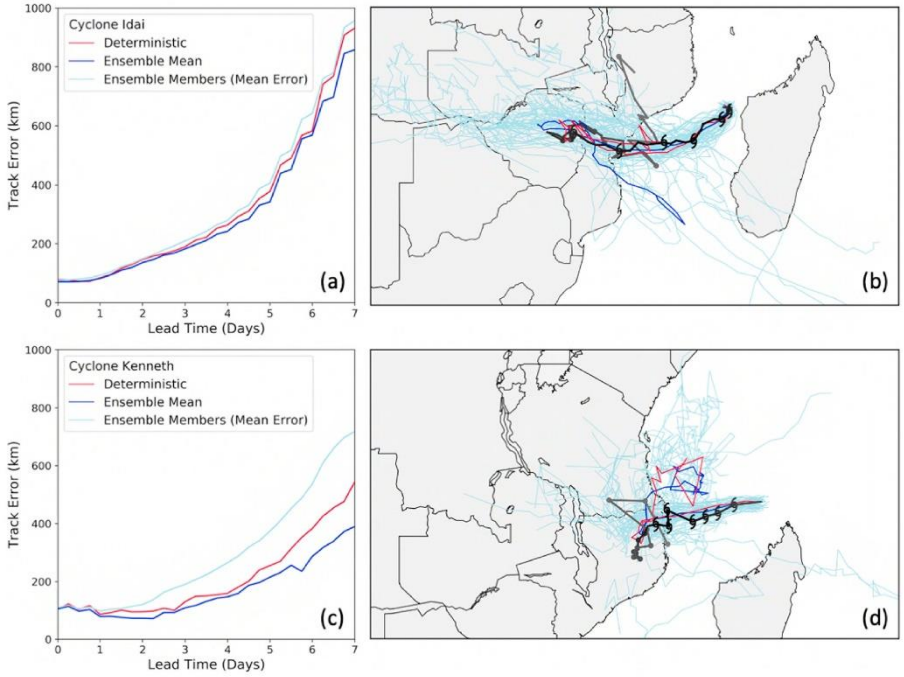


Fig. 7. Track location errors with lead time for ECMWF forecasts of Cyclones (a) Idai and (c) Kenneth. Errors are the mean error across all forecasts (produced twice daily at 00 and 12 UTC) for the tropical cyclone stages of each storm, for the high-resolution deterministic (red) and ensemble mean (dark blue) forecasts, and the mean error across all 50 individual ensemble members (light blue). Forecast tracks are verified against the IBTrACS observed best tracks. An example forecast for Cyclone Idai is shown in (b), issued on 10th March 2019 at 00 UTC, and for Cyclone Kenneth in (d), issued on 23rd April at 12UTC. These maps indicate the forecast track for the deterministic (red)and all 50 individual ensemble members (light blue), alongside the track of the ensemble mean (dark blue). The observed tracks of Cyclones Idai and Kenneth are shown in black, where tropical cyclones symbols denote the cyclone-strength stages of the storm, followed by a grey solid line representing the post-cyclone stages. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

across much of the affected area of central and northern Mozambique. These errors in the rainfall can be tied to the forecasts of the cyclone's track, which predicted the storm to continue moving west rather than the observed stalling over central Mozambique, and the impact of this is seen in the GloFAS flood forecasts as the aforementioned drop in the probability of severe flooding before the stalling was picked up.

The locations and rivers affected by the flooding were correctly predicted by GloFAS with a 10-day lead time. However, for severe flooding the probabilities were relatively low (<30%) until 9th March, with large uncertainties in the expected flood peak timing. The exposure estimates began to highlight the potential severity of the event from 4th March. However, at this point the areas with highest exposure estimates were predicted to be in the Mutatara District, on the border with Malawi. In line with the track and flood forecasts, as time progressed the exposure estimates shifted southwards as the landfall location of cyclone Idai became more certain. As a result, districts such as Nhamatanda and Buzi were forecast to be at risk of flooding at or shortly after landfall.

Comparison of exposure estimates with post-disaster reports are challenging as these principally report the total number of affected people, while the bulletins provided estimates of the number of people affected by flooding, rather than by other/all aspects of the cyclone. According to a UN OCHA situation report [62], 198,300 houses were partially or totally destroyed by the cyclone (while many of these may be due to flooding, it is not possible to say if this was the sole or primary cause), with a further 15,794 households flooded. This suggests that our estimates of the number of people exposed were likely reasonable, as for the 20-year flood hazard (5% AEP) the total estimated exposure was ~200,000 people. An assessment of 14 districts in the Sofala and Manica provinces estimated the total affected population to be ~1 million [2], which is at the upper end of our estimates (see Fig. 9). However, the authors of the report state "it is possible that there was some misunderstanding around the terminology used in Portuguese, and that the floods were understood as a synonym of rain", suggesting a potential overestimation of people flooded, and highlighting the complexities involved in comparing such exposure estimates to the available post-disaster assessments.

4.2. Cyclone Kenneth

Ahead of Cyclone Kenneth, the ENS forecasts began to indicate that a tropical system may develop north of Madagascar and impact Tanzania or northern Mozambique, from 18th April onwards. The system was declared a TC by the RSMC on 23rd April. Forecasts of the landfall location in northern Mozambique became much more accurate after the storm's genesis, from 22nd April, and the ensemble spread (i.e. forecast uncertainty) continued to decrease with each new forecast until Kenneth's landfall on 25th April.

Track location errors for Cyclone Kenneth are shown in Fig. 7, and indicate that at 1 day ahead, forecast skill was similar to Cyclone Idai, with an error of ~75 km. However, at 3 days ahead, track location errors

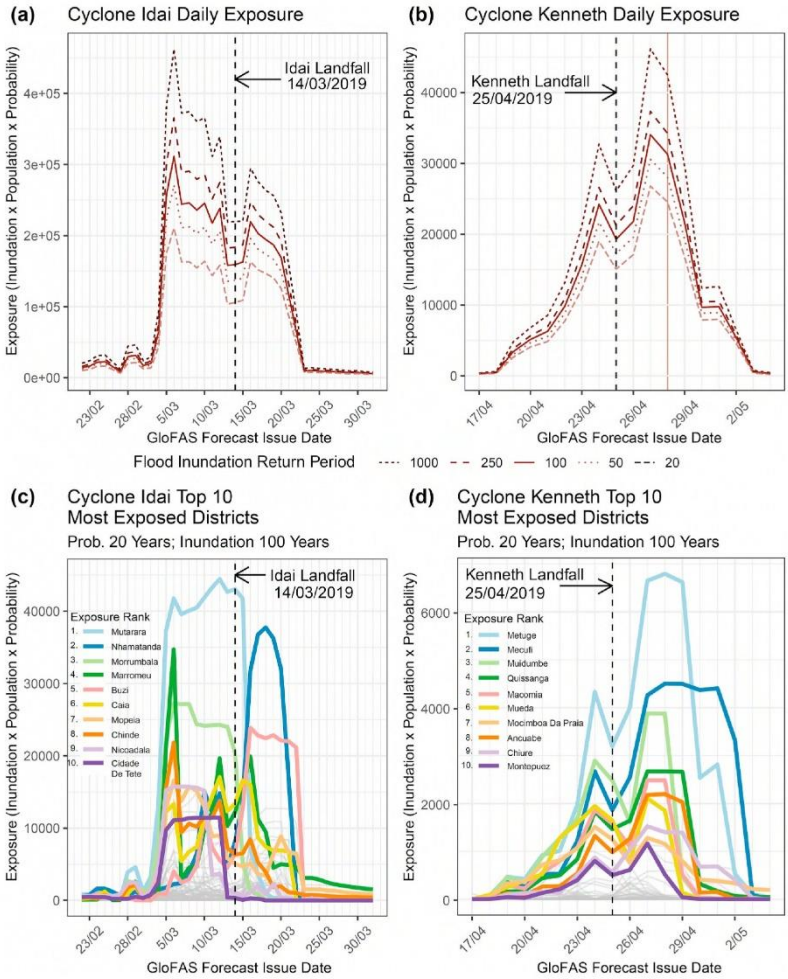


Fig. 8. Daily total exposure estimates for Mozambique for (a) Cyclone Idai and (b) Cyclone Kenneth, for five different flood inundation return periods (20, 50, 100, 250 and 1000-year return periods, equivalent to 5%, 2%, 1%, 0.4% and 0.1% AEPs, respectively, indicated by different line styles), and exposure per district for (c) Cyclone Idai and (d) Cyclone Kenneth. The ranking is based on the total number exposed during the period shown on the graph. The faded grey lines are other districts in Mozambique, outside of the 10 districts with the highest exposure. The exposure per district is calculated based on the severe flood level of GloFAS (20-year return period/5% AEP), the 100- year (1% AEP) inundation return period and the HRSL population dataset.

were much smaller for Kenneth, at ~100 km (compared to ~200 km for Idai). This is also significantly smaller than typical location errors for ECMWF forecasts in the SWIO, which are ~200 km at 3 days ahead, based on the average error across 35 recent TCs (2014–2018; not shown). The errors increased more rapidly with lead time for Idai than Kenneth, implying that Kenneth's track was much more predictable. Typically, forecast location errors are smaller where TCs tend to move more zonally (such as was the case with Kenneth) compared to those which meander or recurve [63,64].

This is reflected in the GloFAS flood forecasts, which, coinciding with

Table 1
Overview of estimated population exposed to river flooding from Cyclone Kenneth, from the bulletin produced on 26th April 2019 (see Appendix), for the Cabo Delgado province, alongside the total number of people reported affected in each district [30]. It is important to note that the definition of affected also covers many more aspects of the impacts than river flooding, such as extreme winds, food insecurity, previous heavy rains and other factors.

District	Flood Bulletin Estimated Population Exposed to River Flooding from Cyclone Kenneth		Total Number of People Affected
	(10% probability)	(50% probability)	
Pemba	9952	3164	9366
Mecufi	5386	4213	1645
Macomia	3906	338	85225
Mueda	3631		2568
Muidumbe	3430		16994
Ancuabe	3184	2475	7515
Quissanga	2805	2805	21154
Montepuez	2519		163
Chiure	1644	853	24435
Meluco	1356	576	5451

the increasing confidence of the landfall location in the ENS forecasts, consistently indicated an increasing probability of severe flooding in the Messalo, Montepuez and Megaruma Rivers, from 18th to 24th April (Fig. 6b). The expected exposure began to increase on 19th April (6 days before landfall), with the most rapid increase also occurring on 22nd April. Similarly to the forecasts for Idai, a drop in the GloFAS probability of severe flooding is seen on 25th– 26th April, due to the ENS track forecasts indicating the storm may continue to move west, rather than stalling over the Cabo Delgado province of northern Mozambique, as was observed. The peak expected exposure occurred 2 days after landfall and ranged from 25,000 people for the 20-year return period (5% AEP) flood inundation to 45,000 for the most extreme 1000-year return period (0.1% AEP) flooding. Fig. 8d shows expected exposure per district for the severe flood (20-year return period/5% AEP) probability and the 100-year (1% AEP) flood inundation. Unlike Cyclone Idai, the ranking of the most exposed district does not significantly alter during the event, due to the more predictable track of Cyclone Kenneth.

Comparing these estimates for population exposed per district, based on the bulletin produced on 26th April (see Appendix, Table 2), with a post-disaster assessment [30] from the Global Facility for Disaster Reduction and Recovery (GFDRR), indicates that these estimates correctly predicted which districts were at risk. The districts listed in the bulletin with a probability of flooding (based on the 250-year flood inundation/0.4% AEP) exceeding 10% are the same districts that were indeed affected by the cyclone, and the districts estimated to be at risk with a higher (50%) probability of flooding generally correspond to those with the highest number of people affected [30]. Table 1 provides a district-level comparison between the exposure estimates provided in the bulletin, and the number of people affected per district, in the Cabo Delgado province. While the estimates in the bulletin are somewhat lower than the total number of people affected, this is to be expected as the definition of affected covers many more aspects of the impacts than river flooding, such as extreme winds, food insecurity and disease, and these numbers are also "superimposed on previous heavy rains at the beginning of the year, the effects of Cyclone Idai in some districts, and vulnerable population groups that had been resettled as part of the conflict stabilisation efforts of the previous year" [30]. This poses a significant challenge in evaluating such exposure estimates, as even the best available data on the number of people affected have drawbacks, such as to what degree these data indicate impacts of the storm itself, and, for example, information may be provided in terms of the number of households affected, but it is not clear how many people are assumed per household.

5. Were the emergency flood bulletins useful?

In this section we use evidence from reports, interviews, conversations, letters, emails and written commentary at post-event meetings,[4] to review the use, usefulness and the potential impact of the bulletins. We critically assess to what extent we can be sure those receiving them found them useful, and were able to take better decisions based on the forecast information, or whether they were just an addition to the overload of information for humanitarian actors and governments involved, distracting from the priorities on the ground.

5.1. Making the best use of scientific forecasts of natural hazards

Using science actively in planning and responding to natural hazards is the 'holy grail' of forecast development. The key is to be able to generate, disseminate and communicate the information in meaningful ways to different users who can actively use it early enough for decisions to be taken. In our case, this was a request from DFID following the declaration of a state of emergency in Mozambique and a request for international assistance, and therefore there was a lot of active discussion between the forecast producers and those responsible for passing on the information to h umanitarians on the ground (see Figs. 2 and 5 for an overview of the bulletin production and feedback process with DFID, and the national disaster management structure in Mozambique, respectively).

> "This is the first time we have been able to use science so early in both planning for and responding to the devastating impact of cyclones. Your expert analysis, collaborative effort across your organisations and with DFID colleagues, and willingness to tailor and communicate the analysis to the needs of the humanitarian agency end users was well received." [Professor Charlotte Watts, Chief Scientific Advisor for DFID]

> "The real innovation of these bulletins lies in the fact that this information has been produced in real-time, but of course many challenges remain." [DFID]

Feedback received from our international humanitarian partners noted that this was the first time that flood risk information had been provided in real-time to them, and that the type of information was perceived as extremely valuable, innovative and promising for future interventions, particularly due to the move from weather forecasts to more impact-based forecasts. Access to the meteorological forecasts used as input to GloFAS allowed the provision of the meteorological context of the flood hazard and risk, and the inclusion of probabilistic meteorological, hydrological and exposure information in one document was found to be extremely valuable and useful. Nevertheless, despite the novelty of the type of information that was produced, it is clear from the series of events that led to the request for these emergency

[4] A Discussion Meeting on Cyclones Idai and Kenneth was organised by the Universities of Reading (Rebecca Emerton, Andrea Ficchi and Hannah Cloke), Bristol (Laurence Hawker) and Oxford (Sara de Wit), and hosted by the Universidade Técnica de Moçambique (Rui da Maia, Benedita Nhambiu and Joaquim Cuna) in Maputo, Mozambique. The meeting took place on 20th September 2019 and brought together representatives from key national agencies (INAM, DNGRH, INGC and the Mozambique Red Cross) involved in the forecasting and response to the cyclones, hydrologists from regional water agencies, and academics from various institutions and scientific backgrounds, to discuss their experiences during Cyclones Idai and Kenneth, barriers and challenges in forecasting and response, differences between the two events, the use and usefulness of the flood bulletins, and ways to move forward through new collaborations and strengthening existing collaborations. The meeting was followed by a GloFAS training workshop for a group of academics and technicians in Mozambique, and FATHUM collaborators from Uganda and Mali, from 23–25 September 2019.

flood bulletins that we do not yet have adequate systems in place to make the best use of scientific forecasts of natural hazards for international humanitarian actions both in terms of their real time nature and the content.

5.2. Cascading information to decision makers

The information provided in the bulletins was cascaded to high-level international organisations, the government of Mozambique, and local partners and emergency response coordination centres (but not the public), in a number of ways. The government of Mozambique declared a state of emergency and formally requested international support shortly after Cyclone Idai's landfall. The humanitarian response was led by the Mozambique Disaster Management Agency (INGC), which worked closely with UN OCHA, and the UN clusters. The bulletins were provided as an additional information resource to inform situational awareness, preparedness and response planning, initially through OCHA, which is mandated to coordinate humanitarian assistance with the consent of the national authorities (UN General Assembly Resolution 46/182 [65]). UN OCHA's situational reports (SitReps) drew directly from the bulletins. These SitReps are public documents (available via reports.unocha.org) and shared with the Government. The INGC initially received the bulletins indirectly from OCHA and subsequently directly from DFID who commissioned them and were responsible for their dissemination. Through the provision of information to DFID and onwards to the UN OCHA, who included key points from the bulletins in their daily situation reports, the information was able to reach a wide range of decision-makers at international and local levels, in both government and humanitarian organisations. This led to UN OCHA formally requesting reactivation of the bulletin production when forecasts indicated a second TC would impact Mozambique, and the same team were able to provide these emergency forecast bulletins before Cyclone Kenneth's landfall.

> "UN humanitarian response actors stated that the reports produced were "tremendously helpful as we continue to analyse the risks in the days ahead". UN OCHA extracted the key analysis to include into their daily sitreps, which all humanitarian actors and the GoM [Government of Mozambique] use as a key reference point." [Professor Charlotte Watts, Chief Scientific Advisor for DFID]

> "The information was presented to WHH's Emergency Response director on the ground in Mozambique and to the "Emergency Decision Panel" – senior Management in Bonn, Germany, to facilitate the decision" (to send part of the team to conduct an assessment in/around Pemba) [Welthungerhilfe (WHH) via DFID]

In addition to providing the information to DFID and UN OCHA, we were able to share the bulletins with national humanitarian and government organisations directly, through SHEAR collaborations with in-country partners. This provided the opportunity for decision-makers to ask questions directly to the team involved in producing the bulletins, and to receive the information faster than may have been possible through the information cascade from high-level organisations. Feedback received from decision-makers and operational organisations was also useful for the team producing the bulletins and allowed us to refine the methodology and format with each new bulletin produced. Through this process, we were also made aware of some key aspects of the situation on the ground, which could be further incorporated into the following flood bulletins and passed on to DFID, such as knowledge of a dam in the area that may be at risk. This was important information to highlight in the bulletins, as not all reservoirs are represented in the GloFAS hydrological model, resulting in uncertainty in the flood forecasts around this location.

> "We/I only started receiving the reports when Kenneth had made landfall in Cabo Delgado. Personally I found them very informative and with relevant information and details. The reports were widely circulated here in Mozambique (by different UN organizations etc)." [Hanne Roden, Programme Coordinator, FbF Project Delegate, German Red Cross – Mozambique]

5.3. How were the bulletins used in taking decisions?

A key objective of the bulletins was to facilitate decision-making and increased understanding of the situation and nature of the risk.[5] While we learn from partners that the ground-breaking element of the bulletins was the fact that it was "produced, shared and it informed" in real-time, it is more challenging to find out how this type of information directly informed decision-making. It is not always easy for organisations to articulate how the bulletins were helpful. In emergency situations, decision-makers are required to consider numerous and varying pieces of information in order to take a balanced decision, and as such, a specific contribution to a complex decision will always be difficult to convey. Discussing the use of big data (and the so-called four Vs: Volume, Variety, Velocity and Value), for emergency decision-making in the context of natural disasters, Zhou et al. [66] state "one of the important contents of natural disaster emergency decision lies in the way to describe the data with different sources, data mapping and fusion, feature extraction and classification, quick and accurate access to valuable information and intelligent decision in emergency response". The bulletins were therefore one piece of information amidst an array of other types of information within a wider system and in a complex situation. Some operating organisations incorporated the bulletins into their existing knowledge dissemination products (UN OCHA), yet for others it was the first time they had received real-time information and might simply not yet know what to do with it. Furthermore, it is difficult to evaluate whether the use of the bulletins enabled organisations to take better decisions than if they hadn't had the information.

Feedback from partners, both directly and through DFID, indicates that a key contribution of the bulletins was to assist in creating an overview of the situation; where and when flooding was likely to occur, where there were more people at risk, and when the floods were likely to recede. This was best done using a range of information from both the bulletins and other sources of local data.

> "Ahead of Cyclone Kenneth, WHH was present in Mozambique responding to Cyclone Idai in Beira & Nhamatanda. The […] flood risk analysis was used shared together with other data to understand the situation in Cabo Delgado and get a first idea of the potential flood impact." [Welthungerhilfe (WHH) via DFID]

> "The bulletins were very helpful. They gave us an overview of which rivers were at greatest risk of flooding, and this helped inform where we gave the greatest attention to. We used them to help inform our daily briefings to partners, as well as in our public information products. All of this meant that the humanitarian community had far greater information, in real-time, about flood risks, than we have often had access to in the past." [Gemma Connell, Head of Regional UN OCHA in Southern and Eastern Africa]

> "Whether they specifically 'redirected' measures, I don't know, but I am fairly sure that they assisted in creating the overview [of the situation]." [Hanne Roden, Programme Coordinator, FbF Project Delegate, German Red Cross – Mozambique]

Through personal communication with DFID, we were informed that

[5] It is important to note that national authorities have the mandate for early warning and civil protection. Triggers for taking early humanitarian action should always be based on forecasts and warnings from mandated national authorities. In practice, information from international organisations and global forecasting systems can be used to support the decision-making process.

the bulletins enabled decision-makers to understand the flood risk, move assets and equipment, and release supplies. This enabled an early response to take place into the locations that were likely to be at greater risk, and which also may have become cut off when the flooding reached its peak, meaning that aid would have been further delayed until after flooding had receded. This was particularly the case for Cyclone Kenneth, as for Idai these bulletins were produced in response to the cyclone's landfall, when some of the most-affected areas had already been cut off. Partners have told DFID that they relied on that information and that it helped them to make quick, informed decisions rather than more subjective decisions that would have had to be made without the information provided in the bulletins.

> "The information was used to decide whether to send part of the team to conduct an assessment in/around Pemba. […] A decision was made to send an assessment team and the analysis informed the composition of the assessment team as well as the preparation: - preparedness activities to be considered such as contingency stock, - measures considered: immediate availability of wash kits / hygiene kits, water treatment & filters and tarpaulins" [Welthungerhilfe via DFID]

5.4. The need for future bulletins

Further communication with partners has highlighted the need for this type of information to be available for future events. Providing all relevant hydro-meteorological hazard and risk information in one place, including expert interpretation of the data and forecast products, proved to be essential for informing decision-making and providing a more in-depth overview of the situation. Both DFID and the Mozambique Red Cross (Cruz Vermelha de Moçambique, CVM) would like to continue to explore the utility of the bulletins and possible ways forward, including whether there is potential to establish a standing capability for similar emergency briefings alongside collaboration and training with local partners, to better co-design future bulletins and build capacity for their use. CVM has requested further collaboration with the team, in order to develop simulation exercises and similar analysis of flood hazard and risk in relation to other river basins and past flood events. This could be relevant for further developing their flood and cyclone Early Action Protocols in Mozambique, and to enhance preparedness in the face of future events.

> "In the FbF context in Mozambique it could be very interesting and relevant to imagine a case with a certain constellation (wind speed, expected rainfall etc.) making landfall in, for example, Sofala and affecting the Buzi river basin […] you could produce the same type of maps, possible impacted areas, etc. […] and together we would be able to make some fairly accurate maps, and prognosis of the impact should such an event occur. And if we had this kind of data, beforehand, then it could for sure inform the measures taken by various actors including RCRC/CVM in preparation of the next cyclone/rainy season." [Hanne Roden, Programme Coordinator, FbF Project Delegate, German Red Cross – Mozambique]

> "Given the demonstrated utility of such analysis, we intend to learn lessons and examine options to better enable this type of science input in future humanitarian responses." [Professor Charlotte Watts, Chief Scientific Advisor for DFID]

6. Challenges, lessons learnt & recommendations

In this section, we discuss the key challenges faced and lessons learnt during Cyclones Idai and Kenneth in Mozambique, and provide recommendations for the provision of such information for future events. In doing so, we advocate careful consideration of the differences in humanitarian response, management set-ups and different actors, in extrapolating these experiences to other potential scenarios.

During the Discussion Meeting on Cyclones Idai and Kenneth, held in Maputo on 20th September 2019, participants were asked to reflect on the usefulness of the bulletins, and following this, were also asked to discuss challenges that were faced, and how the forecast information and response could be improved. Sixteen participants, including representatives from INAM, DNGRH, INGC, regional water agencies (ARAs), the Mozambique Red Cross (CVM) and the Technical University of Mozambique (UDM), provided their perspectives through active discussion, interviews and written commentaries. The responses, alongside observations from the team that produced the bulletins and those facilitating the Discussion Meeting, are incorporated throughout the following subsections.

6.1. Towards co-production and embedding of social science knowledge

6.1.1. Challenges

Considering that the bulletins were produced in an ad hoc fashion during a state of emergency, challenges around the co-production of knowledge in a complex landscape inevitably emerged, which warrants reflection on how knowledge production and use can be more systematically integrated as an ongoing effort for future collaborations and capacity building.

While the real innovation was the production of this kind of information for humanitarians and other interested organisations, as social science research on the use of science and technology for meeting human-development needs has demonstrated, one of the remaining challenges is to better link science and decision making in more socially and institutionally embedded ways from the beginning [10]. This means that more research is needed to understand which institutional factors promote or inhibit the use of uncertain forecasts, and how organisations can be better prepared to make use of real time information during emergencies. One of the reasons for international humanitarian organisations setting up early action protocols, is to pre-assess the skill of various forecasts and decide which are the most appropriate to use for a certain region or event. If new forecasts are then produced during an emergency, how can decision-makers judge the quality of the forecasts? When humanitarian decision-makers begin receiving information from organisations outside the mandated national authorities, how do they know whether this information is trustworthy? During Cyclones Idai and Kenneth, an overwhelming amount of information was received by the Red Cross, such that it was necessary to designate a 'gatekeeper' to filter the information and avoid excessive amounts of information reaching decision-makers on the ground.

As research on the use and societal uptake of forecasts has demonstrated, the question of whether scientific information is useful and useable or not, is not only bound to epistemological concerns (is the science 'good enough', or do users understand the information?), but can largely be explained by nontechnical or institutional factors like mandate, capacity, accountability, how success is defined and measured, and by regulatory frameworks and attitudes to risk [67,68]. In other words, less scientific uncertainty does not automatically lead to less policy uncertainty. Moreover, scientific information is likely to be more effective if it is perceived to be not only scientifically credible but also salient and legitimate, which in turn is closely linked to producing knowledge that is socially robust and can be used within the context in which it is intended for [10].

6.1.2. Lessons learnt and recommendations

To address these critical challenges, four institutional functions are proposed as a means to effective coproduction [10]: (1) face-to-face contact between stakeholders, (2) translation, both literal in terms of language and jargon, and metaphorical, (3) bring together experts and decision-makers from all relevant disciplines (collaboration), and (4) represent different interests to ensure fairness. Additionally, because the representation of scientific uncertainty is largely shaped by social

relations among scientists and those they advise [69], more research is needed to understand this kind of 'boundary work' for different stakeholders and how to deal with the question of uncertainty and accountability in the context of emergencies. Since the emergency bulletins are so-called "boundary objects", which are outputs that "are both adaptable to different viewpoints and robust enough to maintain identity across them" [70], we have to take the *continuous* co-production of knowledge across boundaries of science and policy more seriously, rather than viewing it merely as an ad-hoc endeavour.

The Weather and Climate Services for Africa (WISER) and Future Climate for Africa (FCFA) programmes have produced detailed and clear guidance for co-production of weather and climate services [71], aimed at those involved in co-production "ranging from the academic/practitioner project manager to national meteorological services and government officials wanting to integrate co-production principles into their own work processes". The manual sets out guidelines, recommendations and tips for initiating successful and lasting collaborations that work towards the co-production of weather and climate services that are relevant and useful (https://futureclimateafrica.org/coproduction-manual/). Effective collaboration and co-production should begin with identifying all the relevant actors, typically within three categories: producers (for example: national hydromet services, local forecasters, regional or international organisations, research institutes), intermediaries (government ministries, NGOs, media, research institutes), and users (government ministries, humanitarians, citizens, private sector, local leaders), and work towards building common ground, and co-developing and co-delivering solutions [71].

Recommendation 1: Based on the guidance set out in the Manual for Co-production in African Weather and Climate Services [71], identify the full range of relevant actors (producers, intermediaries and users) to initiate, or develop existing, partnerships and build trust-based relationships and collaborations that go beyond an emergency event situation.

Ideally, the format of such bulletins would be agreed, through collaboration between the full range of relevant actors, prior to the occurrence of an event, providing a template outlining the scientific information required in each section, and the terminology that would allow the content to be universally understood. This process would allow those producing the bulletins to be sure that the information being provided is really what is needed by decision-makers, and would also work towards establishing a collaboration and mutual understanding between forecast users and producers.

While it would be best to agree ahead of time on the key information to be included, it was imperative that the team remained flexible to changing user needs in order to adapt the information provided in response to the changing situation. The real-time feedback that was provided by humanitarian partners, which may be characterized as an emergent form of co-production, was immensely helpful to understand user needs and improve the bulletins. But the feedback received after the events by national authorities, during stakeholder engagement meetings, also identifies further research gaps related to institutional barriers for using forecasts; how to reduce access limitations to forecast data and information, and how best to involve local communities in preparedness activities, both of which would help to increase the uptake of forecast information for decision-making and could also be applied to other regions and types of event.

Recommendation 2: Work towards understanding which institutional factors promote or inhibit the use of uncertain forecasts, and support organisations to be better prepared to make use of real-time information during emergencies, for example through training that involves all actors.

Recommendation 3: Through collaboration between all relevant actors (producers, intermediaries and users), agree on the most effective and useful format for future emergency bulletins, including both scientific content and terminology, while allowing room for flexibility during an emergency situation.

6.2. Operationalisation

6.2.1. Challenges

A key consideration and challenge arising from the production of these emergency bulletins, is the systematic production of this type of information for future events. There have since been several discussions around the challenges faced by decision-makers who received and used this information during Cyclones Idai and Kenneth, but they do not know if they will receive this information ahead of future floods, cyclones or other natural hazards. These bulletins were produced primarily by a team of research scientists in collaboration with model developers, which is unsustainable and cannot realistically be replicated every time an extreme event occurs. It is imperative to find a way to ensure that such information can be relied upon ahead of future emergencies. There are however many barriers and challenges to providing an operational forecast product, which requires effective identification, production and use of the best science-based information, collaboration between authorities and actors at a range of levels, and effective co-production and dissemination of information in real-time.

Another key challenge was the impact of the cyclones on communication infrastructures, that meant it was not always possible to access the internet in order to download the bulletins, raising the question of whether it may be possible to provide a very brief overview of key information, for example, by text/SMS message. However, this has further potential challenges, not least to ensure that any such messages do not conflict with those sent by national authorities, further highlighting the importance of close collaboration between international organisations and mandated national authorities.

A further question is whether the time that the bulletins were sent out, typically in the evening, was useful for those on the ground, or whether a different timeline would have been more efficient. While new forecasts are available from GloFAS each morning, it takes time to provide this data to a team working at another institution who need to run another model, and to write an expert interpretation and summary of the forecast information including new maps and figures, before incorporating feedback and clarifications. For high-level organisations such as DFID and UN OCHA, post-event feedback was that this timeline worked well, as they were able to look into the information at night, ready to incorporate into daily briefings that would be shared first thing in the morning. A potential issue with this, however, is that the forecast models are updated again shortly after these briefings are circulated in the morning, meaning that information is potentially out-of-date within a short time after those on the ground are receiving it. Of course, this timeline and experience would look different in regions of the world where the time difference is more significant than in this particular case. Finally, one of the key challenges for many involved was that the bulletins were provided in English only, when many of the national actors require information in Portuguese.

6.2.2. Lessons learnt and recommendations

The information produced and provided through these emergency bulletins was shown to be valuable and useful for decision-making, but provision of this information by research scientists is not sustainable nor is it the best way to co-produce and disseminate information. It is imperative to find a way to ensure that such information can be relied upon ahead of future emergencies. One project working towards systematically producing forecast bulletins for a range of natural hazards in Europe is the Aristotle Consortium (http://aristotle.ingv.it/tiki-index.php), which produces emergency bulletins for the European Emergency Response Coordination Centre. At the global scale, ideally an organisation such as the World Meteorological Organisation (WMO), which was established to represent an authoritative voice for meteorological and hydrological hazards globally, would take a role in supporting collaborations that work towards the operational production and dissemination of such bulletins by national authorities, in collaboration with international centres.

In section 6.1, it was highlighted that a template for the bulletins should be agreed ahead of time, and institutions should work together to establish the level of detail required and the types of maps, tables and diagrams that are most useful. Additionally, it is important to note that terminology such as "fairly likely" or "likely" and the implied differences between these terms, may not be universally accepted and understood internationally; this should ideally be clarified and terminology agreed such that it will be correctly understood by all actors involved. This could be achieved through training and discussion workshops involving all actors, and standalone FAQ documents or guidance for interpreting the information, which could be available for decision-makers during an emergency situation. In an ideal situation, there would also be a systematised chain of communication in order to cascade the information from high-level organisations to decision-makers and local communities, in a much faster and more organised way, that would also make the process of obtaining feedback and communicating with different actors much clearer and more efficient.

Recommendation 4: Operationalise the co-production of forecast bulletins by the previously identified producers, and the dissemination of forecast information and bulletins to intermediaries and users, to ensure that the information can be relied upon during future events.

In terms of the forecasting information provided, based on feedback, we would revise the way in which some aspects were presented in the bulletins. Throughout the two cyclones, the bulletins were regularly improved based on feedback from DFID and other partners. This included adding labels to figures and maps highlighting key points, including maps of the rainfall forecasts that are used as input to GloFAS, and discussing some of the background information on why the forecast had changed and how the movement of the cyclone was impacting the flood risk. Based on feedback and discussions post-event, we would further recommend including in each updated bulletin a brief summary of changes relative to the previous bulletin. This would provide a way for decision-makers to rapidly update their understanding of the situation, and for the team producing the bulletins to explain why any major changes have occurred. It could also be potentially useful to provide a national overview of the information, such as to provide the probability of severe flooding and the number of people likely to be affected across the country, which may be key information for high-level decision-makers.

A key knowledge gap identified during this process is in understanding who the users are and what information is required for each, alongside the best way to tailor the information to user needs and how to translate the forecasts into useful impact-focussed information. There is a desire to move towards impact-based forecasting, and indeed, future operationalisation of bulletins such as these should aim to incorporate this, and could consider the potential benefits or drawbacks of providing different bulletins tailored to different groups of users. Additionally, while rapid developments in automisation and artificial intelligence mean that it may in future be possible to generate bulletins such as these automatically and directly from forecasting centres, a key aspect of such bulletins is also the human element - the expert interpretation of the forecasts and the changing situation, and continuous dialogue between different institutions, decision-makers and other actors.

Recommendation 5: Provide forecast information that is tailored to the needs of the users: include impact-based forecasts, provide language translations, and engage in two-way communication between forecast producers and users to incorporate real-time information and respond to queries.

6.3. Effective forecast communication

6.3.1. Challenges

From the perspective of decision-makers and actors making use of the bulletins, and indeed those tasked with synthesising and disseminating the information, it was found that some of the scientific method behind producing the bulletins could be complex to convey and understand. In addition, where communications infrastructure was impacted during the disaster, there were technological limitations in how much complex information could be visualised in the field. The team producing the bulletins, in a similar way, found that a key challenge was the need to reduce the complexity of the information without losing the nuance of the forecast limitations. This is a well understood issue in environmental forecasting but remains a frustrating challenge. The bulletins were produced making several assumptions regarding the required level of complexity and without exact knowledge of who the forecast users were, or an understanding of their level of existing forecasting knowledge. Furthermore, to make a binary decision based on probabilistic forecast information is a known challenge, and implementing decision-making while accounting for uncertainty, both known and unknown, can be complex. A recent study by Arnal et al. [72] makes several recommendations for successful transition to probabilistic forecasting and decision-making, based on interviews with flood forecasters at the UK's Environment Agency, but with wider significance and applicability. For example, the need to provide "appropriate and custom designed training to all key players", including clear guidance on how to make decisions using new products, and for "everyone using the forecast products and systems […] to have a say in how the system will look and function, through a mutual design strategy", as any new system which does not reflect the complex decision-making landscape, may be mis- or under-used.

Our partners also highlighted that while the bulletins did provide some impact-based forecast information, the bulletins were not impact-focussed enough, and the population exposure information presented was challenging to interpret and overly precise. The GloFAS flood exceedance probability thresholds were also arbitrarily selected, potentially failing to identify at-risk zones that have a low flood forecast probability. There were some substantial challenges in conveying upcoming flood risk with the bulletins because there was no coverage of coastal or flash flooding, which is a key limitation when providing hazard and risk information for tropical cyclones. A lack of data available for both calibration and validation of forecast models [73,74] was also problematic. Access to more observation data would allow the models to be evaluated more thoroughly, but there are often barriers to accessing data, and across large parts of the world, data collection is scarce and data records contain significant gaps.

6.3.2. Lessons learnt and recommendations

When using a complex chain of environmental models to provide probabilistic information about upcoming flood hazard and risk, understanding who all the end users actually are, their level of knowledge regarding such forecasts, and what information is required on the ground for decision-making, is essential. The partners highlighted that it would be desirable to include a short dedicated 'impact summary' in the bulletins. This could provide information on key infrastructure in the region that may be at risk or may impact the response, such as hospitals and key transport routes. An improved method is also needed to communicate population exposure; one suggestion is that each person could be weighted by their probability of experiencing a flood threshold exceedance, favouring areas with a higher GloFAS probability of exceedance when the exposure is summed across localities. An alternative visualisation using this method is presented in Fig. 9, which gives a ranking of the most exposed localities and the range of exposure for multiple flood hazard return periods. Another suggested improvement could be to display multiple flood inundation return periods, alongside the GloFAS probability of exceedance. Beyond this, in order to better understand how reliable (and trustworthy) the forecast information provided is, information on how the forecast models typically perform in the affected region could be provided, alongside remotely sensed data of flood extent in near-real time, which would provide invaluable information on how the events were unfolding, and allow a real-time preliminary evaluation of how well the forecasts were predicting the event.

Recommendation 6: Share knowledge and understanding surrounding forecast uncertainty, quality and limitations, for all relevant actors, for example through training workshops and guidance documents.

The provision of better information in emergency bulletins relies upon the greater availability of data which is a major barrier to forecasting activities. Observational infrastructure networks need to be supported and their value appreciated. In addition, the need for better impact information provides an impetus to improve routine data collection on key infrastructure, exposure and vulnerability, which requires greater coordination between ministries and other national and local authorities alongside international organisations.

Recommendation 7: Greater coordination between ministries and other national and local authorities alongside international organisations is essential, particularly in order to improve routine data collection that would ensure the best emergency decisions can be taken using accurate and up-to-date information.

Currently, neither GloFAS or the flood inundation model are able to account for coastal or flash flooding, which is a key limitation. While these limitations were communicated in the bulletins, in future we hope to have forecast models that are able to forecast compound flooding from combined rainfall and coastal effects [75–77]. In addition, the shift towards whole Earth System modelling and forecasting means that in the next decade we should start to see multi-hazard forecasts available including all types of flooding, wind and other hazards such as landslides.

Recommendation 8: In the long-term, a move towards an Earth System approach to forecasting would allow a holistic inclusion of the relevant flood hazard and risk information from all sources of flooding (riverine, pluvial and storm surge), combined with other hazards arising from tropical cyclones such as wind damage, landslides and thunderstorms.

7. Conclusions

In order to take effective decisions, humanitarian and civil protection agencies need appropriate information on upcoming flood hazards and risk. In this paper we have critically evaluated the collaborative production of emergency bulletins for Cyclones Idai and Kenneth in Mozambique in support of the international humanitarian community. These were produced using global river flood forecasts from the Copernicus Emergency Management Service's Global Flood Awareness System (GloFAS), together with flood inundation modelling and impact risk assessment for population exposure estimates.

The provision of real time hazard and risk information in this way has provided a technically successful proof of concept with a positive real-world impact: information on different components of flood hazard and risk were integrated, provided in real time and informed decision-making. There is evidence that the bulletins supported critical actions such as sending an assessment team to the region most likely to be affected and considering the availability of hygiene kits, water treatment kits and tarpaulins ahead of the response.

The forecast information provided in the bulletins was evaluated and the interaction between the different components of the forecast chain discussed. While it is possible to predict the track of a tropical cyclone with relative certainty, the path of Idai was much more challenging to predict than that of Kenneth. Despite this, feedback from partners indicated that the uncertainty in Kenneth's track and associated flood risk in the earlier forecasts, which showed that the cyclone may make landfall in Tanzania, posed further challenges associated with a potential transboundary response. Evaluating the flood hazard and exposed population was challenging based on the data available, as post-event reports often indicate the total number of people affected by a wide range of impacts, including extreme wind, food insecurity and disease, while the exposure estimates provided in the bulletins were for river flooding only. While the evaluation indicated that the districts at greatest risk of flooding were successfully identified, increased collaboration with in-country partners could facilitate the provision of improved risk information, through access to more detailed data and information on population and infrastructure, that could better support decision-making.

So are we making the best use of the best science for humanitarian emergencies and resilience? There is a great potential value in using global operational forecasting models (such as GloFAS) for supporting the international humanitarian community to take actions for TCs in south-east Africa and elsewhere in the world. There is clear scope for

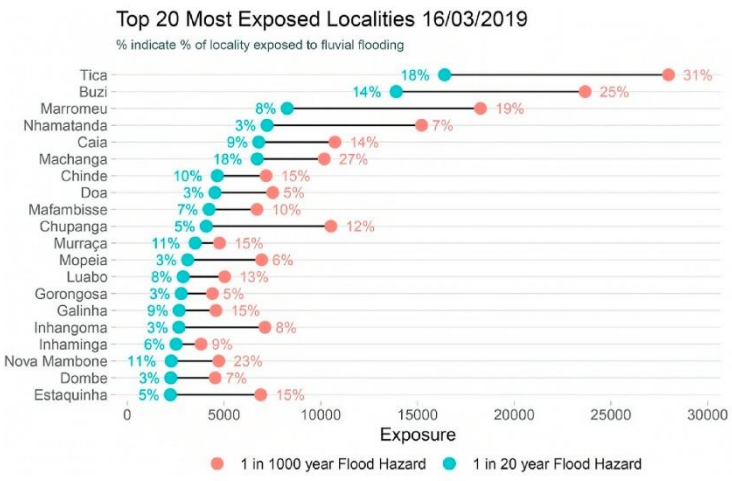


Fig. 9. Top 20 most exposed localities for Cyclone Idai on 16/03/2019. The blue circles indicate the exposure to the 20-year return period (5% AEP) flood inundation and red the exposure to the most extreme 1000-year (0.1% AEP) flood inundation. The percentage of the total locality population exposed to flooding is also shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

improving the provision of bulletins such as these - tailoring the information and making it clearer and more concise. However there is a clear need to not only work more closely with the mandated national authorities responsible for disseminating forecast and warning information, to improve the two-way sharing of information that would have the most impact on the ground, but also to support capacity development for a national, operational "End-to-End Multi-Hazard Early Warning System in the context of disaster risk management" [14]. In the interim, forecast information produced by international organisations can be a useful tool to support anticipatory action and complement current forecasting capacity from national authorities, but must navigate the "murky landscape" of national and international mandates, and capacities and collaborations for forecasting, early warning and anticipatory action. To do this successfully, much closer collaboration between international organisations and national authorities, forecast producers and forecast users, is essential. Section 6 outlined key recommendations for the future production of forecast bulletins, focussing on co-production, operationalisation and communication.

In order to be truly impactful, forecast information must not only inform the decisions taken rather than distracting from key actions, but must be co-produced in socially and institutionally embedded ways from the very beginning. The flood bulletins discussed in this paper were produced by University researchers responding to a request for help; responsively building a collaboration in order to provide the international humanitarian community with real-time information that wasn't already available. So how can this methodology be operationalised so that users can begin to rely on this information, and to trust the scientific method and indeed those scientists and institutions providing the information? Certainly University researchers do not have the required 24/7 operational capabilities to reliably produce such forecasts for every extreme event. This work has shown the technical requirements for producing such information, but the processes for producing these bulletins should be developed in a way that provides international cooperation to complement existing capacity (in line with the Sendai Framework [5]), while working towards the goal of building existing capacity of the mandated national authorities in a sustainable way [10,71,78,79].

Forecast producers and those using early warnings need to spend more time together understanding each other and move beyond the 'loading-dock approach' of science towards genuine co-production that counters the idea of technocratic solutions in which scientists should be isolated from decision makers [10,78]. To better support early humanitarian action with the best science and better integrate forecast and impact information produced both nationally and internationally, there is a clear requirement for embedded collaboration between forecast producers (national, regional and international, mandated and research institutes), intermediaries (governments, media, NGOs and research institutes) and users (governments, humanitarians, citizens, private sector, local leaders) [71]. Having a One Voice Principal for early warnings of natural hazards is important, but as of now the question remains: who will take responsibility for delivering reliable, tailored and comprehensive information that integrates all relevant aspects of forecast, risk and impact information? Who should?

Data availability

Real-time GloFAS forecast products are freely available at www.globalfloods.eu, and GloFAS data can be obtained through the dedicated GloFAS data service (https://confluence.ecmwf.int/display/COPSRV/04.+GloFAS+services). ECMWF ENS forecast data are available through the TIGGE archive after a short delay (https://www.ecmwf.int/en/research/projects/tigge), and HRES data can be accessed through ECMWF's Meteorological Archival and Retrieval System (MARS), subject to licensing. The IMERG satellite rainfall data can be downloaded from NASA (https://pmm.nasa.gov/data-access/downloads/gpm) and the observed tropical cyclone data from IBTrACS (https://www.ncdc.noaa.gov/ibtracs/). The LISFLOOD-FP code is freely available from http://www.bristol.ac.uk/geography/research/hydrology/models/lisflood/, and the global flood inundation maps used in this study are available upon request from Dr Jeffrey Neal or Dr Laurence Hawker.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Ervin Zsoter (ECMWF), Hanne Roden (German Red Cross), Janio Dambo (CVM) and Jonathan Stone (DFID/REAP, IFRC) for their invaluable advice and input to the bulletins, Calum Baugh (ECMWF) for providing the raw GloFAS forecast data to the University of Bristol both in real-time during the two cyclones and for the analysis presented here, Fathom for sharing Fathom Global flood hazard (inundation) data and Kevin Hodges (NCAS) for use of the tropical cyclone tracking scheme. We are also grateful to Erin Coughlan de Perez (Red Cross Climate Centre), Gemma Connell (UN OCHA ROSEA),Mussa Mustafa (INAM), Nyree Pinder, John Ward and Jake Brown (UK Met Office) for many interesting and insightful discussions related to this work, to José Malanco, Isac Filimone and David Mucambe (DNGRH) and Joaquim Comboio (Ara-Sul) for providing the national hydrological bulletins, and to Benedita Nhambiu and Rui da Maia (UDM) for their support during the discussion meeting in Maputo. The authors acknowledge support from the UK's Natural Environment Research Council (NERC) and Department for International Development (DFID) under three projects within the Science for Humanitarian Emergencies and Resilience (SHEAR) research programme; Hannah Cloke, Liz Stephens, Andrea Ficchi, Linda Speight and Sara de Wit through FATHUM (Forecasts for AnTicipatory HUManitarian action, grant number NE/P000525/1), Rebecca Emerton through PICSEA (Predicting Impacts of Cyclones in South East Africa, grant number NE/S005897/1), and Jeffrey Neal and Laurence Hawker through HYFLOOD (Next generation flood hazard mapping for the African Continent at hyper-resolution, grant number NE/S006079/1). A SHEAR Integration Project is also underway, (Learning from Cyclone Idai and Kenneth: Presenting and using flood forecast information to support humanitarian response), involving FATHUM, HYFLOOD, the DFID-GFDRR Challenge Fund 2 Project and ECMWF. Christel Prudhomme and Shaun Harrigan were supported by the Copernicus Emergency Management Service – Early Warning Systems (CEMS-EWS; GloFAS). Nicholas Klingaman was funded by a NERC Independent Research Fellowship (NE/L010976/1).

Appendix. Supplementary data

This appendix provides an example of one of the emergency flood bulletins produced for Cyclone Kenneth on 26th April 2019.

University of Reading · Copernicus Europe's eyes on Earth · ECMWF · University of Bristol

Flood hazard & impact Emergency Report - Cyclone Kenneth

Event start:	Expected 25 April 2019	Report date (first):	24 April 2019
Area	Mozambique, Tanzania	Update #1:	26 April 2019

Key points

Cyclone Kenneth made landfall in northern Mozambique approximately 90km north of Pemba (near the town of Olumbua) at ~18:00 local time on 25 April (source: WMO). It has significantly weakened since landfall, and is no longer "hurricane" strength, but will bring significant rainfall and flood hazard over the coming days, particularly in the Cabo Delgado province.

Meteorological forecast

- The cyclone has "stalled" over the Cabo Delgado province of northern Mozambique, after making landfall on 25 April, and is forecast to remain mostly over this region for at least the next 2 days.
- The stalling of the cyclone means that significant amounts of rain will fall over a more localised region, rather than a larger region further inland. This is likely to cause significant and impactful flooding in Cabo Delgado.
- ECMWF forecasts show a signal of likely high rainfall also for Southern Tanzania over the next 10 days. Thus, the evolution of flooding likelihood for this larger region should be monitored over the next 5-10 days
- Predicted rainfall totals are reduced slightly since forecasts issued on 24 April, but are still forecast to be in excess of 500mm (over the 5-day period 26-30 April) over the Cabo Delgado province (ECMWF).

Flood forecast

- GloFAS forecasts (issued 26 April) indicate that river flows will increase beyond the severe alert threshold (20-year return period) with moderate to high probability, on the Messalo, Montepuez and Megaruma rivers in the Cabo Delgado province.
- Exceedance of (at least) the high alert threshold (5-year return period) is estimated from 26 (Messalo), 27 (Montepuez) and 29 (Megaruma) April onwards. The flood peak for these rivers is currently estimated to occur around 30 April.
- GloFAS flood probabilities have decreased to the north and west of Cabo Delgado province, compared to previous forecasts issued on 24 and 25 April, due to the stalling of the cyclone.

Flood Impact

- Exposure is for river flow risk only (i.e. excludes urban, storm surge and windstorm risk). Out of the estimated ~700k people at risk, over 65k people across Cabo Delgado and Nampula are exposed to river flooding of at least 10% probability, and more than 14k people are exposed in areas with at least 50% chance of river flooding
- In Cabo Delgado province the districts most at risk are Pemba (~10k people) and Mecufi (>5k people). Significant exposure (1k-4k each) is forecast in Macomia, Mueda, Muidumbe, Ancuabe, Quissanga, Montepuez, Mocimboa da Praia, Chiúre, Meluco and Balama.
- In Nampula province the districts most at risk are Mossuril, Monapo and Memba (4k-6k people each). Significant exposure (1k-4k people each) is forecast in Nacala Velha, Muecate, Meconta, Namapa.

Note: GloFAS is designed to simulate large scale hydrological systems, so predictions for smaller watercourses should be evaluated with caution. GloFAS also does not simulate dam release or dam breaks which could be a major problem in the affected region. Estimates of exposure only account for river flooding over the next 30 days.



Figure 1 Mozambique provinces (left) and inset showing the main rivers and towns of Cabo Delgado, expected to be the most likely area to be impacted by flooding from Cyclone Kenneth, (right)

Part 1 - Flood hazard briefing - University of Reading / ECMWF

Background

Cyclone Kenneth formed to the north of Madagascar on 23 April 2019, moving south-west towards northern Mozambique. The cyclone impacted Comoros overnight on 24-25 April, resulting in at least 3 deaths and widespread power outages, according to media reports. It made landfall in northern Mozambique at ~18:00 local time on 25 April, north of Pemba, as a category 4 (equivalent) cyclone with winds of up to 220km/h (source: Joint Typhoon Warning Centre JTWC). The cyclone has weakened significantly after making landfall, and has stalled over the Cabo Delgado province of northern Mozambique. The threat from storm surge has now passed, and the remaining hazard will be from significant amounts of rainfall over a localised region, with flooding predicted to begin in several rivers from 26 April onwards.

Meteorological Forecast Summary

ECMWF's Extreme Forecast Index (EFI) for 26-30 April (Figure 2a) indicates high probabilities of extremely high rainfall (red/orange) compared to typical conditions for the region. This forecast indicates the area most likely to be severely impacted by high rainfall, based on probabilistic forecasts from ECMWF. The ECMWF high-resolution single forecast run (not probabilistic) indicates that >300mm of rainfall (dark blue) is expected across much of the Cabo Delgado province (total rainfall forecast for the 5-day period 26-30 April), and potentially >500mm (orange) in some places (Figure 2b). The majority of the rainfall is forecast to occur over the Cabo Delgado province over the next 5 days, after which the forecast is much more uncertain, due to uncertainty in the direction in which the remnants of cyclone Kenneth will eventually move (ECMWF probabilistic track forecast). However, ECMWF forecasts show a signal of likely extremely high rainfall also for Southern Tanzania over the next 10 days. Thus, the evolution of flooding likelihood for this larger region should be monitored over the next 5-10 days. The JTWC also indicate the large spread in the track forecasts beyond 24 hours, and note that they are monitoring the system for any signs of the cyclone regenerating.

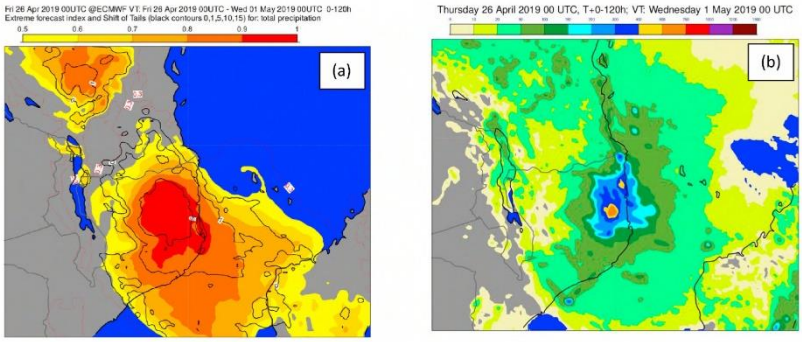


*Figure 2: **(a)** ECMWF Extreme Forecast Index, indicating anomalous rainfall conditions over the period 26-30 April 2019. Source: ECMWF, www.ecmwf.int. Forecast issued 26-04-2019 00 UTC. **(b)** ECMWF high resolution deterministic rainfall forecast for total rainfall accumulation (mm) across the period 26-30 April 2019. Source: ECMWF, www.ecmwf.int. Forecast issued 26-04-2019 00 UTC.*

Current hydrological situation and GloFAS Flood Forecasts

GloFAS forecasts indicate that rivers in the Cabo Delgado province are likely to see flows exceeding the severe alert threshold (see Table 1 and Figure 3), with the Messalo, Montepuez and Megaruma rivers indicating 43-63% probability of exceeding the severe alert threshold (Figure 3), and much higher probabilities (>69%) of exceeding the high alert threshold (see Figure 5 and Figure 6).

The high alert threshold is forecast to be exceeded today, 26 April, in the Messalo river, and during the 27-28 April for the Montepuez and Megaruma rivers. Flooding is expected to peak in all three rivers on 30 April, with flow receding slowly through mid-May. Probabilities of exceeding the severe alert threshold have slightly reduced since the previous report issued 24 April, likely due to the slight reduction in rainfall totals over the coming days in the

ECMWF probabilistic forecasts compared to those issued on 24 April; however, probabilities of severe flooding remain moderate to high (>50%). Hydrographs for the three aforementioned rivers in the Cabo Delgado province are shown in Figure 4 to Figure 6. Lower probability (10%-30%) of ensemble streamflow predictions to exceed the severe threshold (20 year return period) persist over a larger area (than these three river basins), along the coast of Northern Mozambique (Nampula, Cabo Delgado and Niassa provinces) and Tanzania (Mtwara and Lindi regions), as shown in the GloFAS map in Figure 3.

For a map of locations mentioned here, please see Figure 1.

Table 1: Correspondence of flood alert level thresholds with return period, calculated based on GloFAS climatology, colour in GloFAS maps and hydrographs and hazard description (source: GloFAS; www.globalfloods.eu).

Flood alert level / GloFAS threshold name	Return period	Colour	Hazard description
Low	1.5-year return period	Green (-)	Water levels higher than normal or up to bankfull condition but no flooding is expected
Medium	2-year return period	Yellow (-)	Bankfull condition or slightly higher expected; potential (minor) flooding
High	5-year return period	Red	Significant flooding is expected
Severe	20-year return period	Purple	Severe flooding is expected

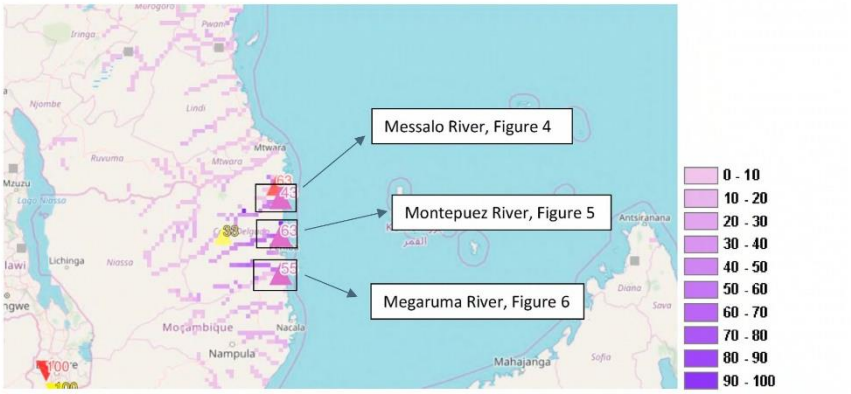


Figure 3: GloFAS forecast indicating rivers likely to see floods exceeding the severe alert threshold (probability highlighted in purple). Source: GloFAS, www.globalfloods.eu. Forecast issued 26-04-2019 00 UTC.

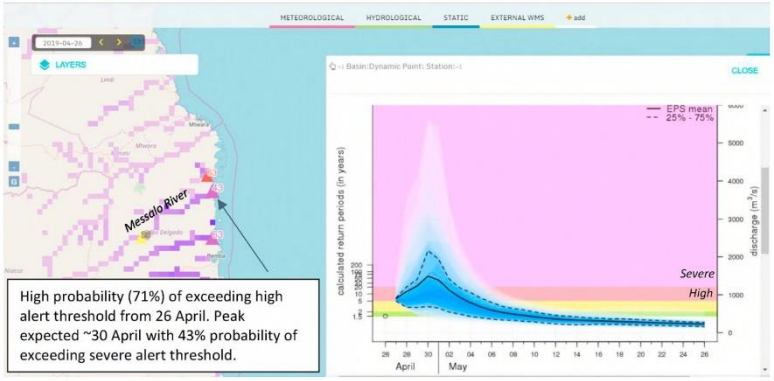


Figure 4: Forecast hydrograph for warning point (lat/lon -11.65/40.45) on the Messalo river in the Cabo Delgado province of Mozambique. Source: GloFAS, www.globalfloods.eu. Forecast issued 26-04-2019 00 UTC. Note: the y-axis indicates that the forecast exceeds the severe alert threshold (1 in 20 yr return period) - any return period values above this are extrapolated from a simulated climatology and are unlikely to be realistic, and as such, the y-axis is intentionally not extended.

Increasing trend above the severe alert threshold (43% exceedance probability)

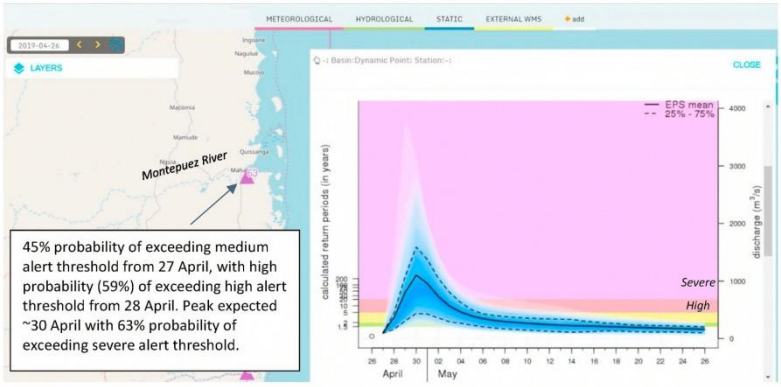


Figure 5: Forecast hydrograph for warning point (lat/lon -12.55/40.45) on the Montepuez River in the Cabo Delgado province of Mozambique. Source: GloFAS, www.globalfloods.eu. Forecast issued 26-04-2019 00 UTC.

Increasing trend above the medium alert threshold (63% exceedance probability)

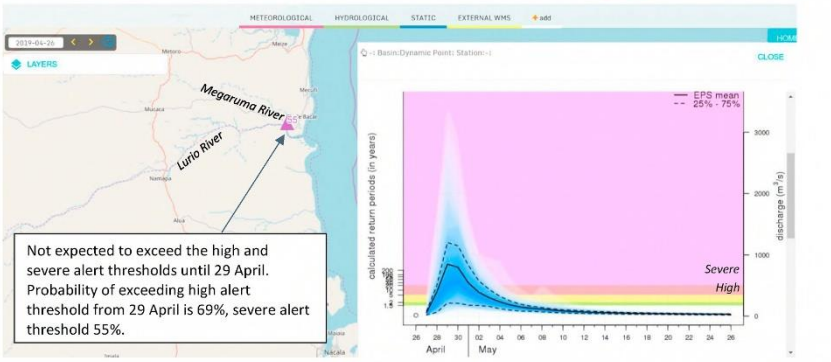


Figure 6 Forecast hydrograph for warning point (lat/lon -13.45/40.45) on the Megaruma River in the Cabo Delgado province of Mozambique. Source: GloFAS, www.globalfloods.eu. Forecast issued 26-04-2019 00 UTC. Note: the warning point appears to be located between the Megaruma and Lurio rivers, due to the grid used in GloFAS. This has been cross-checked with other information, and this warning point is confirmed to be for the Megaruma river.

Increasing trend above the medium alert threshold (55% exceedance probability)

Flood Hazard Emergency Report - Cyclone Kenneth 5 *www.globalfloods.eu*

University of Reading Copernicus Europe's eyes on Earth ECMWF University of BRISTOL

Part 2 - Flood hazard forecasts and population exposure - Bristol University

Background

Cyclone Kenneth made landfall in Mozambique on the afternoon of the 25/09/2019. This report maps flood zones from a global scale flood inundation model and forecasts of potential flooding over the next 30 days. This information is used to make estimates of where at-risk population from flooding are located. Exposure estimates are for fluvial flooding (e.g. flooding directly from rivers overtopping their banks) and may be lower than estimates from other providers that include multiple hazards (e.g. windstorm, pluvial (urban) flooding and coastal flooding).

Exposed population

Table 2 estimates the exposed population to river flooding exceeding the severe flood level (1 in 20-year return period) for given significant probability levels over the next 30 days given forecasted river flows from the ECMWF GloFAS system overlaid on the 1 in 250 year flood zone from a global scale flood inundation model. The table is split into districts with the exposure estimated when the probability of river flooding exceeding the severe flood level (1 in 20-year return period) over the next 30 days exceeds 50% (high probability) and 10% (low to medium probability). These population figures do not include any IDP (internally Displaced People) populations.

Across Cabo Delgado and Nampula over 65k people are forecasted to be exposed to greater than 10% probability of flooding and over 14k people are exposed to greater than 50% change of flooding. In Cabo Delgado province the districts most at risk are Pemba (~10k people) and Mecufi (>5k people). Significant exposure (1k-4k each) is forecasts in Macomia, Mueda, Muidumbe, Ancuabe, Quissanga, Montepuez, Mocimboa da Praia, Chiúre, Meluco and Balama. In Nampula province the districts most at risk are Mossuril, Monapo and Memba (4k-6k people each). Significant exposure (1k-4k people each) is forecasts in Nacala Velha, Muecate, Meconta, Namapa. Significant numbers of people exposed to an extreme flood in Bandar and Nangua in Pemba district and to the south of Mecufi district.

Forecast exposure by district is presented below in Table 2. Exposure maps for each district/town are available on request with selected high-risk areas presented here.

Table 2: Population exposed by district (Mozambique) to the flood hazard zone for the severe alert threshold (1 in 20 year return period) where GloFAS forecasts a probability greater than 10% or 50% of exceeding the severe flood level. Locations are ranked by population exposed to 1 in 250 year flood zone when GloFAS probability of exceedance over 50%.

Province	District	Population exposed to 1 in 250 year flood zone when GloFAS probability of exceedance over 10%	Population exposed to 1 in 250 year flood zone when GloFAS probability of exceedance over 50%
Cabo Delgado	Pemba	9952	3164
Cabo Delgado	Mecufi	5386	4213
Nampula	Mossuril	5223	0
Nampula	Monapo	4724	0
Nampula	Memba	4549	0
Cabo Delgado	Macomia	3906	338
Cabo Delgado	Mueda	3631	0
Nampula	Erati	3467	0
Cabo Delgado	Muidumbe	3430	0
Cabo Delgado	Ancuabe	3184	2475
Cabo Delgado	Quissanga	2805	2805
Cabo Delgado	Montepuez	2519	0
Cabo Delgado	Mocimboa da Praia	2204	0
Nampula	Nacala Velha	1994	0
Nampula	Muecate	1815	0
Cabo Delgado	Chiúre	1644	853
Nampula	Meconta	1439	0
Cabo Delgado	Meluco	1356	576
Cabo Delgado	Balama	1336	0
Nampula	Namapa	1291	0
	Total	65852	14424

Flood maps

An overview map of the area of interest in Mozambique is shown in Figure 7. Selected areas that are most at risk are shown in Figure 8 to Figure 9, with other areas available on request (see contact details below). The probability of flooding exceeding the severe flood level (20-year return period) over the next 30 days is overlaid with the extent of a 1 in 250 year flood event (worst-case scenario) is shown in greens/purple. The 1 in 250 year flood extent is shown in these maps to depict an extreme flooding scenario. Potentially exposed population is shown in oranges/red with the redder the colour showing a higher density of exposed population.

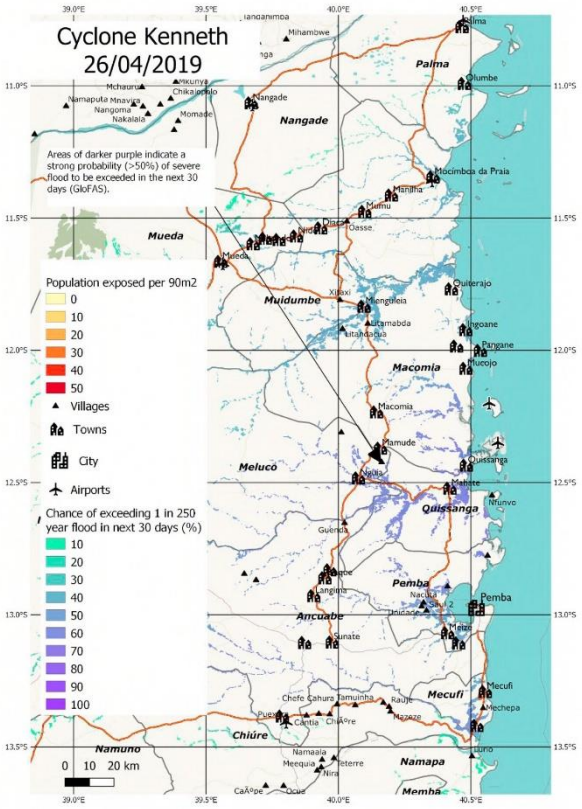


Figure 7: Flood hazard map of Mozambique (mostly Cabo Delgado Province). The probability of flooding exceeding the severe flood level (250-year return period) over the next 30 days is shown in purple. Potentially exposed population is shown in red by district.

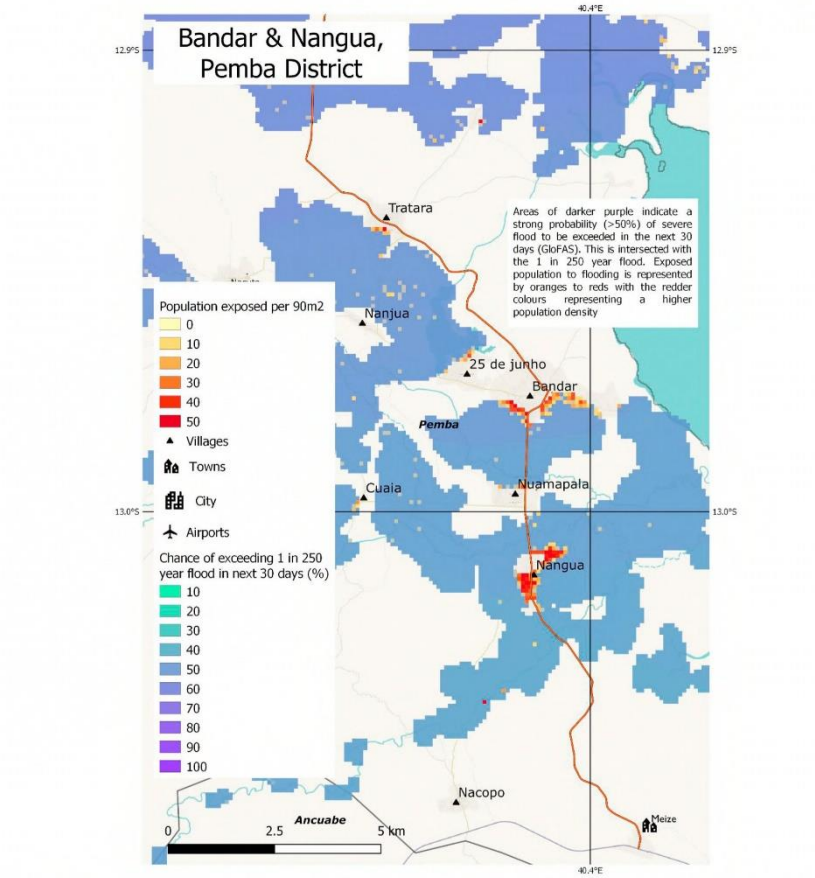


Figure 8: Flood hazard map around the villages of Bandar and Nangua, Pemba district. The probability of flooding exceeding the severe flood level (20-year return period) over the next 30 days for the extent of a 1 in 250 year flood event is shown in greens/purple. Potentially exposed population is shown in oranges/red with the redder the colour showing a higher density of exposed population.

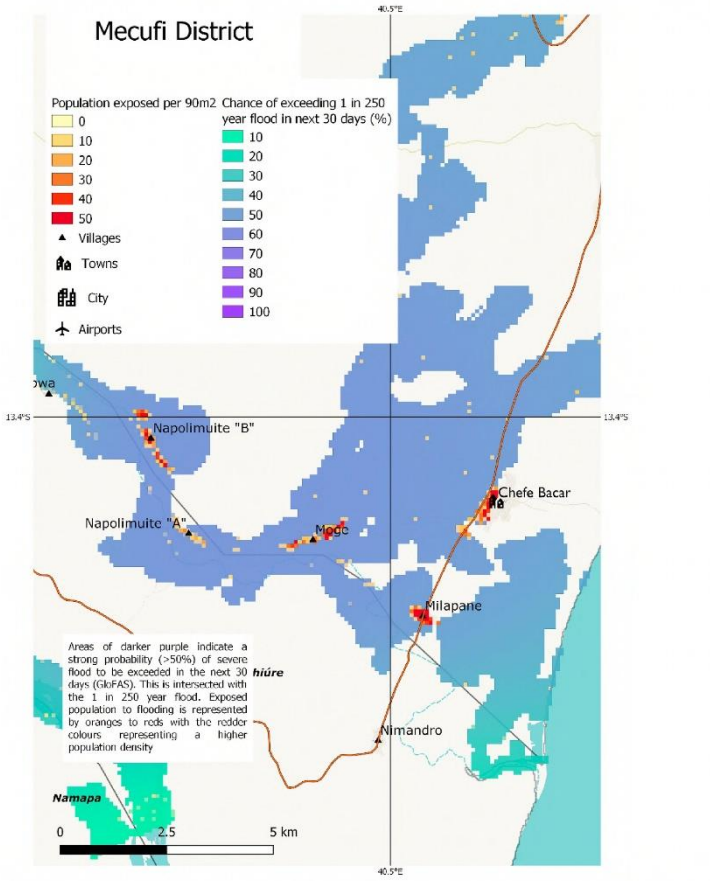


Figure 9: Flood hazard map around the town of Chefe Bacar, Mecufi district. The probability of flooding exceeding the severe flood level (20-year return period) over the next 30 days for the extent of a 1 in 250 year flood event is shown in greens/purple. Potentially exposed population is shown in oranges/red with the redder the colour showing a higher density of exposed population.

Part 3 - Additional Information

Contributors

Part 1: Main report author: Dr Rebecca Emerton (National Centre for Atmospheric Science, University of Reading, rebecca.emerton@reading.ac.uk) and Dr Andrea Ficchì (University of Reading); other contributors: Shaun Harrigan, Christel Prudhomme, Damien Decremer, Calum Baugh, Ervin Zsoter (ECMWF); Hannah Cloke, Liz Stephens, Helen Titley (University of Reading).

Part 2: Dr Jeffrey Neal, Dr Laurence Hawker, Prof Paul Bates (School of Geographical Science, University of Bristol, Bristol, UK. BS8 1SS), Dr Andrew Smith, Dr Chris Sampson (Fathom Global, Temple Meads, Engine Shed, Bristol, UK. BS1 6QH). We acknowledge the GloFAS team at ECMWF for provision of forecast exceedance probabilities.

Data sources

Part 1: ECMWF and GloFAS forecasts; GloFAS is part of the Copernicus Emergency Management Service (CEMS). Contributors: ECMWF / JRC / University of Reading. Please see http://www.globalfloods.eu. We also acknowledge support from the NERC/DFID SHEAR Programme, projects FATHUM (Forecasts for Anticipatory Humanitarian Action) and PICSEA (Predicting Impacts of Cyclones in South East Africa).

Part 2: Source data are located at the University of Bristol. Please contact Dr Jeffrey Neal (j.neal@bristol.ac.uk) or Dr Laurence Hawker (Laurence.hawker@bristol.ac.uk) for further information. Population data from the Facebook Connectivity Lab and Center for International Earth Science Information Network - CIESIN - Columbia University. 2016. High Resolution Settlement Layer (HRSL). https://www.ciesin.columbia.edu/data/hrsl/ (Accessed 23/03/2019)

Correspondence:

Please direct technical enquiries to Prof Christel Prudhomme (christel.prudhomme@ecmwf.int) or Prof Hannah Cloke (h.l.cloke@reading.ac.uk). For all other enquiries, please contact Philip Rundell (philip-rundell@dfid.gov.uk) or Dr Rosalind West (r-west@dfid.gov.uk)

References

[1] UN OCHA, Cyclones Idai and Kenneth. Southern and Eastern Africa (ROSEA). https://www.unocha.org/southern-and-eastern-africa-rosea/cyclones-idai-and-kenneth, 2019. (Accessed 21 November 2019).

[2] INGC, Multi-sectoral rapid assessment of post-cyclone Idai, 14 districts in Sofala and Manica provinces, Mozambique. https://reliefweb.int/sites/reliefweb.int/files/resources/Mozambique MRA 20190426 EN def.pdf, 2019. (Accessed 11 February 2020).

[3] T. Jurlina, C. Baugh, F. Pappenberger, C. Prudhomme, Flood hazard risk forecasting index (FHRI) for urban areas: the Hurricane Harvey case study, Meteorol. Appl. (2019) 1–10, https://doi.org/10.1002/met.1845.

[4] Meque, A., Ficchi, A., Emerton, R., Cloke, H., Speight, L., Dambo, J., Stephens E. and Coughlan de Perez, E.: A multi-source approach to understanding tropical cyclone risk in Mozambique, under review.

[5] UNISDR Sendai framework for disaster risk reduction 2015–2030. www.preventionweb.net/files/43291_sendaiframeworkfordrren.pdf. (Accessed 6 February 2020).

[6] A. Zia, C.H. Wagner, Mainstreaming early warning systems in development and planning processes: multilevel implementation of Sendai framework in Indus and Sahel, Int. J. Disasters Risk Sci. 189–199 (2015).

[7] E. Coughlan de Perez, B. van den Hurk, M.K. van Aalst, I. Amuron, D. Bamanya, T. Hauser, B. Jongman, A. Lopez, S. Mason, J. Mendler de Suarez, F. Pappenberger, A. Rueth, E. Stephens, P. Suarez, J. Wagemaker, E. Zsoter, Action-based flood forecasting for triggering humanitarian action, Hydrol. Earth Syst. Sci. 20 (2016) 3549–3560.

[8] RCCC, Second humanitarian distribution in Uganda under forecast-based financing, red Cross climate centre. https://www.climatecentre.org/news/723/second-humanitarian-distribution-in-uganda-under-forecast-based-financing. (Accessed 5 December 2019).

[9] A. Tall, et al., Using seasonal climate forecasts to guide disaster management: the red Cross experience during the 2008 west Africa floods, Int. J. Geophys. 1–12 (2012).

[10] W.D. Cash, J.C. Borck, A.G. Patt, Countering the loading-dock approach to linking science and decision making: comparative analysis of El Nino/Southern Oscillation (ENSO) Forecasting Systems, Sci. Technol. Hum. Val. 31 (4) (2006) 465–494.

[11] M. Mahony, M. Hulme, Epistemic geographies of climate change: science, space and politics, Prog. Hum. Geogr. 42 (3) (2018) 395–424.

[12] M. Lahsen, Distrust and participation in international science and environmental decision making: knowledge gaps to overcome, in: M. Pettinger (Ed.), The Social Construction of Climate Change, Ashgate, Aldershot, UK, 2007, pp. 173–196.

[13] Skelton, et al., Customising global climate science for national adaptation: a case study of climate projections in UNFCCC's National Communications, Environ. Sci. Pol. 101 (2019) 16–23.

[14] WMO, Reducing Vulnerability to Extreme Hydro-Meteorological Hazards in Mozambique after Cyclone Idai, WMO Mission Report Following Tropical Cyclone Idai (29 April – 7 May 2019), 2019. https://library.wmo.int/doc_num.php?explnum_id=6259 Last access. (Accessed 13 July 2019).

[15] WHO, Tropical Cyclone Idai, Mozambique, Donor Alert, 2019. https://www.who.int/emergencies/cyclone-idai/mozambique-donor-alert.pdf?ua=1#:~:text=A%20category%204%20tropical%20cyclone,6%20provinces%20with%20severe%20flooding.&text=The%20President%20of%20Mozambique%20declared,people%20need%20urgent%20humanitarian%20assistance. (Accessed 13 July 2019).

[16] L. Alfieri, S. Cohen, J. Galantowicz, G.J. P. Schumann, M.A. Trigg, E. Zsoter, C. Prudhomme, A. Kruczkiewicz, E. Coughlan de Perez, Z. Flamig, R. Rudari, H. Wu, R.F. Adler, R.G. Brakenridge, A. Kettner, A. Weerts, P. Matgen, S.A.K. M. Islam, T. de Groeve, P. Salamon, A global network for operational flood risk reduction, Environ. Sci. Pol. 84 (2018) 149–158, https://doi.org/10.1016/j.envsci.2018.03.014.

[17] G. Schumann, G. Brakenridge, A. Kettner, R. Kashif, E. Niebuhr, Assisting flood Disaster Response with Earth Observation data and products: a critical assessment, Rem. Sens. 10 (8) (2018) 1230.

[18] Flood Forecasting Centre. http://www.ffc-environment-agency.metoffice.gov.uk/services/FGS_User_Guide.pdf. (Accessed 6 February 2020).

[19] R.H. Carr, B. Montz, K. Semmens, K. Maxfield, S. Hoekstra, E. Goldman, Motivating action under uncertain conditions: enhancing emergency briefings during coastal Storms, Weather Clim. Soc. 8 (2016) 421–434, https://doi.org/10.1175/WCAS-D-16-0028.1.

[20] E. Stephens, H. Cloke, Improving flood forecasts for better flood preparedness in the UK (and beyond), Geogr. J. 180 (4) (2014) 310-316, https://doi.org/10.1111/geoj.12103.

[21] F. Pappenberger, H.L. Cloke, C.A. Baugh, Cartograms for use in forecasting weather driven natural hazards, Cartogr. J. 56 (2) (2019) 134–145, https://doi.org/10.1080/00087041.2018.1534358.

[22] L. Speight, S. Cole, R. Moore, C. Pierce, B. Wright, B. Golding, M. Cranston, A. Tavendale, J. Dhondia, S. Ghimire, Developing surface water flood forecasting capabilities in Scotland: an operational pilot for the 2014 Commonwealth Games in Glasgow, J. Flood Risk Manag. 11 (2018) S884–S901, https://doi.org/10.1111/jfr3.12281.

[23] R.A. Pielke Jr., Asking the right questions: atmospheric sciences research and societal needs, Bull. Am. Meteorol. Soc. 78 (2) (1997) 255–264.

[24] E. Stephens, T.L. Edwards, D. Demeritt, Communicating probabilistic information from climate model ensembles—lessons from numerical weather prediction, WIREs Clim. Chang. 3 (5) (2012) 409–426, https://doi.org/10.1002/wcc.187.

[25] E. Coughlan de Perez, B. van den Hurk, M.K. van Aalst, B. Jongman, T. Klose, P. Suarez, Forecast-based financing: an approach for catalyzing humanitarian action based on extreme weather and climate forecasts, Nat. Hazards Earth Syst. Sci. 15 (2015) 895–904.

[26] Météo France La Reunion Archives Advisories List. http://www.meteo.fr/temps/domtom/La_Reunion/webcmrs9.0/anglais/index.html. (Accessed 27 November 2019).

[27] DNGRH: Boletim Hidrologico Nacional. Ministério das obras públicas, habitaçao e recursos hidricos, Direçao Nacional de Gestao de Recursos Hidricos, Ediçao N° 055-142. 01/02/2019 – 30/04/2019.

[28] ERCC, Tropical cyclone Idai impact overview, European Commission emergency response coordination centre (ERCC) DG ECHO daily map 18/3/2019. https://reliefweb.int/sites/reliefweb.int/files/resources/ECDM_20190318_TC_IDAI_update.pdf. (Accessed 5 February 2020).

[29] World Heath Organization Africa, Weekly Bulletin on Outbreaks and Other Emergencies, 26 May 2019. https://reliefweb.int/disaster/ep-2019-000026-moz. (Accessed 6 February 2020).

[30] GFDRR, Mozambique Cyclone Idai Post Disaster Needs Assessment, 2019. https://www.gfdrr.org/en/publication/mozambique-cyclone-idai-post-disaster-needs-assessment full report. (Accessed 14 February 2020).

[31] H.L. Cloke, F. Pappenberger, Ensemble flood forecasting: a review, J. Hydrol. 375 (3-4) (2009) 613–626, https://doi.org/10.1016/j.jhydrol.2009.06.005.

[32] INGC: Disaster risk management in Mozambique, presentation at the FATHUM annual meeting at the Universidade Tecnica de Mocambique (UDM), September 2019.

[33] FEWS NET, Improving Flood Risk Management through GIS Flood Tool (GFT), 2018. https://fews.net/sites/default/files/Mozambique_ND_Flyer_Snapshot.pdf Last. (Accessed 13 July 2019).

[34] UKRI, Predicting Impact of Cyclones in South-East Africa (PICSEA), 2018. https://gtr.ukri.org/projects?ref=NE%2FS005897%2F1 Last. (Accessed 13 July 2019).

[35] R. Emerton, Predicting the Impacts of Tropical Cyclones - when Should We Trust the Forecasts, and when Should We Not? University of Reading, Department of Meteorology, 2019. https://blogs.reading.ac.uk/weather-and-climate-at-reading/2020/predicting-the-impacts-of-tropical-cyclones-when-should-we-trust-the-forecasts and when should we not/Last. (Accessed 13 July 2019).

[36] H.A. Titley, R.L. Bowyer, H.L. Cloke, A global evaluation of multi-model ensemble tropical cyclone track probability forecasts, Q. J. R. Meteorol. Soc. 146 (2020) 531–545, https://doi.org/10.1002/qj.3712.

[37] K.I. Hodges, A general method for tracking analysis and its application to meteorological data, Mon. Weather Rev. 122 (1994) 2573–2586, https://doi.org/10.1175/1520-0493(1994)122,2573: AGMFTA.2.0.CO;2.

[38] K.I. Hodges, Feature tracking on the unit sphere, Mon. Weather Rev. 123 (1995) 3458–3465, https://doi.org/10.1175/1520-0493(1995)123,3458:FTOTUS.2.0.CO;2.

[39] K.I. Hodges, Adaptive constraints for feature tracking, Mon. Weather Rev. 127 (1999) 1362–1373, https://doi.org/10.1175/1520-0493(1999)127,1362:ACFFT.2.0.CO;2.

[40] K.I. Hodges, N.P. Klingaman, Prediction errors of tropical cy- clones in the western North Pacific in the Met Office global forecast model, Weather Forecast. 34 (2019) 1189–1209, https://doi.org/10.1175/WAF-D-19-0005.1.

[41] K.R. Knapp, M.C. Kruk, D.H. Levinson, H.J. Diamond, C.J. Neumann, The international best track archive for climate stewardship (IBTrACS) project: unifying tropical cyclone data, Bull. Am. Meteorol. Soc. 91 (2010) 363–376, https://doi.org/10.1175/2009BAMS2755.1.

[42] F. Vitart, J.L. Anderson, W.F. Stern, Simulation of interannual variability of tropical storm frequency in an ensemble of GCM integrations, J. Clim. 10 (1997) 745–760, https://doi.org/10.1175/1520-0442(1997)010<0745:SOIVOT>2.0.CO;2.

[43] F. Vitart, D. Anderson, T. Stockdale, Seasonal forecasting of tropical cyclone landfall over Mozambique, J. Clim. 16 (2003) 3932–3945, https://doi.org/10.1175/1520-0442(2003)016<3932:SFOTCL>2.0.CO;2.

[44] S.C. Peatman, N.P. Klingaman, K.I. Hodges, Tropical cyclone-related precipitation over the northwest tropical pacific in Met Office global operational forecasts, Weather Forecast. 34 (2019) 923–941, https://doi.org/10.1175/WAF-D-19-0017.1.

[45] L. Guo, N.P. Klingaman, P.L. Vidale, A.G. Turner, M.-E. Demory, A. Cobb, Contribution of tropical cyclones to atmospheric moisture transport and rainfall over East Asia, J. Clim. 30 (2017) 3853–3865, https://doi.org/10.1175/JCLI-D-16-0308.1.

[46] G. Huffman, D. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, P. Xie, Integrated multi-satellitE Retrievals for GPM (IMERG), v05B. NASA's precipitation processing center. ftp://arthurhou.pps.eosdis.nasa.gov/gpmdata/, 2014. (Accessed June 2019).

[47] F. Chen, H. Huang, Comparisons of Gauge, TMPA and IMERG products for monsoon and tropical cyclone precipitation in southern China, Pure Appl. Geophys. 176 (2019) 1767–1784, https://doi.org/10.1007/s00024-018-2038-z.

[48] C. Huang, J. Hu, S. Chen, A. Zhang, Z. Liang, X. Tong, L. Xiao, C. Min, Z. Zhang, How well can IMERG products capture typhoon extreme precipitation events over southern China? Rem. Sens. 11 (2019) 70, https://doi.org/10.3390/rs11010070.

[49] E. Omranian, H.O. Sharif, A.A. Tavakoly, How well can global precipitation measurement (GPM) capture hurricanes? Case study: hurricane Harvey, Rem. Sens. 10 (7) (2018) 1150, https://doi.org/10.3390/rs10071150.

[50] G. Balsamo, A. Beljaars, K. Scipal, P. Viterbo, B. van den Hurk, M. Hirschi, A. K. Betts, A revised hydrology for the ECMWF model: verification from field site to terrestrial water storage and impact in the integrated forecast system, J. Hydrometeorol. 10 (3) (2009) 623–643. https://doi:10.1175/2008jhm1068.1.

[51] J.M. Van Der Knijff, J. Younis, A.P.J. De Roo, LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation, Int. J. Geogr. Inf. Sci. 24 (2) (2010) 189–212, https://doi.org/10.1080/13658810802549154.

[52] S. Harrigan, E. Zsoter, L. Alfieri, C. Prudhomme, P. Salamon, F. Wetterhall, C. Barnard, H. Cloke, F. Pappenberger, GloFAS-ERA5 v2.1 operational global river discharge reanalysis 1979-present, Earth Syst. Sci. Data (2020), https://doi.org/10.5194/essd-2019-232.

[53] ECMWF, GloFAS v2.0 description. https://confluence.ecmwf.int/display/COPSRV/GloFAS+v2.0. (Accessed 6 February 2020).

[54] L. Alfieri, P. Burek, E. Dutra, B. Krzeminski, D. Muraro, J. Thielen, F. Pappenberger, GloFAS – global ensemble streamflow forecasting and flood early warning, Hydrol. Earth Syst. Sci. 17 (3) (2013) 1161–1175. https://doi:10.5194/hess-17-1161-2013.

[55] F.A. Hirpa, P. Salamon, H.E. Beck, V. Lorini, L. Alfieri, E. Zsoter, S.J. Dadson, Calibration of the global flood awareness system (GloFAS) using daily streamflow data, J. Hydrol. 566 (2018) 595–606, https://doi.org/10.1016/j.jhydrol.2018.09.052.

[56] Facebook connectivity lab and center for International Earth Science Information Network (CIESIN). http://ciesin.columbia.edu/data/hrsl/#acknowledgements. (Accessed 14 March 2019).

[57] C.C. Sampson, A.M. Smith, P.D. Bates, J.C. Neal, L. Alfieri, J.E. Freer, A high-resolution global flood hazard model, Water Resour. Res. 51 (9) (2015) 7358–7381.

[58] J. Neal, G. Schumann, P. Bates, A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas, Water Resour. Res. 48 (11) (2012).

[59] A. Smith, C. Sampson, P. Bates, Regional flood frequency analysis at the global scale, Water Resour. Res. 51 (1) (2015) 539–553.

[60] S. Leyk, A.E. Gaughan, S.B. Adamo, A. de Sherbinin, D. Balk, S. Freire, A. Rose, F. R. Stevens, B. Blankespoor, C. Frye, J. Comenetz, A. Sorichetta, K. MacManus, L. Pistolesi, M. Levy, A.J. Tatem, M. Pesaresi, The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use, Earth Syst. Sci. Data 11 (3) (2019) 1385–1409.

[61] A. Smith, P.D. Bates, O. Wing, C. Sampson, N. Quinn, J. Neal, New estimates of flood exposure in developing countries using high-resolution population data, Nat. Commun. 10 (1) (2019) 1814.

[62] UN OCHA, Mozambique: cyclone Idai & floods, Situation Report 2: https://reliefweb.int/sites/reliefweb.int/files/resources/MOZAMBIQUE%20CYCLONE%20IDAI%20%26%20FLOODS%20SITUATION%20REPORT%20NO.2%2003%20April%202019%20FINAL.pdf, 2019. (Accessed 14 February 2020).

[63] M. Yamaguchi, J. Ishida, H. Sato, M. Nakagawa, WGNE Intercomparison of tropical cyclone forecasts by operational NWP models, Bull. Am. Met. Soc. (November 2017) 2337–2349, https://doi.org/10.1175/BAMS-D-16-0133.1, 2017.

[64] K.I. Hodges, R. Emerton, The prediction of Northern Hemisphere tropical cyclone extended life cycles by the ECMWF ensemble and deterministic prediction systems. Part I: tropical cyclone stage, Mon. Weather Rev. 143 (2015) 5091–5114, https://doi.org/10.1175/MWR-D-13-00385.1.

[65] UN OCHA, UN General Assembly Resolution 46/182, 1991. https://www.unocha.org/sites/unocha/files/dms/Documents/120402_OOM-46182_eng.pdf Last. (Accessed 13 July 2019).

[66] L. Zhou, X. Wu, Z. Xu, H. Fujita, Emergency decision making for natural disasters: an overview, Int. J. Disaster Risk Reduct. 27 (2018) 567–576, https://doi.org/10.1016/j.ijdrr.2017.09.037.

[67] S. Rayner, D. Lach, H. Ingram, Weather forecasts are for wimps: why water managers do not use climate forecasts, Climatic Change 69 (2005) 197–227.

[68] A. Lopez, S. Haines, Exploring the usability of probabilistic weather forecasts for water resources decision-making in the United Kingdom, Weather Clim. Soc. 9 (2017) 701–715.

[69] S. Shackley, B. Wynne, Representing uncertainty in global climate change science and policy: boundary-ordering devices and authority, Sci. Technol. Hum. Val. 21 (3) (1996), https://doi.org/10.1177/016224399602100302.

[70] S.L. Star, J.R. Griesemer, Insititutional ecology, 'translations' and boundary objects: amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39, Soc. Stud. Sci. 19 (3) (1989), https://doi.org/10.1177/030631289019003001.

[71] FCFA and WISER, A Manual for Co-production in African Weather and Climate Services, 2019. https://futureclimateafrica.org/coproduction-manual/Last. (Accessed 13 July 2019).

[72] Arnal, L., Anspoks, L., Manson, S., Neumann, J., Norton, T., Stephens, E., Wolfenden, L. and Cloke, H. L.: "Are we talking just a bit of water out of bank? Or is it Armageddon?" Front line perspectives on transitioning to probabilistic fluvial flood forecasts in England, Geoscience Communication, in press, https://doi.org/10.5194/gc-2019-18, 2020.

[73] D.A. Lavers, S. Harrigan, E. Andersson, D.S. Richardson, C. Prudhomme, F. Pappenberger, A vision for improving global flood forecasting, Environ. Res. Lett. 14 (2019) 121002, https://doi.org/10.1088/1748-9326/ab52b2.

[74] Beven, K., Asadullah, A., Bates, P., Blyth, E., Chappell, N., Child, S., Cloke, H., Dadson, S., Everard, N., Fowler, H. J., Freer, J., Hannah, D. M., Heppell, K., Holden, J., Lamb, R., Lewis, H., Morgan, G., Parry, L. and Wagener, T.: Developing observational methods to drive future hydrological science: can we make a start as a community?.

[75] R.A. Davidson, L.K. Nozick, T. Wachtendorf, B. Blanton, B. Colle, R.I. Kolar, S. DeYoung, et al., An Integrated Scenario Ensemble-Based Framework for Hurricane Evacuation Modeling: Part 1—Decision Support System, Risk Analysis, 2018, https://doi.org/10.1111/risa.12990.

[76] B. Blanton, K. Dresback, B. Colle, R. Kolar, H. Vergara, Y. Hong, N. Leonardo, R. Davidson, L. Nozick, T. Wachtendorf, An Integrated Scenario Ensemble Based Framework for Hurricane Evacuation Modeling: Part 2—Hazard Modeling, Risk Analysis, 2018, https://doi.org/10.1111/risa.13004.

[77] K. Yang, R.A. Davidson, H. Vergara, R.L. Kolar, K.M. Dresback, B.A. Colle, B. Blanton, T. Wachtendorf, J. Trivedi, L.K. Nozick, Incorporating inland flooding into hurricane evacuation decision support modeling, Nat. Hazards (2019), https://doi.org/10.1007/s11069-019-03573-9.

[78] D.L.A. Flack, C.K. Skinner, L. Hawkness-Smith, G. O'Donnell, R.J. Thompson, J. A. Waller, A.S. Chen, J. Moloney, C. Largeron, X. Xia, S. Blenkinsop, A. J. Champion, M.T. Perks, N. Quinn, L.J. Speight, Recommendations for improving integration in national end-to-end flood forecasting systems: an overview of the FFIR (Flooding from Intense Rainfall) Programme, Water 11 (4) (2019) 725, https://doi.org/10.3390/w11040725.

[79] UN, Sustainable Development Goal 17, 2019. https://sustainabledevelopment.un.org/sdg17. (Accessed 10 February 2020).

A9: Submitted Article: Recommendations to improve the interpretation of global flood forecasts to support international humanitarian operations for tropical cyclones

This appendix contains the submitted version of the following paper (summarised in Section 7.5.3):

Speight, L., Stephens, E.M., Hawker, L., Baugh, C.A., Neal, J., Cloke, H.L., Grey, S., Titley, H.A., Marsden, K., Simner, A., Ficchì, A., Prudhomme, C., Archer, J., Bazo, J., Dambo, J., Dolan, S., Hunn, A.L., Moschini, F., Savage, J., Smith, J., Towner, J. and Wanzala, M. (2023) Recommendations to improve the interpretation of global flood forecasts to support international humanitarian operations for tropical cyclones. Journal of Flood Risk Management, (Submitted).

Recommendations to improve the interpretation of global flood forecasts to support international humanitarian operations for tropical cyclones

Linda Speight[1,2], Elizabeth Stephens[1,3,4], Laurence Hawker[5], Calum Baugh[6], Jeffrey Neal[5,15], Hannah Cloke[1,3,7], Stephen Grey[8], Helen Titley[1,10], Katherine Marsden[9,10], Tim Sumner[9], Andrea Ficchi[1,16], Christel Prudhomme[6, 11, 12], Leanne Archer[5], Juan Bazo[4], Jânio Dambo[13], Siobhan Dolan[1], Anna Lena Huhn[14], Francesca Moschini[6], James Savage[15], Andy Smith[15], Jamie Towner[1], Maureen Wanzala[1]

1. Department of Geography and Environmental Science, University of Reading, Reading, RG6 6AB, UK
2. School of Geography and the Environment, South Parks Road, Oxford, OX1 3QY, University of Oxford, UK
3. Department of Meteorology, University of Reading, Reading RG6 6UR, UK
4. Red Cross Red Crescent Climate Centre, the Hague, the Netherlands
5. School of Geographical Sciences, University of Bristol, University Road, Clifton, Bristol, BS8 1SS, UK
6. European Centre for Medium-Range Weather Forecasts (ECMWF), Shinfield Park, Reading, RG2 9AX, UK
7. Department of Earth Sciences, Uppsala University, SE-751 05 Uppsala, Sweden
8. HR Wallingford Ltd, Howbery Park, Wallingford, Oxfordshire, OX10 8BA, UK
9. Foreign, Commonwealth & Development Office, King Charles Street, London SW1A 2AH, UK
10. Met Office, Fitzroy Road, Exeter, EX1 4PB, UK
11. UK Centre for Ecology and Hydrology, Wallingford, OX10 8BB, UK
12 Geography Department, Loughborough University, Loughborough, LE11 3TT, UK
13. Mozambique Red Cross (CVM), 284, Avenue Agostinho Neto, Maputo, Mozambique
14. German Red Cross, Carstennstrasse 58, Berlin, Germany
15. Fathom, Square Works, 17-18 Berkeley Square, Clifton, Bristol, SB8 1HB, UK
16. Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milano, Italy

Abstract

During major events decision makers are faced with multiple sources of information. International humanitarian organisations increasingly turn to forecast teams to support the coordination of efforts to respond to disasters caused by hazards such as tropical cyclones and large-scale fluvial floods. Such disasters often occur where there is limited local capacity or information available to support decision making and so global forecasting capacity is utilised to provide impact-based flood forecast bulletins.

A multidisciplinary team of meteorologists, hydrologists, coastal, inundation, and exposure modellers joined together to provide forecast bulletins and expertise for such events through the UK Foreign and Commonwealth Development Office (FCDO). In this paper we capture the successes and challenges of using global models to generate information to support humanitarian operations for two cyclones: Hurricane Iota in Central America (November 2020) and Cyclone Eloise in Mozambique (January 2021). From this experience we draw recommendations to improve global forecasting systems for forecast producers, forecast translators, humanitarian users, and the wider international community of researchers and practitioners involved in disaster prediction and response. We consider the requirement for additional data and expertise to support the interpretation of global models and we discuss how global forecasting can help build local forecasting capability.

Key words (up to 8): Tropical Cyclones, Iota, Eloise, Global flood models, anticipatory action, impact-based forecasts

1. Introduction

An appeal for international humanitarian assistance is made by the government of the disaster-affected country when the magnitude and / or duration of an emergency is beyond their response capacity (UN General Assembly, 46th Session, December 1991). In these situations, international humanitarian organisations require impact-based forecasts and assessments to support coordination, logistics and decision-making (Nauman, 2021).

Tropical cyclones (also known regionally as hurricanes or typhoons) are associated with strong winds, heavy rainfall and in some cases storm surges. There are also often impacts from fluvial flooding which can extend far from the landfall location and last long after the initial landfall (Titley *et al.*, 2021). When responding to such disasters humanitarian organisations may have access to a variety of products and services such as local maps, data feeds and translators. The availability and quality of this information varies by location. Global-scale flood forecast models, though often at a course resolution, can provide a consistent and timely source of information when forecasts from national mandated authorities are not easily accessible, where transboundary impacts are possible or where longer lead times or probabilistic forecasts are required (Emerton *et al.*, 2016; Hirpa *et al.*, 2018a; Lavers *et al.*, 2019; Ward *et al.*, 2015). This approach was first piloted to produce impact-based flood forecast bulletins to support humanitarian operations for the devastating floods in Mozambique caused by Cyclones Idai and Kenneth in 2019 (Emerton *et al.*, 2020). Through co-ordination and interaction between forecast producers, translators, intermediaries and users (see Table 1) the approach has been improved and systematised.

This paper aims to share our experiences of using global flood models to support humanitarian decision making in Central America and Mozambique through a Flood Early Warning Pilot funded by the UK Foreign and Commonwealth Development Office (FCDO). Specific examples from case studies of Hurricane Iota (November 2020) and Cyclone Eloise (January 2021) are drawn on to identify recommendations to provide more robust and actionable impact-based flood forecasts from global models. The users in these examples were largely humanitarian organisations however the discussion and recommendations presented in this paper are also valid for other local, national and international emergency management organisations who use global models to inform their decision making. It should be noted that although we used the terminology “Flood Early Warning Pilot”, users held the responsibility for communicating any resulting flood warnings and taking action.

1.1. Humanitarian Anticipatory Action

The impact-based flood forecast bulletins have been developed alongside a growing use of forecasts by humanitarians to support anticipatory action. Acting in advance of disasters is a cost-effective way to reduce the impacts of major weather and climate driven events by increasing the resilience of local communities and targeting humanitarian response to the areas most in need (Coughlan de Perez et al., 2015; Gros et al., 2019; Gros et al., 2020). The success of this approach is contingent upon the increasing sophistication of impact-based flood forecasts and warnings (Merz *et al.*, 2020), driven by technological advances in data availability (Nauman et al., 2021), continued development of hydro-meteorological models (Emerton *et al*. 2016; Wu *et al.*, 2020) and a growing commitment to collaborative and multi-disciplinary ways of working (Golding, 2022).

Warning systems are only effective if appropriate action is taken on the basis of the warning. In the humanitarian anticipatory action framework, Early Action Protocols (EAPs) are formal plans

produced by Red Cross National Societies (other humanitarian organisations have also developed similar procedures) which outline the early actions to be taken, and the organisations responsible for them, when a specific hazard is forecasted to impact communities. When the forecast reaches a pre-defined trigger, funding for these forecast based actions is released from the International Federation of Red Cross and Red Crescent Societies (IFRC) Disaster Response Emergency Fund (DREF) (IFRC, 2022).

The provision of impact-based flood forecast bulletins supports humanitarian anticipatory action by providing early notification of the potential activation of EAPs, real-time information to support decision making before and during major events, and helps builds familiarisation with forecasts. Evaluating the utility of global forecasts in real-time for specific events (as opposed to skill evaluation using reforecast datasets) provides an alternative way of highlighting where further research or investment is needed to improve their usability.

2. Approach to producing and evaluating the impact-based flood forecast bulletins

2.1. Organisational roles

There are multiple roles involved in the provision of these impact-based flood forecast bulletins (hereafter referred to as flood bulletins). These are detailed in Table 1 using the definitions developed in the Weather and Climate Information Services for Africa (WISER) co-production manual (Carter, *et al.*, 2019), but with the important role of forecast interpretation explicitly acknowledged with the addition of a forth category for 'forecast translator'.

Table 1 Organisation roles in the provision of flood bulletins during the Flood Early Warning Pilot

Organisational role	Definition	Organisations involved
Forecast producers	"Those who produce weather and climate data and information" and who are responsible for long term improvements to the forecast through scientific and technological development	Joint Research Centre (JRC), European Centre for Medium Range Forecasting (ECMWF), HR Wallingford, FATHOM, national hydro-met agencies, Joint Typhoon Warning Centre (JTWC), Regional Specialised Meteorological Centres (RSMC)
Forecast translators	"Those who have the responsibility for converting this data into a form that is appropriate for the user" by translating the complex forecasts into key decision relevant messages and informative graphics	University of Reading (UoR), University of Bristol (UoB), HR Wallingford, ECMWF, FATHOM
Intermediaries	"Those who support engagement between producers and users… Intermediaries have content knowledge and play the role of a knowledge broker, or connector, in co-production"	UK Foreign, Commonwealth & Development Office (FCDO), Red Cross Red Crescent Climate Centre, Start Network

Organisational role	Definition	Organisations involved
Users	"Those who will take action based on the weather and climate information."	FCDO, United nations Office for Coordination of Humanitarian Affairs (UN OCHA), National Red Cross / Red Crescent Societies, World Food Programme (WFP), Start Network Members

Note: Definitions in quotation marks are as described by Cater (2019)

2.2. Operating procedure

The trigger to produce the bulletins is a collaborative decision based on a preliminary assessment of likely impact as well as the likely information needs of the humanitarian community, and the feasibility of using the available forecasting tools given the hydrometeorological situation for the event (Figure 1). Following the trigger, a pre-agreed operating procedure is followed which addresses the flow of data and information, timing of calls and production of the bulletin (Figure 1). Examples of the bulletin content and presentation, and a discussion of its evolution, are provided in Emerton *et al.* (2020) and Budimir *et al.* (2022). After the event a post event wrap-up provides a process for documenting learning and deciding on any ways forward. As the organisations involved in producing the bulletins included research organisations who do not have a 24/7 operational remit (see Emerton *et al.*, 2020), bulletins during this pilot project are not produced on weekends or public holidays.

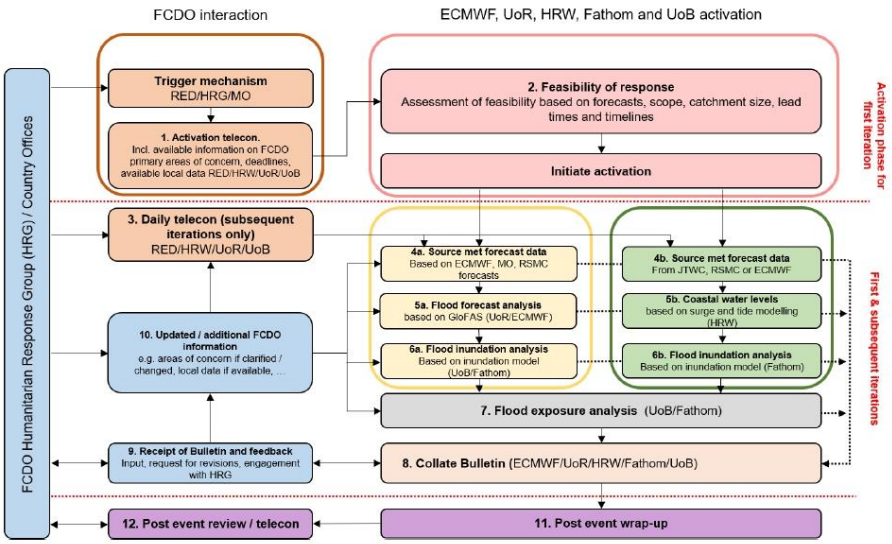


Figure 1 Flood Early Warning Pilot operating procedure

Definitions: Foreign, Commonwealth & Development Office (FCDO) Humanitarian Response Group (HRG), FCDO Research and Evidence Directorate (RED), Met Office (MO), University of Reading (UoR), University of Bristol (UoB), HR Wallingford

(HRW), European Centre for Medium Range Forecasts (ECMWF), Joint Typhoon Warning Centre (JTWC), Regional Specialised Meteorological Centres (RSMC)

Colours: Orange – trigger process, pink – feasibility assessment, yellow – fluvial analysis, green – coastal analysis, grey – exposure analysis, cream – reporting, feedback and incorporation of additional information, blue – FCDO input to bulletin focus and content, purple - post event evaluation

2.3. Forecast methods

Impact-based forecasts should provide details of the timing, location and severity of potential impacts to inform appropriate action (Harrowsmith, *et al.*, 2020; Robbins, *et al.*, 2022; WMO, 2015; WMO, 2021).

The impact-based forecasts bring together global scale methods for flood forecasting, hazard mapping and exposure analysis. Meteorological forecasts of precipitation, wind speed and tropical cyclone track locations are sourced from the Regional Specialised Meteorological Centres (RSMC), the Severe Weather Impact Assessment report produced by the UK Met Office, and ECMWF data.

The fluvial flood forecast uses the Global Flood Awareness System (GloFAS, www.globalfloods.eu), an early warning component of the European Commission Copernicus Emergency Management Service (emergency.copernicus.eu). Fluvial flooding is estimated by combining GloFAS forecast probabilities of exceeding the flood alert thresholds, with pre-computed 3 arc second (~90m) flood hazard maps from Fathom (Sampson et al., 2015). These flood inundation maps are produced by a hydrodynamic model based on the numerical scheme of LISFLOOD-FP (Neal *et al.*, 2012) and a global flood inundation model framework (Sampson *et al.*, 2015). Population exposure is estimated by intersecting flood inundation extents with population counts constrained to buildings from WorldPop (Lloyd et al., 2015; Bondarenko et al., 2020a; Bondarenko et al., 2020b).

Cyclone track forecasts from the appropriate operational Tropical Cyclone (TC) forecasting centre are interpreted to create time-varying wind and pressure fields which are applied to a regional or global TELEMAC-2D model (Hervouet, 2007, www.opentelemac.org) to predict storm surge. To estimate storm-surge flooding, water levels along the affected coast are extracted from TELEMAC2D model and then applied to the boundary of the LISFLOOD-FP inundation model run in near real time.

The methodology is described in detail by Emerton *et al.* (2020) and summarised in Appendix 1.

2.4. Post event review

Post event review is an integral component of the Flood Early Warning pilot (Figure 1). There are two stages to the process, an immediate post event wrap-up to collate initial thoughts and feedback, followed by the production of a post event review report and review meeting. Input is sought from all organisations identified in Table 1.

The post event review combines an assessment of the forecast evolution alongside available observations and impact data, and feedback on the use and usefulness of the bulletins (Table 2). The aim is to document the strengths and limitations of each activation and identify recommendations for both operational and scientific improvements. The approach taken for the post event reviews has been largely qualitative. This reflects both the limited observations available (especially immediately) to support the review, and, most importantly, that while post event reviews can be a useful learning process, analysis of any one individual event provides limited information about the skill of a model in forecasting other events (either in the same region or elsewhere). For these reasons a more quantitative approach to the post event reviews is not considered appropriate. The content of this paper is based on the post event reviews completed following Cyclones Iota and Eloise.

Table 2 Post event review topics and example questions

Areas covered	Example questions
Approach	• Were there any difficulties encountered with the modelling approach? • Were there any difficulties encountered with the bulletin production?
Dissemination	• Did the bulletins reach the right people? • Was any additional information shared that could be incorporated into the methodology?
Forecasts	• How far in advance was the event forecasted? • Could the trigger process have been initiated earlier? • Were the areas most impacted correctly highlighted? • Was the event more or less extreme than anticipated in the bulletin?
Further activities	• Is this work leading to any subsequent activities? (e.g. has data been found that will be incorporated elsewhere, what collaborations have developed?).

3. Summary of case study events

3.1. Tropical Cyclone Iota

The 2020 Atlantic Hurricane season (June-November) was the most active Atlantic hurricane season on record with 30 named storms and 7 major hurricanes (NOAA, 2021). The last of these, Hurricane Iota, affected Nicaragua, Honduras, Guatemala, and parts of other bordering countries. The category five hurricane made landfall on 17 November 2020 03:40 UTC (16 November, 10:40 EST local time), south of Puerto Cabezas in northeast Nicaragua. River levels in the region were already high and the humanitarian community was already in response mode following Eta, a category 4 hurricane which had made landfall just 24km north of Iota's landfall location two weeks earlier. The reported maximum event rainfall from Iota was 510mm in Guatamala – including 94mm falling in a 6hr period on 18th November (Stewart, 2021) and the resulting storm surge was reported to be up to 8m around the village of Haulover, Nicaragua (Stewart 2021). Figure 2 shows the maximum flood return period exceeded in the GloFAS-ERA5 reanalysis.

The humanitarian response to Eta and Iota was combined into a single operation and it is difficult to differentiate the compound impacts of the two events separately (UN OCHA, 2020a). In total 5.2 million people were affected across nine Latin America and Caribbean countries (OCHA, 2020a). Widespread mud and rockslides and flooding of up to 2m was reported across central America causing loss of life and damage to property, roads and infrastructure (IFRC, 2020a; OCHA, 2020a; OCHA 2020b). A storm surge caused coastal flooding to air and sea ports across the region hampering response efforts (Stewart, 2021; Shultz *et al.*, 2021). The immediate response from the Red Cross included activating emergency response centres, supporting evacuation and rescue missions, distributing humanitarian aid included food packages and hygiene kits, and damage assessment (IFRC 2020a). Longer term humanitarian response is ongoing and expected to last for many years to come (OCHA., 2020b).

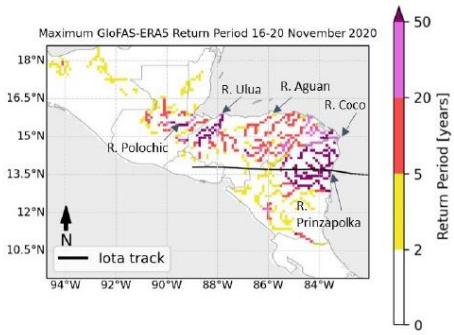


Figure 2 Iota Maximum GloFAS-ERA5 Return Period 16th - 20th November 2020

The purple colours indicate the expected locations of the most severe inland flooding. Labels identify major river of interest.

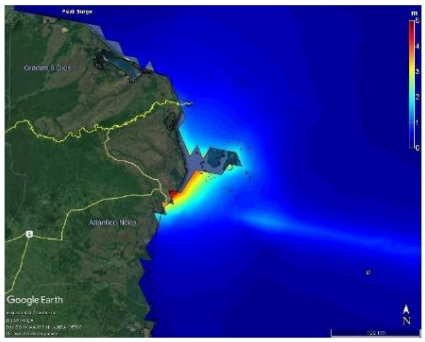


Figure 3 Peak surge from Hurricane Iota predicted by the TELEMAC-2D global model

3.2. Iota forecast bulletins

The evolution of the cyclone forecast for Iota and resulting rainfall is shown in Figure 5. A consistent signal for severe flooding in GloFAS (coinciding with the increase in confidence in the TC position and rainfall accumulations) was identified on 12th November, five days before landfall. At this point an initial discussion started around the possible need to provide briefings (Table 3). This extended lead time allowed forecast translators time to collate useful data sets and other resources before the official activation. The first flood bulletin was produced the day before landfall (16 November) highlighting the exposure in Honduras, Guatemala and Nicaragua (Figure 4, Table 3).

The surge forecast was based on the NHC forecast released at 0900 16/11/2020 UTC and predicted a maximum surge of up to 6m north of Puerto Cabezas (Figure 3). The coastal model bathymetry was based on General Bathymetric Chart of the Oceans data (GEBCO 2020) which is often poor in coastal areas away from main navigation routes where there is unlikely to have been a dedicated survey. For the stretch of coast on which Hurricane Iota made landfall near Puerto Cabezas the combination of coarse regular grid (node spacing of approximately 16 km) and likely sparse source bathymetry data produced a poor representation of the coastline (Figure 3). The small islands northeast of Puerto Cabezas are represented in the TELEMAC-2D model as connected to the mainland forming a peninsular. The coarse regular mesh creates triangular elements along the coast which lead to localised enhancement of surge at the apex of the triangles. In addition, the global model does not predict tide well, although in this case, the local tide has a small range and is relatively unimportant.

The hurricane made landfall further south than forecast. The predicted peak surge of 6m was less than the reports of 8m at Haulover and Wawa Bar (Stewart, 2021). It is not known how the surge quoted in Stewart (2021) was estimated or whether it is relative to the astronomical tide or a fixed vertical datum. Both Haulover and Wawa Bar are located near river mouths and it is possible that the surge combined with river flooding to produce a very high compound flood level. The global coastal global model was useful in identifying the overall area of coastline at risk and the approximate time of landfall but only a qualitative indication of the magnitude of the surge.

The bulletins were successfully used by humanitarians to inform actions on the ground prior to the floods occurring:

> *"The bulletin was very helpful in communicating to IFRC colleagues and National Societies who were deployed on the ground conducting humanitarian operations, the potential impact that Hurricane Iota could bring to communities that had already been struck by Eta. With this information we were able to give early warning so that Start Network volunteers and staff as well as IFRC could be prepared in areas such as San Pedro de Sula in Honduras where Eta had already caused extensive damage and losses and where forecasts were once again predicting severe impacts".*

Juan Bazo, Red Cross Red Crescent Climate Centre.

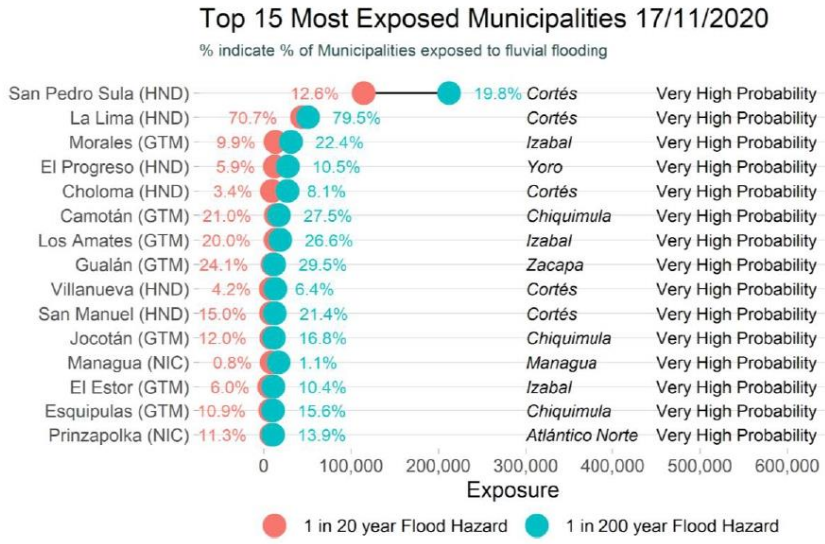


Figure 4 Forecast exposure of population by municipality from Iota

Estimated exposure per municipality for a less extreme 20-year return period (red dot) to the more extreme 200-year return period (blue dot). The percentages indicate the proportion of the municipality population that could be exposed. The probabilities refer to the probability of exceeding the GloFAS severe alert threshold (20-year return period with the following probability thresholds: Very-High (Expected) > 70%; High (Probable) 50%-70%; Moderate (Possible) 25%-50%; Low (Unlikely) < 25%

Table 3 Timeline of Tropical Cyclone Iota evolution and bulletin activity 12th – 19th November 2020

Date	Hydro-meteorological situation	Bulletin Activity
12th November	• Broad low-pressure system develops over Caribbean Sea • Signal for severe flooding first seen in GloFAS	Internal discussion on potential need for bulletins started
13th / 14th November	• Tropical storm develops • Hurricane forecast advisory issued by the National Hurricane Centre • River catchments already very wet following Eta	Continued monitoring of forecasts
15th November	• Hurricane Iota passes over islands of Providencia and Santa Catalina • Notable increase in GloFAS forecast probability of severe flooding and reduced spatial variability in cyclone track forecast (Figure 5)	
16th November	• End of period of rapid intensification • Iota upgraded to category 5 hurricane	First flood bulletin issued. Key messages: • severe flooding likely across northern Honduras and eastern Guatemala (Figure 4) • coastal surge expected around landfall • flash floods and landslides in areas of heavy rainfall
17th November	• Iota makes landfall then weakens as it passed over mountainous terrain in Northern Nicaragua • Storm surge of up to 8m in parts of Nicaragua • Heavy rainfall, landslides and flooding reported across central America	Flood bulletin issued with no major change from previous bulletin
18th November	• On-going inland flooding	Flood bulletin issued. Key messages: • widespread flooding similar to Eta occurring across the region • levels currently above government thresholds and continuing to rise in Guatemala and Nicaragua
19th November	• On-going inland flooding	Last flood bulletin. Key messages: • Further rainfall expected in Guatemala • Flooding ongoing

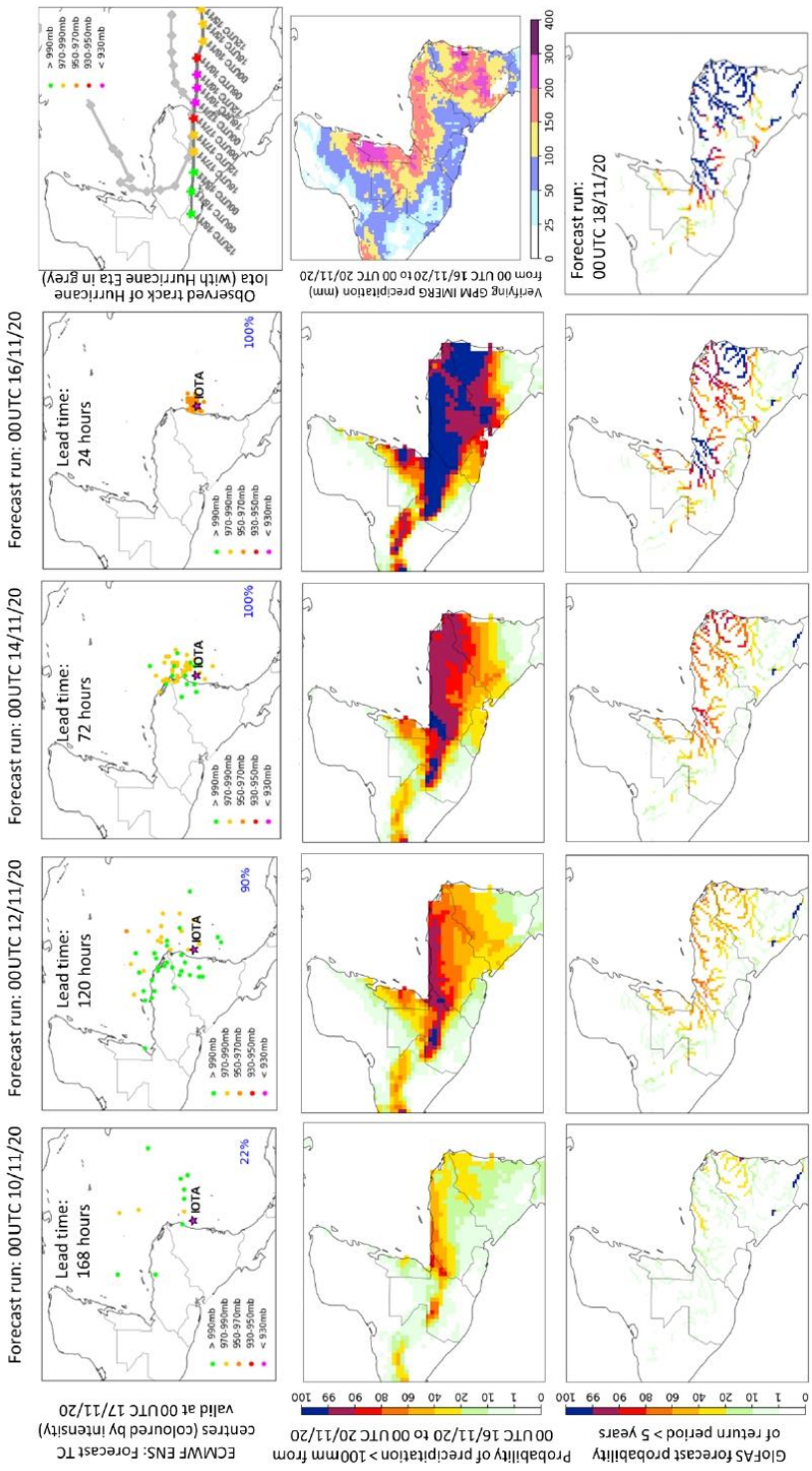


Figure 5 Iota cyclone, rainfall and flood forecast evolution – In the top row the star shows the observed centre position of TC Iota at 00UTC on 17th January (just before landfall), the coloured dots show the forecast positions valid at this time from each EC ENS member. The percentages indicate the percentage of members that had a cyclone centre forecast within the frame. The far-right image shows the observed cyclone tracks of Iota and Eta. The second row shows the precipitation probability forecasts from EC ENS for 4-day event precipitation totals > 100mm, with the verifying GPM data for the same period in the far-right column. The bottom row shows the forecast GloFAS return period of exceeding the 5-year flood threshold, with an additional forecast from a post-landfall GloFAS run on the right.

3.3. Key reflections from the Iota post event review

3.3.1. Balancing skill at longer lead time with operational decision making and response capacity

Figure 5 shows that the risk of heavy precipitation along the north coast of Honduras was predicted four to six days before landfall, indicating that if flood bulletins had been produced at longer lead times they would have identified a similar severity of flood impacts for this area. For areas close to the landfall location (the east coast of Nicaragua), the rainfall forecasts were more sensitive to correctly forecasting the landfall location, and so the lead times at which forecasts signalled the heavy rain was lower (one to three days). For the Río Grande de Matagalpa in Central Nicaragua, the GloFAS signal only stabilised once the rain has fallen. There was a notable increase in the probability of exceeding the 20-year return period in the correct locations from the 15 November run onwards. From this point the main rivers expected to experience severe flooding (>20-year return period) remained consistent in GloFAS over time indicating good forecast confidence (Figure 2 and Figure 5).

One of the advantages of global probabilistic models is providing information at longer lead times compared to local forecast models (Lavers et al., 2019). However there is always an inherent trade-off for humanitarians between the benefit of acting early and the potential of acting in vain or in the wrong location (Bischiniotis et al., 2019). Post event discussion with users indicated that despite the potential to identify areas of high risk four to six days before landfall, it was unlikely producing earlier bulletins would have changed the timing of actions taken by humanitarians. For Eloise an interim step of providing short written forecast briefing paragraphs to FCDO in the lead up to the event was added to the operating procedures, enabling humanitarian to benefit from early indication of potential impacts whilst balancing the resource commitment of producing a full bulletin against the potential actions.

3.3.2. Providing information at decision relevant scales in mountainous catchments

Multiple landslides, surface water flooding and flooding on smaller mountainous rivers occurred during Iota (FloodList, 2020; Amenesty International, 2020). This was of concern to humanitarian responders but these processes are not well represented in GloFAS which is only recommended for use in catchments greater than 1000km^2 (Hirpa *et al.* 2018b). By using knowledge of the large-scale situation from GloFAS and local gauged data (where available) some information could be provided on flood exposure for these regions by interpolating between known points. Information was also provided on flooding on access roads further down the river that could isolate mountainous communities (Figure 6). The focus on access roads was a pragmatic solution as many of the reported impacts of Eta and Iota involved road damage and closure which affects the ability to access and support affected communities (IFRC, 2021c).

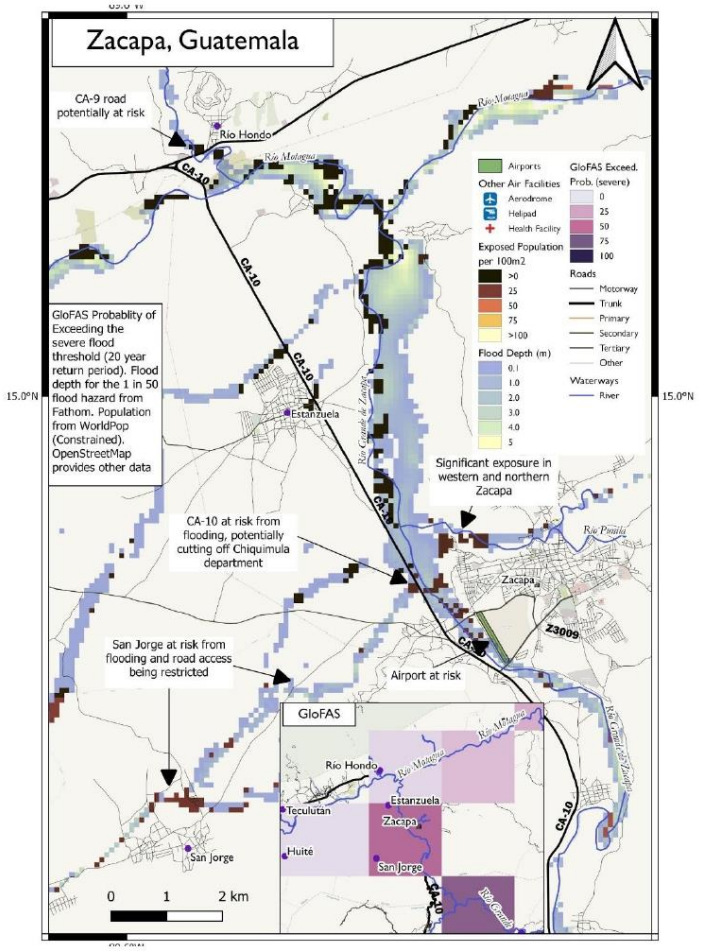


Figure 6 Example flood hazard map for Zacapa, Guatemala on the confluence of the Grande de Zacapa and Motagua rivers from 18 November illustrating identification of flooding to access roads

The map shows the 1 in 50 year flood hazard depth (blue to yellow) and the probability of exceeding the severe flood level (purples), which is high in this region (~50%). Exposed population is shown (black to yellow), as well as roads and airports.

3.3.3.Use of real time observations to support interpretation of forecasts in Central America

The operational skill of GloFAS in Central America was largely untested before Iota. The interpretation of GloFAS forecasts was therefore based on model producers' knowledge of model performance in other catchments with similar hydro-meteorological drivers. An immediate "sense check" of the forecasts was conducted using publicly available rain gauge and hydrometric data from national agencies which was used to compare the observed flood situation with previous forecasts. Inundation and exposure estimates were compared to satellite imagery and available data (news reports, photos and videos) of impacts. Access to this information enabled forecasted producer to include an assessment of forecast confidence in the bulletins. The flood levels forecasted by GloFAS (event return periods) and flood exposures closely matched available observations in large river

basins such as Ulua and Motagua. In addition, the available satellite data showed good agreement with the modelled flood inundation. One of the key risk areas highlighted in the bulletin was the River Chamelecón where floods of up to 2m were reported which totally destroyed many properties along the river (IFRC 2021c).

3.4. Tropical Cyclone Eloise

Tropical Cyclone Eloise made landfall in central Mozambique, near Beira, on 23rd January 2021 around 2am local time (UTC 00:00). The Sofala Province (of which Beira is the provincial capital) was still recovering from tropical cyclone Idai which had made landfall in the same region in 2019. In addition, many people were already displaced across the region following the impact of Tropical Storm Chalane three weeks earlier on 30th December 2020. Due to this, there was already an active humanitarian presence in the area. The World Meteorological Organisations (WMO) Regional Specialised Meteorological Centre La Réunion and the Mozambique National Institute of Meteorology (INAM) issued warnings of high winds, heavy rainfall and coastal flooding before Eloise made landfall (WMO, 2021). Windspeeds of 140km/h with gusts up to 160km/h were reported by INAM along with 250mm of rainfall in 24 hours in Beira which resulted in flooding in the Buzi catchment and surrounding areas. Eloise continued to track southwest across Mozambique and into South Africa and Botswana over the following days (Reliefweb, 2021).

Figure 7 shows the maximum flood return period exceeded in the GloFAS-ERA5 reanalysis for Eloise and Figure 8 shows the forecast peak surge.

Eloise did not meet the trigger specified in the Mozambique Early Action Protocol on Cyclones (IFRC, 2019) at the pre-defined decision-time three days before landfall, however based on the information from national authorities and international forecast models (including the forecast bulletins reported here), the Mozambique Red Cross (CVM) decided to deploy a readiness team to Beira and to the bordering Province of Ihambane (IFRC, 2021a). Flooding affected large parts of the city of Beira and surrounding agricultural regions (Frey, 2021). Although widespread, it was not as severe as the flooding from Cyclone Idai in 2019 (Figure 10).

Following the flooding in Sofala province, there was a second phase of impacts in Southern Mozambique centred on the Gaza province. The heavy rainfall following Cyclone Eloise, combined with existing high water levels and full upstream dams meant that on 1st February the observed water level on the Limpopo reached the trigger of the Limpopo Flood Early Action Protocol (IFRC, 2020b), the flood peak was expected to reach the district of Chokwé on February 4th (IFRC, 2021b). The Mozambique Red Cross (CVW) responded by distributing early warning messages, assisting in the evacuation of vulnerable communities, reinforcing housing structures and distributing jerry cans and WASH and COVID-19 hygiene kits.

OCHA reports there were 12 deaths from Eloise (one in Madagascar and 11 in Mozambique) with more than 467,000 people affected (OCHA, 2021) and 56,000 houses severely damaged or destroyed (IOM, 2021) across Madagascar, Mozambique, South Africa and Zimbabwe. Mozambique's central provinces, which were still recovering from the devastation wrought by Cyclone Idai in 2019, were hardest-hit (UN OCHA, 2021).

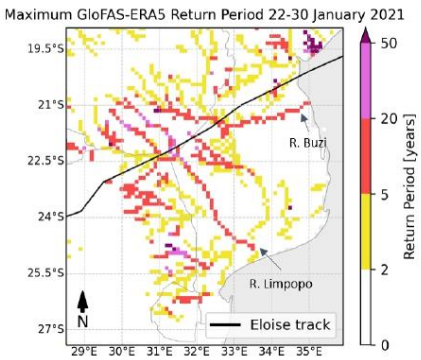


Figure 7 Maximum GloFAS-ERA5 Return Period 22nd January - 30th January 2021

The purple colours indicate the expected locations of the most severe inland flooding. Labels identify major river of interest.

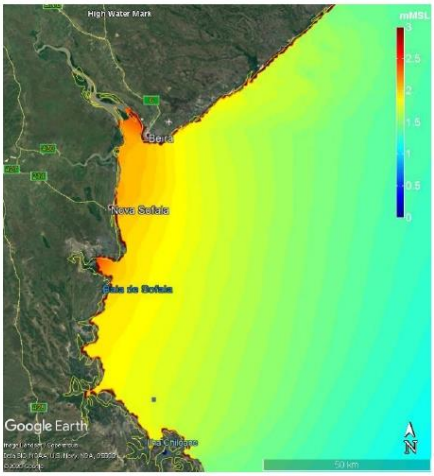


Figure 8 The maximum water level during Cyclone Eloise predicted by the Mozambique Channel TELEMAC-2D model

3.5. Eloise forecast bulletins

GloFAS identified the potential of impactful flooding as early as 14th January (nine days before landfall, Figure 9) and internal discussions with FCDO started at this point. The first full bulletin was produced on 22nd January, the day before landfall (Table 4).

For the forecast of cyclone surge, a regional TELEMAC-2D model was available covering the full east coast of Mozambique and the Mozambique Channel. The cyclone surge forecast was up to 2m in the vicinity of Beira but coincided with neap tide so that the forecast peak water level remained less than the highest astronomical tide (Figure 8). Therefore, widespread coastal flooding was not expected, and, from anecdotal reports, this appears to have been the outcome. No quantitative reports of peak sea levels were available to validate the forecast. In contrast to Central America, the TELEMAC-2D model resolution was 1 km along the Mozambique coast, made use of good quality digitised nautical chart data and was validated in its representation of the tide. This allows greater confidence in the surge forecast for Eloise compared to Iota.

Humanitarian partners on the ground, such as the Red Cross Movement, received the bulletins and were able to utilise them for the purpose of early, informed decision-making alongside other international and national forecasting products. Of particular benefit was the identification of two distinct phases of impacts and the flood extent maps;

> *"The bulletins put the FbF* [Forecast based Financing] *mechanism in Mozambique on alert, signalling early on with significant lead time that precipitation related to TC Eloise would lead to flooding in the Limpopo river basin, long before national hydrological forecasts reflected this. The information allowed the Mozambique Red Cross and Governmental partners to focus multi-hazard anticipatory actions two-fold: On Cyclone impact close to Beira, as well as associated flooding in southern Mozambique. The bulletin's flood extent maps, indicating critical infrastructure forecasted to be impacted by the flooding, were used*

in overlay with Mozambique Red Cross's composite vulnerability index to identify the districts for anticipatory action interventions."

Anna Lena Huhn, German Red Cross

Table 4 Timeline of Tropical Cyclone Eloise evolution and bulletin activity 12th January – 1st February 2021

Date	Hydro-meteorological situation	Bulletin Activity
12th January	• Low-pressure weather system forms in south-west India	Internal discussion on potential need for bulletins started
19th January	• Eloise intensifies into severe storm • Makes landfall in Madagascar • Glofas identifies potential for high-impact flooding (Figure 9)	First briefing paragraph issued
21st January		Briefing paragraph issued
22nd January	• Eloise intensifies into a Tropical Cyclone • River levels already high due to wet start to rainy season • Mozambique Red Cross (CWM) send preparedness teams to Beira and Inhambane, and start to warn and support communities along the River Limpopo	First flood bulletin issued. Key messages: • flooding not expected to be as severe as Idai • Two stages of impacts expected, 23-26th in Sofala province, 25th onwards from the River Limpopo with the highest exposure in Gaza province (Figure 11)
23rd January	• Tropical Cyclone Eloise makes landfall near Beira • 24 hours rainfall 250mm • Wind speeds of 140km/h (gusts 160km/h)	Briefing paragraph issued
24th January	• Tropical Storm Eloise reaches South Africa	Briefing paragraph issued. key message: • GloFAS forecasting a decrease in severity of flooding for the River Limpopo. For example the forecast percentage of population exposed in Chokwe district fell from a range of 14.3-24.1% (Figure 11) to 7.9-11.3%
25th January	• Tropical Storm Eloise reaches Botswana where it weakens	Flood bulletin. Key message: • Flooding on the Limpopo not expected to be as severe as 2013 floods
27th January		Last flood bulletin issues. Key massages: • River levels in Sofala province starting to fall. • Flooding on Limpopo expected in coming days • Upstream dams at capacity
1st February	• Limpopo Early Action protocol trigged from local model	

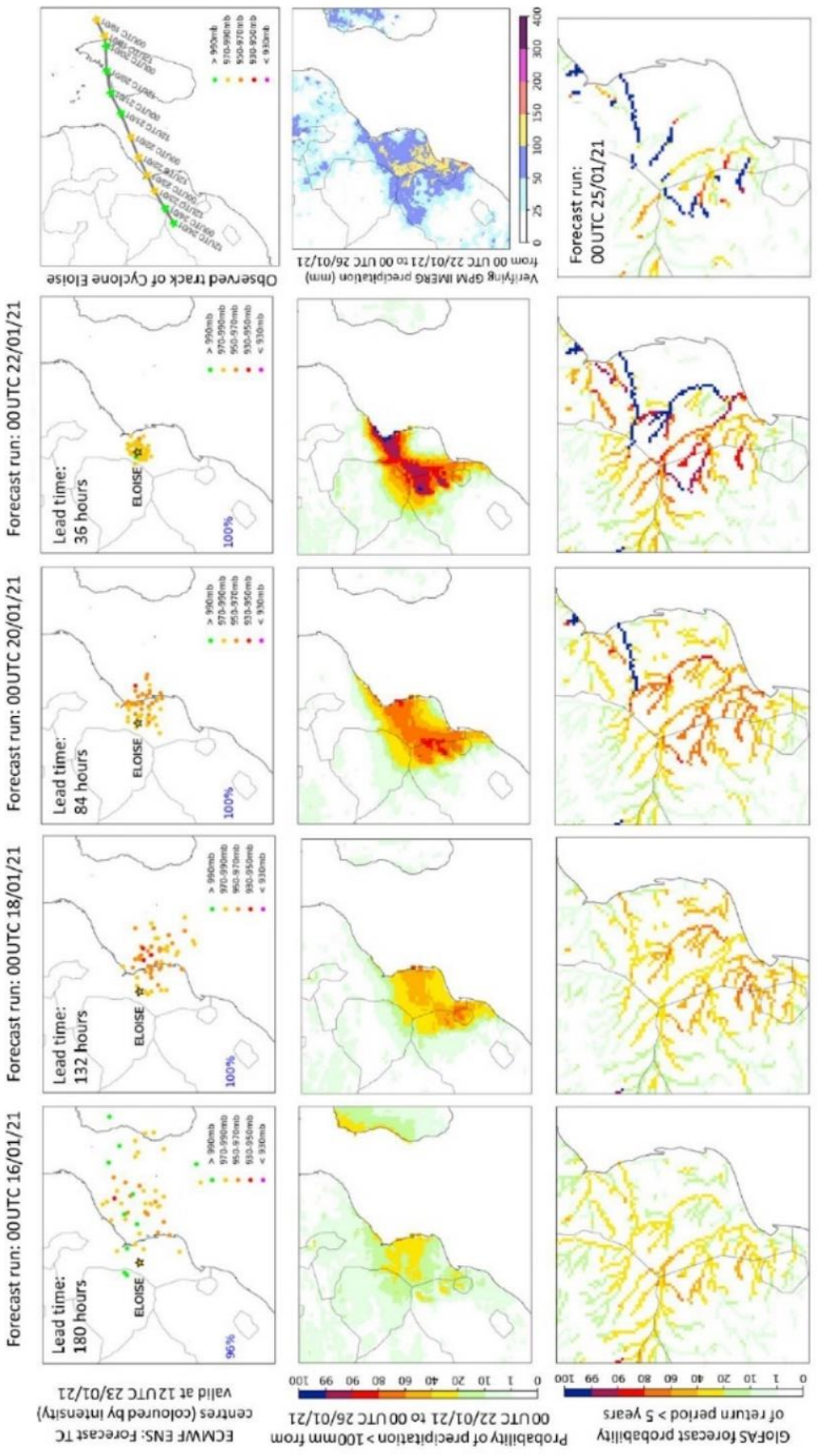
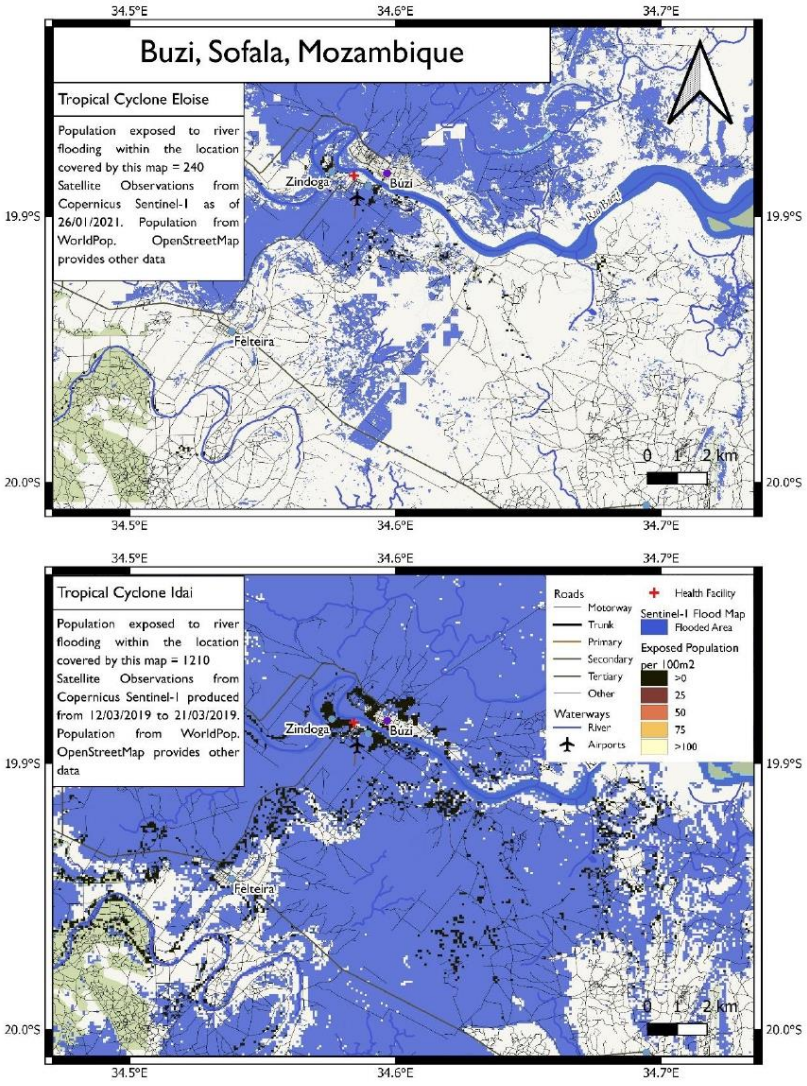


Figure 9 Eloise cyclone, rainfall and flood forecast evolution - In the top row the star shows the observed centre position of TC Eloise at 12UTC on 23rd January (12 hours after landfall), the coloured dots show the forecast positions valid at this time from each EC ENS member. The percentages indicate the percentage of members that had a cyclone centre forecast within the frame. The far-right image shows the observed cyclone track. The second row shows the precipitation probability forecasts from EC ENS for 4-day event precipitation totals > 100mm, with the verifying GPM data for the same period in the far-right column. The bottom row shows the forecast GloFAS return period of exceeding the 5-year flood threshold, with an additional forecast from a post-landfall GloFAS run on the right.

1



2

3 *Figure 10 Flood inundation maps from Sentinel-1 for cyclones Eloise (top) and Idai (bottom) for the area in and around Buzi.*
4 *Estimates of exposed population from the Fathom global flood model and WorldPop are overlaid on the Sentinel-1 imagery*

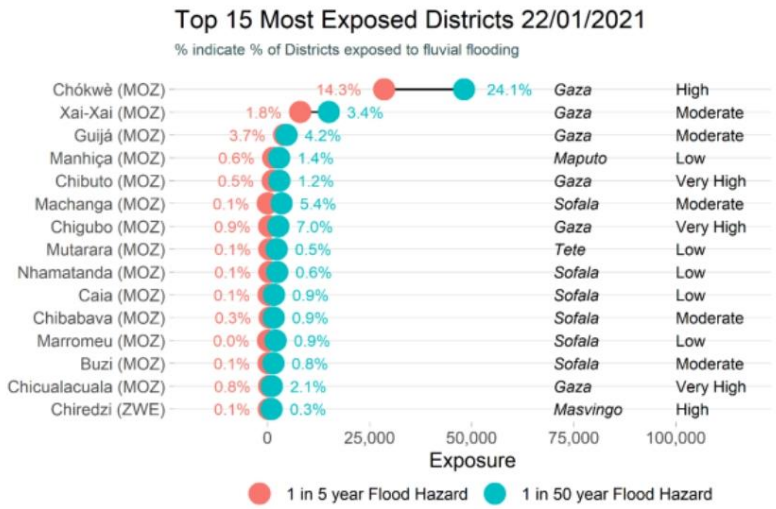


Figure 11 Forecast exposure of population by district from Eloise

Estimated exposure per municipality for a less extreme 5-year return period (red dot) to the more extreme 50-year return period (blue dot). The percentages indicate the proportion of the district population that could be exposed. The probabilities refer to the probability of exceeding the GloFAS severe alert threshold (20 -year return period) and should be interpreted as: Very-High (Expected) > 80%; High (Probable) 50%-80%; Moderate (Possible) 25%-50%; Low (Unlikely) < 25%

3.6. Key reflections from the Eloise post event review

3.6.1.Building networks benefits local and global partners

Following the experience of producing flood bulletins for previous cyclones Idai and Kenneth in Mozambique (see Emerton *et al.*, 2020), the bulletins benefitted from the existing relationships with national hydrometeorological organisations and humanitarian teams on the ground;

> *"We were in direct contact with the scientific team behind the bulletins and were able to be informed by them, as well as feedback information on national-level hydrological forecasts we received from our focal point at the national hydrological services DNGRH from our end. It was good to see how the bulletins were evolving and started to reflect additional information coming from the team in the field."*
>
> Jânio Dambo, Mozambique Red Cross

Eloise was the fourth major cyclone to make landfall in Mozambique in less than three years. It was evident that humanitarians had an existing awareness of the type of issues they would face and the important questions to ask the forecast translators. Demonstrating a growth in capacity since 2019 when a post event review of Idai reported that warnings were not taken sufficiently serious as individuals and organisations failed to appreciate the magnitude of the upcoming event (Norton, 2020).

3.6.2.Flexibility helps to respond to user requests for contextual information

A key question from humanitarians during Eloise was how bad would the flooding be compared to Idai. The GloFAS forecasts (Figure 9) and exposure analysis (Figure 11) showed that Eloise was a much smaller event than Idai and this was communicated in the first and subsequent flood bulletins.

The International disaster charter was activated to acquire satellite imagery for the Beira region by FCDO. Once these images were available this distinction could be clearly communicated by overlaying the satellite observations of inundated area with population and Open Street map data for both events (Figure 10). Feedback from FCDO indicated this comparison was very useful as it made it clear that the response to Eloise was something that could be managed within the existing routine humanitarian response framework.

Meeting the immediate request for event comparison required flexibility from the bulletin producers, additional data processing and the availability of relevant datasets. Scientifically the comparison between Idai and Eloise was complicated, feedback from FCDO suggests this complexity was not fully appreciated by bulletin users. Eloise was a much smaller event than Idai, therefore there was more uncertainty in the inundation and exposure analysis as small differences in the topography or location of buildings and population would lead to large difference in the forecast exposure (Smith et al., 2019 and Hawker et al., 2020). Furthermore, a substantial area of flooding related to seasonally flooded wetlands that are not considered by the global flood model, requiring auxiliary remote sensing data to characterise the floodplain pre-event.

3.6.3.The value of combining global and local forecasts for the River Limpopo

GloFAS forecast the timing and magnitude of the flood peaks on the Buzi river well however in the Limpopo catchment the forecast was more challenging. Firstly, the lead time of flooding on the Limpopo was much longer than on the Buzi. Recent research has shown that there is a lead time effect on flood thresholds in GloFAS (Zsoter *et al.*, 2020) meaning events with a longer lead time can initially appear less or more extreme in some locations, and the bulletin producers manually incorporated this emerging research into their assessments. Secondly a shift in the forecast on 24th January saw the severity of expected flooding on the Limpopo reduce. Thirdly, and most significantly, GloFAS only has a rudimentary representation of dam management (ECMWF, 2022). The river Limpopo has multiple large dams upstream in South Africa and Zimbabwe which, if not at capacity, act to slow the timing of the flood peak and reduce the magnitude (Alfieri *et al.*, 2013). Whilst the bulletin producers had access to some publicly available information on dam levels, there are national hydrological forecasts models available for the Limpopo which include real time observations of water level and therefore provide more reliable forecasts at short leadtimes. The bulletins clearly signposted readers to this stating “Please refer to DNGRH for the most accurate forecast information for flood magnitude and timing along the Limpopo River, as this incorporates observations of upstream river levels”. Although GloFAS gave an early indication of the potential need to trigger the Limpopo EAP, local models were more reliable for predicting the timing and magnitude of flood peaks.

4. Discussion and recommendations to improve the use of global flood models

The collaborative post event review process (Figure 1) highlighted areas where further research or investment is needed to make better use of global flood models to support humanitarian decision making. The recommendations in this section were included in a Learning and Recommendations report for FCDO in 2021 (Speight *et al.*, 2021) and seek to identify specific actions and research requirements related to the use of global models. Recommendations for improving other aspects of forecast bulletin design, production and dissemination are discussed in Emerton *et al.* (2020) and Budimir *et al.* (2022).

4.1. Work in partnership

Building effective flood forecasting systems requires working together, understanding user needs, being transparent about scientific limitations and developing strategies to effectively communicate risk (Harrowsmith *et al.*, 2020; Hirons *et al*., 2021; Robbins *et al.*, 2022; WMO, 2021).

To provide useful information, forecast translators need to understand how forecasts are used in practice. For example, what are the key challenges for decision makers (Section 3.6.2), what lead time is needed to take anticipatory action (Section 3.3.1), what spatial scale do humanitarians need forecast information at and where are the main areas of concern (Section 3.3.2)? As advocated by (Emerton *et al.*, 2020), maintaining some flexibility in being able to respond to specific questions is valuable, however it is more efficient to identify these questions in advance to enable collation of the required data sets and necessary processing time and skills (Section 3.6.2).

For many users it is difficult to articulate what forecast information is needed without understanding what can be provided given the logistical, scientific and technical constraints (Section 3.6.1). Similarly, for effective post event review, producers need to be clear about what type of feedback is needed and how it will be used. Methods of co-production (see Hirons *et al.*, 2021), such as joint workshops outside of busy operational periods, are key to developing pragmatic solutions that can support decision making within the scientific constraints of global models (WMO, 2021). For some regions, these topics are already being discussed through the joint development of Early Action protocols and globally partnerships are developing through platforms such as the Anticipation Hub (https://www.anticipation-hub.org/) and research programmes such as SHEAR (http://shear.org.uk/). This network building activity is critical to support the increased effective use of humanitarian forecast based decision making.

4.2. Provide accessible documentation to support interpretation of bulletins

During major events humanitarians are faced with multiple sources of information. Decisions were not made based on the flood bulletins in isolation. Training and supporting documentations around the strengths and limitations of the global models used and the supporting data sets is required (Alfieri *et al.*, 2018; Ward *et al.*, 2015). The strengths and limitations of any model or data set won't be the same for every event or for all regions of the world. Where local forecast models, which are calibrated based on observations, are available they will usually perform better than global models and this should be clearly signposted to users (Section 3.6.3). The daily discussions with FCDO were invaluable to communicate this local and event specific context.

There were multiple users of the flood bulletins with varying skills in forecast-based decision making. For both Eloise and Iota strategic decision makers used the summary information on the front page whilst the data specialists were interested in the technical figures and asked questions about model uncertainty and comparisons between different population data sets. Technical guidance documents were produced containing information about the methodology, model skill and underlying datasets to share with bulletin users interested in this additional level of detail. To support informed decision making and help increase capacity for interpreting global forecasts this information should be easily accessible alongside the flood bulletins.

4.3. Improve synergies between global models and local contexts

Information which could support the development and interpretation of impact-based forecasts is from disparate sources, often not easily accessible, and information holders are not necessarily aware of its value to others. Improved global data sharing is essential for the continued improvement of global flood models (Lavers *et al.*, 2019; 2020). During bulletin production access to available in-country data provided valuable information to inform interpretation of the global model

forecasts, for example data on vulnerable populations, local knowledge of catchment characteristics, the current operational status of dams and reservoirs and observed river levels before and during the event (Sections 3.3.3, 3.6.1 and 3.6.3). The availability of real time observations of river levels and reported impacts offered a pragmatic approach to incorporate an immediate sensibility check of the model forecasts and helped support post event review. The time forecast translators spent collating such information during events was considerable. In Mozambique the process benefitted from existing established links with the German Red Cross who were receiving information from DNGRH (Section 3.6.1). The dissemination of the flood bulletins also played a valuable role in communicating in country information that was not freely or readily available to all decision makers (Section 3.6.3).

For global forecasting groups building relationships with organisations who may hold this information will help increases access to data. All parties would benefit from building and maintaining a meta-database of links to in country data and integrating this into existing global platforms for data sharing.

4.4. Improve application of existing and emerging research and knowledge

A strength of the bulletins is their ability to flexibly incorporate emerging developments into global forecast systems, for example TC Eloise demonstrated the advantage of using fine resolution, validated regional surge models where available and highlighted the need to develop higher resolution global coastal models. Understanding of the lead time dependent thresholds of Zsoter *et al.* (2020) illustrated in (Section 3.6.3) improved the usability of GloFAs forecasts. Further work by Titley *et al*. (2023) analyses the link between cyclone track and flood forecast skill for Iota. Ongoing updates to the flood hazard model at Fathom and UoB (e.g. Bates *et al.*, 2021, Neal *et al.*, 2021; Zhao *et al.*, 2021) bring regular incremental improvements in accuracy, and joint work with ECMWF as part of this project enabled the inundation maps to be based on the full GloFAS ensemble rather than percentage chance of exceeding a threshold.

One of the challenges of using global flood models in this context is that the existing skills metrics do not necessarily reflect the criteria that are important to decision makers (see section 4.5). A real time expert assessment of confidence should therefore be included in forecast bulletins based on the predictability of the meteorological situation (Lavers *et al.*, 2021; Harrigan *et al.*, 2020; Titley et al, 2021) and understanding of how well the model represents hydrological processes (Alfieri *et al.*, 2013) for the region of interest, the consistency of the forecast over time (Section 3.3.1), and any observations on how the model forecasts compare to observed flows or real time impacts (Section 3.3.3).

Whilst the bulletins have attempted to communicate this assessment, the process was ad hoc and much of the expertise remains tacit. The depth of understanding of GloFAS contained within the team, alongside active involvement in ongoing joint research on the operational use of GloFAS (e.g. Budimir *et al.*, 2022; Coughlan de Perez *et al.*, 2016; Emerton *et al.*, 2020; Ficchi *et al.*, 2021; Hirpa *et al.* 2018a; Hossain *et al.*, 2022; Robbins *et al.*, 2022) and global flood modelling (e.g. Emerton *et al.*, 2016; Harrigan *et al.*, 2020a; Harrigan *et al.*, 2020b; Lavers *et al.*, 2020; Towner *et al.*, 2019; Zsoter *et al.*, 2020) enabled valuable expert interpretation. This contrasts to other areas of hydro-meteorology where models are more developed and there are limits to how much additional value forecast translators can add to model output using their own expertise compared to the inbuilt skill of the forecast models (Pagano *et al.*, 2016). However, the small group of experts that currently process this knowledge limits the number of events that forecast bulletins can be produced for and potentially acts as a gatekeeping process for the wider effective use of global models by decision

makers. Therefore, research that has been shown to have operational impact should be communicated widely and prioritised in model updates.

Funding for future model developments is contingent on evidence of impact. The increasing use of global models to support decision making provides an opportunity to identify this. By combining available data (section 4.3) with focused user feedback (section 4.1) and emerging best practice on methodologies for post event reviews in data sparse contexts (HiWeather, 2022; Magnusson, 2019; Venkateswaran *et al.*, 2020), a valuable learning database is developing which will help build confidence and skills in the use of global forecasts for decision making.

4.5. Develop action-based forecast skill assessment to support improved decision making

The applied nature of the bulletin development has offered a unique opportunity to identify directions for future research informed by operational needs. The most significant is the need to develop an operationally informed forecast skill assessments in regions with limited data.

The decision to trust available global forecasts would be easier if users could easily compare the output and skill of different approaches (Alfieri *et al.*, 2018; Towner *et al.*, 2019, and Section 4.2). A major limitation of global models such as GloFAS is that although they have been calibrated for hundreds of river catchments covering a range of global climates (Hirpa *et al.*, 2018b), they have not been calibrated for all countries where a humanitarian response might be activated as observation data are not easily available (Lavers *et al.*, 2019). Where data are available, attempts have been made to assess global skill (Harrigan, *et al.*, 2020a; Harrigan *et al.*, 2020b) and skill metrics are published on the GloFAS website (ECMWF, 2021). However, these global skill assessments often focus on the skill of the forecast model over the full range of flows from droughts to floods and can be of limited value to assess the skill during large flood events. Global flood models are increasingly being used by humanitarians to support the development of Early Action Protocols (Coughlan de Perez *et al.*, 2016; Coughlan de Perez *et al.*, 2015; Nauman, *et al.*, 2021). In these situations, a decision-based skill assessment for the country or location of interest is undertaken (Coughlan de Perez *et al.*, 2016; Ficchi *et al.*, 2021; Hossain *et al.*, 2022; Lala *et al.*, 2021; Lopez *et al.*, 2020).

Impact based forecasts require linking hydrological and coastal models with flood inundation models to simulate hazard. As such flood forecast inundation maps embody multiple sources of uncertainty, one specific example is the downscaling of GloFAS to the 90m resolution used in the inundation mapping to provide information at a decision relevant scale (see Appendix). The performance of inundation models is often very site specific and building up a representative sample of events to make general statements on model performance is a long and ongoing process (see Trigg *et al.*, 2016; Bernhofen *et al.*, 2018 and Hawker *et al.*, 2020).

The continued development of pragmatic solutions to assess the suitability of global models for countries of interest in the absence of observation data will require collaboration (see section 4.1). Any such assessment should consider forecast skill from a user perspective (Parker, 2020), covering aspects of consistency, quality and value (Murphy, 1993) and considering how the derived metric could be used to improve decision making during events (Lopez *et al.*, 2020; Werner *et al.*, 2016). Such an assessment is further complicated by the need to assess the skill of both the hydro-meteorological components and the inundation and exposure assessments. This is especially challenging as there is no systematic approach to impact reporting on a global scale (Robbins & Titley, 2018). Recent innovations (e.g. Mitheu *et al.* 2022) are starting to use historical re-forecasts and observed flood reports to evaluate skill for past events in ungauged basins for both the flood and inundation model components for specific locations. These approaches should be modified for use in global models.

5. Conclusions

In the absence of global hydrological observations, it continues to be difficult to evidence the skill of global flood models. The case studies presented here for Tropical Cyclones Iota and Eloise, as well as the previous experience of producing flood bulletins from Idai and Kenneth (Emerton *et al.*, 2020), show that despite their well-documented limitations, global flood models are a valuable resource for humanitarians. The access to computationally-expensive probabilistic information and increased lead time with respect to local models is appreciated by decision makers and FCDO reported that humanitarian colleagues are appreciative and excited by the expertise and insight this pilot project has enabled during TC flooding events.

The effective use of global models currently requires understanding of their strengths and weakness in conjunction with access to local data to provide information at a scale that can support local humanitarian action. Long term funding is required to support the development of close research and operational collaborations to improve the linkages between expert knowledge and forecast interpretation. The increasing use of global forecasting systems highlights an urgent need for innovative development of user relevant metrics to assess the skill and value of global models, particularly in ungauged catchments.

Multiple global models are available (Emerton *et al.*, 2016), they are a valuable resource and should be used more effectively. With continued communication and co-ordination between people, organisations, models and data, global forecasting systems can improve humanitarian decision making. And in turn, the increased use of such systems will inform and motivate further development of forecasting capabilities at local and global scales. To achieve this, reliable funding to support the long-term development of the skillsets, partnerships and tools advocated in this paper is required.

Acknowledgements

The Flood Early Warning Pilot was funded by the UK Foreign, Commonwealth & Development Office (FCDO) contracted through the EACDS Lot B service 'Strengthening resilience and response to crises'.

In addition the authors acknowledge support from the UK's Natural Environment Research Council (NERC) and FCDO (DFID) under two projects within the Science for Humanitarian Emergencies and Resilience (SHEAR) research programme; Hannah Cloke, Andrea Ficchi, Linda Speight and Liz Stephens, through FATHUM (Forecasts for AnTicipatory HUManitarian action, grant number NE/P000525/1 and Jeffrey Neal and Laurence Hawker through HYFLOOD (Next generation flood hazard mapping for the African Continent at hyper-resolution, grant number NE/S006079/1).

Data Availability

Real-time GloFAS forecast products are freely available at www.globalfloods.eu, and GloFAS data can be obtained through the dedicated GloFAS data service (https://confluence.ecmwf.int/display/COPSRV/GloFAS+services). ECMWF ENS forecast data are available through the TIGGE archive (https://www.ecmwf.int/en/research/projects/tigge). The Global TELEMAC-2D model is open source (http://www.opentelemac.org/). The IMERG satellite rainfall data can be downloaded from NASA (https://gpm.nasa.gov/data/imerg) and the observed

tropical cyclone data from IBTrACS (https://www.ncdc.noaa.gov/ibtracs/). The LISFLOOD-FP code is freely available from http://www.bristol.ac.uk/geography/research/hydrology/models/lisflood/, and the global flood inundation maps used in this study are available upon request from Dr Jeffrey Neal or Dr Laurence Hawker. WorldPop data is available from https://www.worldpop.org/.

Appendix 1: Forecast methods

A1.1Coastal forecasts of storm surge

Since the original bulletin production for Cyclones Idai and Kenneth in 2019 (Emerton *et al.*, 2020), an approach for storm surge modelling was added to the procedures in response for the subsequent need for coastal flooding information for Cyclone Amphan which affected India and Bangladesh in 2020.

The surge forecast uses the open-source hydrodynamic model, TELEMAC-2D (Hervouet 2007, www.opentelemac.org). The model is forced with wind and pressure fields generated from cyclone track forecasts produced by relevant meteorological agencies such as the Joint Typhoon Warning Centre (JTWC) and Regional Specialized Meteorological Centers: the National Hurricane Center in Miami, Météo France in La Réunion and the Indian Meteorological Department in New Delhi. For regional models, tidal water levels and currents are applied at the boundaries of the model. Internal forcing from the gravitational effects of the moon and sun is also included.

A1.2Fluvial forecasts

The fluvial flood forecast uses the Global Flood Awareness System (GloFAS[1], www.globalfloods.eu), an early warning component of the European Commission Copernicus Emergency Management Service (emergency.copernicus.eu). Using a 51-member forecast ensemble from the ECMWF Integrated Forecast System (IFS), the GloFAS web interface produces a probability of exceeding three different flood severity thresholds corresponding to the 2-year, 5-year and 20-year return period (referred to as medium, high and severe respectively) at every grid point along the global river network out to 30 days ahead (Alfieri et al., 2013). GloFAS is designed to simulate large scale hydrological systems, predictions for smaller catchments (generally consider to be less than 1000km^2) should be evaluated with caution due to limited calibration (Hirpa *et al*., 2018b). Skill is also known to be lower in areas where large amounts of water can be stored naturally within the river system (for example in flat waterlogged areas, rivers with inner deltas and braided channels), in arid regions (Alfieri et al., 2013) and where water management features such as dams are utilised (Zajac et al., 2017).

A1.3Inundation modelling

A global flood inundation model framework (Sampson *et al.*, 2015) estimates riverine flooding at ~90m resolution for all basins with an upstream area >50km2 using a sub-grid hydrodynamic model within the LISFLOOD-FP code (Neal *et al.*, 2012). The inundation model has been updated from the original reference by using MERIT DEM (Yamazaki *et al.*, 2019) and MERIT HYDRO (Yamazaki *et al.*, 2019) and most recently FABDEM (Hawker et al., 2022). A regionalised flood frequency analysis (Smith *et al.*, 2015) provides model boundary conditions to produce flood hazard maps for different return periods.

For river flooding direct coupling of GloFAS with the flood inundation model is not possible given the large spatial area and short-time frames needed to produce results. As such, the inundation analysis selects a range of return period(s) inundation maps to align with the spread of estimates produced by the GloFAS ensemble, thus keeping the value of this probabilistic information. To estimate storm-

[1] The GloFAS version used during the case studies was v2.2. As of May 26th 2021, GloFAS has been upgraded to v3.1. For details see https://confluence.ecmwf.int/display/COPSRV/Latest+operational+release%3A+GloFAS+v3.1

surge flooding water levels along the affected coast are extracted from TELEMAC2D model and then applied to the boundary of the LISFLOOD-FP inundation model.

Dams, flood defences and wetland processes are not included in the inundation model as reliable data on these are difficult to obtain at large scale (Hawker *et al.*, 2020). The flood inundation model struggles to model frequent flood events (return periods < 20 years) as it is difficult for the model to capture anthropogenic change, and the model has a greater sensitivity to discharge, channel conveyance and topography errors for smaller flood events (Quinn *et al.*, 2019).

A1.4Exposure analysis

Leyk *et al.* (2019) describe the various available gridded population datasets available and their differences. The techniques to distribute population onto grids have improved markedly in recent years (Stevens *et al.*, 2015; Leyk *et al.*, 2019), especially with the addition of building footprints. For the bulletins, we used the WorldPop Constrained 2020 UN (WorldPop, Stevens *et al.*, 2015; Bondarenko *et al.*, 2020a; Bondarenko *et al.*, 2020b) dataset, chosen as the most up-to-date global dataset constraining population to building footprints, which results in a better estimate of exposure, especially in rural areas (Smith *et al.*, 2019). Despite the latest census data being used, the dataset is limited by the timeframe of the census data (e.g. Mozambique uses 2017 census data), and thus may not take into account the latest population movements, which can contain some of the most vulnerable people that have been displaced.

Population exposure due to fluvial flooding is estimated by combining GloFAS forecast probabilities of exceeding the flood alert thresholds, with flood inundation and population information, following the method developed for Cyclone Idai (Emerton *et al.* 2020). Population exposed due to coastal flooding is calculated by intersecting population data with a binary flood map from the coastal flood simulation.

Table 5 Data sets used in forecast models

Source	Used in	Description	Spatial resolution	Further information
GloFAS	Fluvial flooding forecast	Produced by ECMWF on behalf of the EU Copernicus service, hydro-meteorological forecasts force a global river routing model	0.1°x0.1° (~ 10km)	https://www.globalfloods.eu/
ECMWF IFS forecast model	Fluvial flooding forecast and meteorological summary	51 member numerical weather prediction ensemble used in GloFAS and meteorological summary	~18km	https://www.ecmwf.int/en/publications/ifs-documentation https://www.ecmwf.int/en/forecasts/documentation-and-support
GEBCO bathymetry	Coastal surge modelling	Ocean bathymetry used in surge model	15'×15' (~500m)	https://www.gebco.net/ GEBCO (2020)

Source	Used in	Description	Spatial resolution	Further information
Topography	Coastal and fluvial flood modelling	Global elevation data from MERIT DEM	90m	http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_DEM/ Yamazaki *et al.* 2017
JTWC cyclone track forecasts	Coastal surge modelling	Forecast cyclone tracks in Indian Ocean and western Pacific Ocean	-	https://www.metoc.navy.mil/jtwc/jtwc.html
NHC hurricane track forecasts	Coastal surge modelling	Forecast cyclone tracks in North Atlantic, Caribbean, eastern and central Pacific	-	https://www.nhc.noaa.gov/
MeteoFrance cyclone track forecasts	Coastal surge modelling	Forecast cyclone tracks in southeast Indian Ocean	-	http://www.meteofrance.re/cyclone/activite-cyclonique-en-cours
Fathom Global 2.0	Fluvial flooding forecast	Flood hazard layers used to predict inundation extent	~90m	https://www.fathom.global/fathom-global
WorldPop (Constrained 2020)	Flood exposure analysis	Population per 100m grid-cell, constrained by building footprint. Global coverage	3 arc second (~90m)	https://www.worldpop.org/geodata/listing?id=78
Humanitarian OpenStreetMap	Flood exposure analysis	Geographic data on infrastructure (e.g. roads) and key amenities (e.g. health facilities)	Vector data	https://www.hotosm.org/
Humanitarian Data Exchange	Flood exposure analysis	Geographic data on infrastructure (e.g. roads) and key amenities (e.g. health facilities)	variable	https://data.humdata.org/

References

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., & Pappenberger, F. (2013). GloFAS - global ensemble streamflow forecasting and flood early warning. *Hydrol. Earth Syst. Sci.*, 17(3), 1161-1175. https://doi.org/10.5194/hess-17-1161-2013

Alfieri, L., Cohen, S., Galantowicz, J., Schumann, G. J. P., Trigg, M. A., Zsoter, E., Prudhomme, C., Kruczkiewicz, A., Coughlan de Perez, E., Flamig, Z., Rudari, R., Wu, H., Adler, R. F., Brakenridge, R. G., Kettner, A., Weerts, A., Matgen, P., Islam, S. A. K. M., de Groeve, T., & Salamon, P. (2018). A global

network for operational flood risk reduction. *Environmental Science & Policy*, 84, 149-158. https://doi.org/https://doi.org/10.1016/j.envsci.2018.03.014

Amnesty International (2020). *When it rains it pours: the devastating impact of hurricanes Eta and Iota in Honduras.* Retrieved 16th September from https://www.amnesty.org/en/latest/news/2020/12/devastating-impact-hurricanes-eta-iota-honduras/

Bates, P. D., Quinn, N., Sampson, C., Smith, A., Wing, O., Sosa, J., et.al. (2021). Combined modeling of US fluvial, pluvial, and coastal flood hazard under current and future climates. *Water Resources Research*, *57*(2), e2020WR028673. https://doi.org/10.1029/2020WR028673

Bernhofen M V *et al* 2018 A first collective validation of global fluvial flood models for major floods in nigeria and mozambique *Environ. Res. Lett.* **13** 104007. https://doi.org/10.1088/1748-9326/aae014

Bischiniotis, K., van den Hurk, B., Coughlan de Perez, E., Veldkamp, T., Nobre, G. G., & Aerts, J. (2019). Assessing time, cost and quality trade-offs in forecast-based action for floods. *International Journal of Disaster Risk Reduction*, 40, 101252. https://doi.org/https://doi.org/10.1016/j.ijdrr.2019.101252

Bondarenko M., Kerr D., Sorichetta A., & Tatem, A.J. (2020a). Census/projection-disaggregated gridded population datasets, adjusted to match the corresponding UNPD 2020 estimates, for 183 countries in 2020 using Built-Settlement Growth Model (BSGM) outputs. WorldPop, University of Southampton, UK. doi:10.5258/SOTON/WP00685

Bondarenko, M., Kerr, D., Sorichetta, A., Tatem, A., & WorldPop, (2020b) Census/projection-disaggregated gridded population datasets for 51 countries across sub-Saharan Africa in 2020 using building footprints. University of Southampton doi:10.5258/SOTON/WP00682 [Dataset]

Budimir. M., Sneddon, A., Nelder, I., Brown. S., Donovan, A., & Speight, L. (2022). Development of forecast information for institutional decisionmakers: landslides in India and cyclones in Mozambique. *Geoscience Communication*, 5, 151-175. https://doi.org/10.5194/gc-5-151-2022

Carter, S., Steynor, A., Vincent, K., Visman, E., and Waagsaether, K. (2019). *Co-production of African weather and climate services*. https://futureclimateafrica.org/coproduction-manual

Coughlan de Perez, E., van den Hurk, B., van Aalst, M. K., Amuron, I., Bamanya, D., Hauser, T., Jongma, B., Lopez, A., Mason, S., Mendler de Suarez, J., Pappenberger, F., Rueth, A., Stephens, E., Suarez, P., Wagemaker, J., & Zsoter, E. (2016). Action-based flood forecasting for triggering humanitarian action. *Hydrol. Earth Syst. Sci.*, 20(9), 3549-3560. https://doi.org/10.5194/hess-20-3549-2016

Coughlan de Perez, E., van den Hurk, B., van Aalst, M. K., Jongman, B., Klose, T., & Suarez, P. (2015). Forecast-based financing: an approach for catalyzing humanitarian action based on extreme weather and climate forecasts. *Nat. Hazards Earth Syst. Sci.*, 15(4), 895-904. https://doi.org/10.5194/nhess-15-895-2015

ECMWF. (2021). *GloFAS evaluation*. Retrieved 17th September from https://confluence.ecmwf.int/display/COPSRV/GloFAS+evaluation

ECMWF (2022) Reservoirs, https://ec-jrc.github.io/lisflood-model/3_03_optLISFLOOD_reservoirs/. Model Documentation. Accessed 5/7/2022

Emerton, R., Cloke, H.L., Ficchi, A., Hawker, L., de Wit, S., Speight, L., Prudhomme, C., Rundell, P., West, R., Neal, J., Cuna, J., Harrigan, S., Titley, H., & Magnusson, L., Pappenberger, F., Klingaman, N., & Stephens, E. (2020). Emergency flood bulletins for Cyclones Idai and Kenneth: A critical evaluation of

the use of global flood forecasts for international humanitarian preparedness and response. *International Journal of Disaster Risk Reduction*, 50. https://doi.org/https://doi.org/10.1016/j.ijdrr.2020.101811

Emerton, R. E., Stephens, E. M., Pappenberger, F., Pagano, T. C., Weerts, A. H., Wood, A. W., Salamon, P., Brown, J. D., Hjerdt, N., Donnelly, C., Baugh, C. A., & Cloke, H. L. (2016). Continental and global scale flood forecasting systems. *WIREs Water*, 3(3), 391-418. https://doi.org/https://doi.org/10.1002/wat2.1137

Ficchì, A., Cloke, H., Speight, L., Mulangwa, D., Amuron, I., Ntale, E., & Stephens, L. (2021) Flood forecast skill for Early Action: Results and Learnings from the development of the Early-Action Protocol for Floods in Uganda, EGU General Assembly 2021, EGU21-16169, https://doi.org/10.5194/egusphere-egu21-16169, 2021.

FloodList. (2020). *Central America – Hurricane Iota causes deadly floods and landslides*. Retrieved 16th September from https://floodlist.com/america/central-america-hurricane-iota-floods-landslides-november-2020

Frey, A. (2021). *Mozambique: Cyclone Eloise hits Beira – photos*. Retrieved 16th September from https://clubofmozambique.com/news/mozambique-cyclone-eloise-hits-beira-photos-182771/

GEBCO Compilation Group (2020) GEBCO 2020 Grid (doi:10.5285/a29c5465-b138-234d-e053-6c86abc040b9)

Golding, B. (Ed) (2022). Towards the "perfect" weather warning: bridging disciplinary gaps through partnership and communication. Springer, pp270. https://doi.org/10.1007/978-3-030-98989-7

Gros, C., Bailey, M., Schwager, S., Hassan, A., Zingg, R., Uddin, M.M., Shahjahan, M., Islam, H., Lux, S., Jaime, C., Coughlan de Perez,. E. (2019). 'Household-level Effects of Providing Forecast-based Cash in Anticipation of Extreme Weather Events: Quasi-experimental Evidence from Humanitarian Interventions in the 2017 Floods in Bangladesh'. *International Journal of Disaster Risk Reduction*, Vol. 41, 101275. https://doi.org/10.1016/j.ijdrr.2019.101275

Gros, C., Easton-Calabria, E., Bailey, M., Dagys, K., de Perez, E.C., Sharavnyambuu, M. and Kruczkiewicz, A. (2022). The effectiveness of forecast-based humanitarian assistance in anticipation of extreme winters: a case study of vulnerable herders in Mongolia. *Disasters*, 46: 95-118. https://doi.org/10.1111/disa.12467

Harrigan, S., Cloke, H.L., Pappenberger, F. (2020). Innovating global hydrological prediction through an Earth system approach, *WMO Bulletin,* 69(1) https://public.wmo.int/en/resources/bulletin/innovating-global-hydrological-prediction-through-earth-system-approach

Harrigan, S., Zoster, E., Cloke, H., Salamon, P., & Prudhomme, C. (2020a). Daily ensemble river discharge reforecasts and real-time forecasts from the operational Global Flood Awareness System. *Hydrol. Earth Syst. Sci. Discuss.*, 2020, 1-22. https://doi.org/10.5194/hess-2020-532

Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., Barnard, C., Cloke, H., & Pappenberger, F. (2020b). GloFAS-ERA5 operational global river discharge reanalysis 1979–present. *Earth Syst. Sci. Data*, 12(3), 2043-2060. https://doi.org/10.5194/essd-12-2043-2020

Harrowsmith, M., Nielsen, M., Jaime, C., Coughlan de Perez, Erin., Uprety. M., Johnson. C., van den Homberg, M., Tijssen. A., Mulvihill Page, E., Lux. S. & Comment, T. (2020). *The future of forecasts:*

impact-based forecasting for early action. pp84. https://www.forecast-based-financing.org/wp-content/uploads/2020/09/Impact-based-forecasting-guide-2020.pdf

Hawker, L., Uhe, P., Paulo, L., Sosa, J., Savage, J., Sampson, C., Neal, J. (2022). A 30 m global map of elevation with forests and buildings removed. *Environmental Research Letters*, *17*(2). https://dx.doi.org/10.1088/1748-9326/ac4d4f.

Hawker, L., Neal, J., Tellman, B., Liang, J., Schumann, G., Doyle, C., ... & Tshimanga, R. (2020). Comparing earth observation and inundation models to map flood hazards. *Environmental Research Letters*, *15*(12), 124032. https://doi.org/10.1088/1748-9326/abc216

Hervouet, J-M. (2007) Hydrodynamics of free surface flows with the finite element method, John Wiley & Sons.

Hirons, L., Thompson, E., Dione, C., Indasi, V. S., Kilavi, M., Nkiaka, E., Talib, J., Visman, E., Adefisan, E. A., de Andrade, F., Ashong, J., Mwesigwa, J. B., Boult, V. L., Diédhiou, T., Konte, O., Gudoshava, M., Kiptum, C., Amoah, R. K., Lamptey, B., Lawal, K. A., Muita, R., Nzekwu, R., Nying'uro, P., Ochieng, W., Olaniyan, E., Opoku, N. K., Endris, H. S., Segele, Z., Igri, P. M., Mwangi, E., & Woolnough, S. (2021). Using co-production to improve the appropriate use of sub-seasonal forecasts in Africa. *Climate Services*, 23, 100246. https://doi.org/https://doi.org/10.1016/j.cliser.2021.100246

Hirpa, F. A., Pappenberger, F., Arnal, L., Baugh, C. A., Cloke, H. L., Dutra, E., Emerton, R. E., Revilla-Romero, B., Salamon, P., Smith, P. J., Stephens, E., Wetterhall, F., Zsoter, E., & Pozo, J. T.-d. (2018a). Global Flood Forecasting for Averting Disasters Worldwide. In Schumann, G.J-P., Bates, P.D., Apel, H., & Aronica, G.T. (Eds.) *Global Flood Hazard*. (pp. 205-228). American Geophysical Union https://doi.org/https://doi.org/10.1002/9781119217886.ch12

Hirpa, F. A., Salamon, P., Beck, H. E., Lorini, V., Alfieri, L., Zsoter, E., & Dadson, S. J. (2018b). *Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data*. Journal of Hydrology, 566, 595-606. https://doi.org/https://doi.org/10.1016/j.jhydrol.2018.09.052

HiWeather (2022) Warning value chain project, http://hiweather.net/Lists/130.html. Accessed 5/7/2022

Hossain, S., Cloke, H.L., Ficchì, A., Gupta, G., Speight, L., Hassan, A., & Stephens, E.M. (2022) A decision-led evaluation approach for flood forecasting system developments: An application to the Global Flood Awareness System in Bangladesh. Journal of Flood Risk Management. Submitted.

IFRC. (2019). *Mozambique/Africa: To coordinate early action for preparation and response to cyclones in Mozambique.* https://reliefweb.int/sites/reliefweb.int/files/resources/Mozambique%20Cyclone%20Early%20Action%20Protocol.pdf

IFRC (2020a) Revised emergency appeal: Central America – hurricane Eta and Iota, Red Cross & Red Crescent International, pp36 https://reliefweb.int/report/honduras/central-america-hurricane-eta-iota-revised-emergency-appeal-n-mdr43007-revision-no-1

IFRC. (2020b). *Mozambique: Floods - early action protocol summary.* https://reliefweb.int/sites/reliefweb.int/files/resources/EAP2020MZ02%20EAP%20Summary.pdf

IFRC. (2021a). *Mozambique Red Cross taking early action as Tropical Cyclone Eloise brings strong winds and rains.* https://www.forecast-based-financing.org/2021/01/22/mozambique-red-cross-taking-early-action-as-tropical-cyclone-eloise-brings-strong-winds-and-rains/#

IFRC. (2021b). *Anticipating the flood: taking early action at the Lower Limpopo in Mozambique.* https://www.forecast-based-financing.org/2021/02/04/anticipating-the-flood-taking-early-actions-at-the-lower-limpopo-in-mozambique/

IFRC (2021c). Operation Update N° 2 Honduras / Hurricanes Eta & Iota, International Federation of Red Cross and Red Crescent Societies, pp61 https://adore.ifrc.org/Download.aspx?FileId=380563

IFRC (2022). Practical information for National Societies on forecast-based financing and funding from the DREF. pp 14. Available from https://www.anticipation-hub.org/download/file-2778

Lala, J., Bazo, J., Anand, V., & Block, P. (2021) Optimizing forecast-based actions for extreme rainfall events, *Climate Risk Management*, 34. https://doi.org/10.1016/j.crm.2021.100374.

Lavers, D. A., Harrigan, S., Andersson, E., Richardson, D. S., Prudhomme, C., & Pappenberger, F. (2019). A vision for improving global flood forecasting. *Environmental Research Letters*, 14(12), 121002. https://doi.org/10.1088/1748-9326/ab52b2

Lavers, D.A., Ramos, M.-H., Magnusson, L., Pechlivanidis, I., Klein, B., Prudhomme, C., Arnal, L., Crochemore, L., Van Den Hurk, B., Weerts, A.H., Harrigan, S., Cloke, H.L., Richardson, D.S., & Pappenberger, F. (2020) A Vision for Hydrological Prediction. *Atmosphere*, 11, 237. https://doi.org/10.3390/atmos11030237

Lavers, D. A., Harrigan, S., & Prudhomme, C. (2021). Precipitation Biases in the ECMWF Integrated Forecasting System, *Journal of Hydrometeorology*, *22*(5), 1187-1198. https://doi.org/10.1175/JHM-D-20-0308.1

Leyk, S., Gaughan, A.E., Adamo, S.B., Sherbinin, A.D., Balk, D., Freire, S., Rose, A., Stevens, F.R., Blankespoor, B., Frye, C. & Comenetz, J. (2019). The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use. *Earth System Science Data*, 11(3), 1385-1409. https://doi.org/10.5194/essd-11-1385-2019

Lloyd, C. T., Sorichetta, A. & Tatem, A. J. (2017). High resolution global gridded data for use in population studies. *Scientific data,* 4, 170001, https://doi.org/10.1038/sdata.2017.1.

Lopez, A., Coughlan de Perez, E., Bazo, J., Suarez, P., van den Hurk, B., & van Aalst, M. (2020). Bridging forecast verification and humanitarian decisions: A valuation approach for setting up action-oriented early warnings. *Weather and Climate Extremes*, 27, 100167. https://doi.org/https://doi.org/10.1016/j.wace.2018.03.006

International Organisation for Migration (IOM) (2021). *Mozambique tropical cyclone Eloise response: situation report #1 25 January - 12 February 2021*. https://reliefweb.int/report/mozambique/mozambique-tropical-cyclone-eloise-response-situation-report-1-25-january-12

Magnusson, L. (2019) ECMWF Severe Event Catalogue for Evaluation of Multi-scale Prediction of Extreme Weather, ECMWF Technical Memoranda, 851, ECMWF, Reading, UK. 10.21957/i2pbf6pe

Mitheu, F., Stephens, E., Tarnavsky, E., Ficchi, A., Cornforth, R., & Petty, C. (2022) Towards a community-led approach to improve the design of early warning systems and anticipatory action for flood risk preparedness, EGU General Assembly 2022, Vienna, Austria, 23–27 May 2022, EGU22-9300, https://doi.org/10.5194/egusphere-egu22-9300, 2022.

Merz, B., Kuhlicke, C., Kunz, M., Pittore, M., Babeyko, A., Bresch, D. N., et al. (2020). Impact forecasting to support emergency management of natural hazards. *Reviews of Geophysics*, 58, e2020RG000704. https://doi.org/10.1029/2020RG000704

Murphy, A. H. (1993). What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting. *Weather and Forecasting*, 8(2), 281-293. https://doi.org/10.1175/1520-0434(1993)008<0281:Wiagfa>2.0.Co;2

Nauman, C., Anderson, E.., Coughlan de Perez, E., Kruczkiewicz, A., McClain, S., Markert, A. , Griffin, R. & Suarez. P., (2021). Perspectives on flood forecast-based early action and opportunities for Earth observations. *J. Appl. Rem. Sens*., 15(3). https://doi.org/ https://doi.org/10.1117/1.JRS.15.032002

Neal, J., Hawker, L., Savage, J., Durand, M., Bates, P., & Sampson, C. (2021). Estimating river channel bathymetry in large scale flood inundation models. *Water Resources Research*, *57*(5), e2020WR028301.

Neal, J., Schumann, G., & Bates, P. (2012). A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas. *Water Resources Research*, 48(11). 10.1029/2012WR012514

NOAA. (2021). *Record-breaking Atlantic hurricane season draws to an end* https://www.noaa.gov/media-release/record-breaking-atlantic-hurricane-season-draws-to-end

Norton, R., MacClune, K., & Szönyi, M. (2020). When the unprecedented becomes precedented: Learning from Cyclones Idai and Kenneth. https://www.i-s-e-t.org/perc-cyclone-idai-2019

Pagano, T. C., Pappenberger, F., Wood, A. W., Ramos, M.-H., Persson, A., & Anderson, B. (2016). Automation and human expertise in operational river forecasting. *WIREs Water*, 3(5), 692-705. https://doi.org/https://doi.org/10.1002/wat2.1163

Parker, W., S. (2020). Model Evaluation: An Adequacy-for-Purpose View. *Philosophy of Science*, 87(3), 457-477. https://doi.org/10.1086/708691

Quinn, N., Bates, P. D., Neal, J., Smith, A., Wing, O., Sampson, C., Smith, J. & Heffernan, L. (2019). The spatial dependence of flood hazard and risk in the United States. *Water Resources Research*, 55, 1890– 1911. https://doi.org/10.1029/2018WR024205

Reliefweb. (2021). *Tropical Cyclone Eloise - Jan 2021. Retrieved 15th September from* https://reliefweb.int/disaster/tc-2021-000008-moz#overview

Robbins, J.C., & Titley, H. A. (2018). Evaluating high-impact precipitation forecasts from the Met Office Global Hazard Map (GHM) using a global impact database. *Meteorological Applications*, 25(4), 548-560. https://doi.org/https://doi.org/10.1002/met.1720

Robbins, J., Bee, E., Sneddon, A., Brown, S., Stephens, E., & Amuron, I. (2022) *Gaining user insights into the research-to-operational elements of Impact-based Forecasting (IbF) from within the SHEAR programme : summary of findings.* NERC, pp 21. https://nora.nerc.ac.uk/id/eprint/532837/

Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., Alfieri, L., & Freer, J. E. (2015), A high-resolution global flood hazard model, *Water Resour. Res.*, 51, 7358– 7381, doi:10.1002/2015WR016954.

Shultz, J. M., Berg, R. C., Kossin, J. P., Burkle Jr, F., Maggioni, A., Pinilla Escobar, V. A., Castillo, M. N., Espinel, Z., & Galea, S. (2021). Convergence of climate-driven hurricanes and COVID-19: The impact

of 2020 hurricanes Eta and Iota on Nicaragua. *The Journal of Climate Change and Health*, 3, 100019. https://doi.org/https://doi.org/10.1016/j.joclim.2021.100019

Smith, A., Sampson, C., & Bates, P. (2015). Regional flood frequency analysis at the global scale, *Water Resour. Res.*, 51, 539– 553, doi:10.1002/2014WR015814.

Smith, A., Bates, P. D., Wing, O., Sampson, C., Quinn, N., & Neal, J. (2019). New estimates of flood exposure in developing countries using high-resolution population data. *Nature Communications*, *10*(1), 1814. https://doi.org/10.1038/s41467-019-09282-y

Speight, L., Stephens, L., Grey, S., Neal, J., Hawker, L., Baugh, C., Prudhomme, C., Savage, J., Cloke, H., Ficchi, A., & Heap, R. (2021) *Flood Early Warnings Pilot Study: learning and recommendations*. Report to FCDO. DER6401-RT004-R00-01. HR Wallingford, Wallingford, pp28.

Stevens, F.R., Gaughan, A.E., Linard, C. & Tatem, A.J. (2015) Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PloS one*, 10(2), p.e0107042. https://doi.org/10.1371/journal.pone.0107042

Stewart, S. R. (2021).: Hurricane Iota (AL312020) 13-18 November, *National Hurricane Centre tropical cyclone report*. NOAA. https://www.nhc.noaa.gov/data/tcr/AL312020_Iota.pdf

Titley, H. A., Cloke, H. L., Harrigan, S., Pappenberger, F., Prudhomme, C., Robbins, J. C., Stephens, E. M., & Zsoter, E. (2021). Key Factors Influencing the Severity of Fluvial Flood Hazard from Tropical Cyclones. *Journal of Hydrometeorology*, 22(7), 1801-1817. https://doi.org/10.1175/jhm-d-20-0250.1

Titley, H.A., Cloke H.L., Stephens, E.M., Pappenberger, F, and Zsoter, E. (2023) Analysing predictability links in the tropical cyclone flood forecast chain using an ensemble-based technique. Submitted to *Journal of Hydrometeorology*

Towner, J., Cloke, H. L., Zsoter, E., Flamig, Z., Hoch, J. M., Bazo, J., Coughlan de Perez, E., & Stephens, E. M. (2019). Assessing the performance of global hydrological models for capturing peak river flows in the Amazon basin. *Hydrol. Earth Syst. Sci.*, 23(7), 3057-3080. https://doi.org/10.5194/hess-23-3057-2019

Trigg M A *et al* 2016 The credibility challenge for global fluvial flood risk analysis *Environ. Res. Lett.* **11** 094014. https://doi.org/10.1088/1748-9326/11/9/094014

United Nations (2021). *Mozambique: UN responds as thousands are caught in the wake of devastating Cyclone Eloise*. Retrieved 16th September from https://news.un.org/en/story/2021/01/1082972

UN OCHA (2020a). *Latin America and The Caribbean: 2020 hurricane seasons - situation report no. 4*. https://reliefweb.int/sites/reliefweb.int/files/resources/20201120_CA%20Eta%20SitRep%204%20ENG.pdf

UN OCHA (2021). *Southern Africa: Humanitarian snapshot.* https://reliefweb.int/sites/reliefweb.int/files/resources/ROSEA_20210520_SA_humanitarian_snapshot.pdf

UN OCHA (2020b). *Central America and Mexico: 2020 Hurricane season - situation report no. 5.* https://reliefweb.int/sites/reliefweb.int/files/resources/20201126%20CA%20Eta%20SitRep%205%20%28ENG%29.pdf

Venkateswaran, K., MacClune, K., Keating, A., & Szönyi, M. (2020) The PERC manual Learning from disasters to build resilience: A guide to conducting a Post-Event Review. Zurich Insurance Company

Ltd, Zurich, Switzerland. pp40. https://www.zurich.com/-/media/project/zurich/dotcom/sustainability/docs/the-perc-manual.pdf accessed 5/7/2022

World Meteorological Organisation (2021) WMO Guidelines on Multi-hazard Impact-based Forecast and Warning Services (WMO-No. 1150), Part II: Putting Multi-Hazard IBFWS into Practice. WMO, Geneva, Switzerland. pp63. https://library.wmo.int/?lvl=notice_display&id=21994#.YsVoDnbMLIU

Ward, P. J., Jongman, B., Salamon, P., Simpson, A., Bates, P., De Groeve, T., Muis, S., de Perez, E. C., Rudari, R., Trigg, M. A., & Winsemius, H. C. (2015). Usefulness and limitations of global flood risk models. *Nature Climate Change*, 5(8), 712-715. https://doi.org/10.1038/nclimate2742

Werner, K., Verkade, J. S., & Pagano, T. C. (2016). Application of Hydrological Forecast Verification Information. In Q. Duan, F. Pappenberger, J. Thielen, A. Wood, H. L. Cloke, & J. C. Schaake (Eds.), *Handbook of Hydrometeorological Ensemble Forecasting* (pp. 1-21). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40457-3_7-1

World Meteorological Organisation (2015) WMO Guidelines on Multi-hazard Impact-based Forecast and Warning Services. WMO- No. 1150. Geneva, Switzerland. pp34 https://library.wmo.int/index.php?lvl=notice_display&id=17257#.YsVppnbMLIU

World Meteorological Organisation (2021) WMO Guidelines on Multi-hazard Impact-based Forecast and Warning Services (WMO-No. 1150), Part II: Putting Multi-Hazard IBFWS into Practice. WMO, Geneva, Switzerland. pp63. https://library.wmo.int/?lvl=notice_display&id=21994#.YsVoDnbMLIU

WMO. (2021). *Tropical Cyclone Eloise hits Mozambique*. Retrieved 16th September from https://public.wmo.int/en/media/news/tropical-cyclone-eloise-hits-mozambique

Wu, W., Emerton, R., Duan, Q., Wood, A.W., Wetterhall, F., Robertson, D.E. (2020). Ensemble flood forecasting: Current status and future opportunities. *WIREs Water*. 7:e1432. https://doi.org/10.1002/wat2.1432

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J.C., Sampson, C.C., Kanae, S. & Bates, P.D., 2017. A high-accuracy map of global terrain elevations. *Geophysical Research Letters*, *44*(11), pp.5844-5853. https://doi.org/10.1002/2017GL072874

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P.D., Allen, G.H. & Pavelsky, T.M., 2019. MERIT Hydro: A high-resolution global hydrography map based on latest topography dataset. *Water Resources Research*, *55*(6), pp.5053-5073. https://doi.org/10.1029/2019WR024873

Zajac, Z., Revilla-Romero, B., Salamon, P., Burek, P., Hirpa, F. A., & Beck, H. (2017). The impact of lake and reservoir parameterization on global streamflow simulation. *Journal of Hydrology*, 548, 552-568. https://doi.org/10.1016/j.jhydrol.2017.03.022

Zhao, G., Bates, P., Neal, J., & Pang, B. (2021). Design flood estimation for global river networks based on machine learning models. *Hydrology and Earth System Sciences*, *25*(11), 5981-5999. https://doi.org/10.5194/hess-25-5981-2021

Zsoter, E., Prudhomme, C., Stephens, E., Pappenberger, F., & Cloke, H.L. (2020). Using ensemble reforecasts to generate flood thresholds for improved global flood forecasting. *Journal of Flood Risk Management*, 13(4), e12658. https://doi.org/https://doi.org/10.1111/jfr3.12658