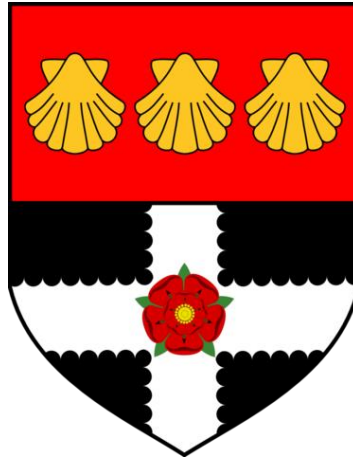


UNIVERSITY OF READING

Department of Geography & Environmental Science



**Understanding global flood hazard
climatology for improved early flood
warnings**

Ervin Zsoter

A thesis submitted for the degree of Doctor of Philosophy

September 2022

Declaration

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

Ervin Zsoter

Abstract

Earth system models have become viable alternatives to traditional hydrological models in supporting global hydrological forecasting of river flow during the last decades. Hydrological forecasting systems rely on a climatology of flood hazard to derive flood thresholds, which are used to generate early flood warnings. However, the derivation of these climatologies is not straightforward and limitations, errors and uncertainties may play a major role and can significantly influence the quality of the flood warnings.

This thesis evaluates some of the crucial characteristics of the reanalysis data sets used to produce the flood hazard climatologies, such as the land data assimilation and snow scheme complexity. Limitations of these data sets are identified, with suggestions presented to further improve the hydrological modelling and threshold generation methodologies. This in turn will lead to improved climatologies as crucial elements in delivering higher quality flood warnings.

It was found that increments produced by the land data assimilation of snow and soil moisture can lead to systematic water budget errors and subsequently contribute to significant errors in river discharge simulations. Results have also shown that a more complex snow scheme with multiple layers can generally improve river discharge unless there is permafrost, where improvements required further adjustments of the snow and soil freezing parametrisations. In addition, the linear trend analysis of a state-of-the-art hydrological reanalysis data set revealed widespread, dominantly negative trends globally, that can adversely impact on the use of thresholds in flood warnings, derived from these reanalyses. In order to improve the quality of the flood hazard climatologies, an alternative threshold generation method, using ensemble reforecasts, has been developed and shown to deliver vastly improved forecast reliability and skill.

This thesis contributes to better understanding of the global flood hazard climatologies in Earth system models and implements a more sophisticated method for producing these climatologies which will deliver better flood warnings. Additional research avenues are also recommended to further improve the hydrological representation of the Earth system models and the generation of the flood hazard climatologies in order to achieve the best possible hydrological forecast quality.

Acknowledgment

I would first and foremost like to thank my supervisors, Hannah Cloke and Liz Stephens. Their support, advice and guidance throughout my PhD was second to none and a huge driving force for me to go all the way in this PhD. They never stopped pushing me forward, for which I will always be indebted to them. Further thanks go to my external supervisor and line manager, Christel Prudhomme, who has always wholeheartedly supported me with my PhD studies and has always helped me fitting the PhD work with my work schedule. I am ever so grateful for the three of them for helping me learn writing better scientific papers and for all the inspiring discussions about the intriguing details of climatologies and global flood forecasting.

Further special thanks go to Florian Pappenberger, to whom I will ever be grateful for inviting me into the flood group at ECMWF and for supporting me all the way in my PhD studies from the management of the Forecast Department. Similarly, I would like to thank my other line manager, David Richardson, who as head of the Evaluation Section has always supported me over the years, even dating back many years before I started my PhD. I will be grateful for them for all the interesting discussions about forecasting in weather and hydrology.

I would also like to thank my friends and colleagues in the flood group and in other sections at ECMWF and in the Water@Reading research group at the University of Reading. I have always felt welcome in both communities and I can only be grateful for the support I was given and for the fun we have had together over the past years during my PhD and even beyond.

I am also grateful to all the friends and family who have had to listen to me talking about weather and in the last few years also floods, for years and years without too much complaint.

Finally, I am incredibly grateful to my family, my devoted parents, Katalin and Mihaly, who sadly passed away quite recently, my wonderful wife, Mariann, and now three fantastic daughters, Lidia, Greta and Alexandra, who have had to suffer the consequences of me being busy 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	ii
Acknowledgment.....	iii
Contents	iv
List of acronyms.....	ix
My personal story of the PhD.....	xiii
Chapter 1 Introduction	1
1.1 Motivation and aims.....	1
1.2 Structure of the thesis.....	5
Chapter 2 Literature review.....	9
2.1 Advances in Earth system modelling and the impact of land data assimilation on hydrology	9
2.2 Importance of snow scheme complexity on the quality of global hydrological simulations .	13
2.3 Trends in hydrological reanalysis data sets	16
2.4 Innovative ways of generating global flood thresholds	19
Chapter 3 Methods and data sets	23
3.1 Hydrological and land-surface models	23
3.1.1 ECLand	23
3.1.2 Offline simulation methodology.....	24
3.1.3 Lisflood.....	24
3.1.4 Lisflood-routing.....	24
3.1.5 CaMa-Flood.....	25
3.2 Meteorological forcing data sets.....	25
3.2.1 ERA-Interim climate reanalysis.....	25
3.2.2 ERA-Interim/Land climate reanalysis.....	25
3.2.3 ERA-20CM climate reanalysis	26
3.2.4 ERA5 climate reanalysis	26
3.2.5 ERA5-Land climate reanalysis	26
3.2.6 ENS forecasts	27
3.2.7 SEAS5 forecasts.....	27
3.2.8 ENS reforecasts.....	27
3.2.9 SEAS5 reforecasts	27
3.3 Hydrological modelling systems.....	28
3.3.1 Global Flood Awareness System (GloFAS).....	28
3.3.2 ECLand/CaMa-Flood system.....	31
3.4 Hydrological data sets	32

3.4.1 Data production	38
3.5 Limitations of the modelling systems and data sets used	41
Chapter 4 How Well Do Operational Numerical Weather Prediction Configurations Represent Hydrology?	43
4.1 Introduction	44
4.2 System description, datasets, and methods	45
4.2.1 Land surface model HTESSEL	46
4.2.2 Land data assimilation.....	46
4.2.3 CaMa-Flood river routing	48
4.2.4 GloFAS	48
4.2.5 Offline land surface modelling	48
4.2.6 ERA5 reanalysis	49
4.2.7 Experimental setup	49
4.2.8 River discharge observations	50
4.2.9 Annual peak river discharge.....	51
4.2.10 Water budget increments	51
4.2.11 Daily 2-m temperature and snow depth.....	51
4.2.12 Climatologies.....	52
4.2.13 Verification statistics	53
4.3 Results	54
4.3.1 Snow depth and 2-m temperature impact	54
4.3.2 Global water budget analysis	55
4.3.3 Catchment-level process examination.....	57
4.3.4 Regionally representative catchments.....	61
4.3.5 Global river discharge analysis	63
4.4 Discussion.....	66
4.5 Conclusions	69
4.5.1 Implications for hydrological forecasting.....	69
4.5.2 Implications for land surface modelling and data assimilation	70
Chapter 5 Hydrological Impact of the New ECMWF Multi-Layer Snow Scheme	73
5.1 Introduction	74
5.2 Materials and methods	76
5.2.1 ECLand land-surface model and offline methodology.....	76
5.2.2 ERA5 reanalysis	77
5.2.3 CaMa-Flood river-routing.....	77
5.2.4 ECLand snow and soil freezing schemes	77

5.2.5 River catchment selection	80
5.2.6 Verification statistics	82
5.2.7 Daily climatology computation.....	82
5.2.8 Experimental setup.....	83
5.3 Results	87
5.3.1 Default multi-layer vs. single-layer snow schemes	87
5.3.2 ML struggles in permafrost.....	90
5.3.3 Improving the multi-layer snow scheme performance in permafrost.....	91
5.3.4 Global impact of the ECLand experiments	95
5.4 Discussion	96
5.5 Conclusions.....	99
5.6 Supplementary figures	101
Chapter 6 Trends in the GloFAS-ERA5 river discharge reanalysis	105
6.1 Introduction.....	106
6.2 Data and methods	106
6.2.1 Global Flood Awareness System.....	106
6.2.2 ERA5 and ERA5-Land	108
6.2.3 GloFAS-ERA5 reanalysis	109
6.2.4 Trend analysis methodology.....	109
6.3 Results	116
6.3.1 Global land average annual mean time series for ERA5 and ERA5-Land	116
6.3.2 Trends in GloFAS-ERA5 river discharge	118
6.3.3 Trends in precipitation, snowfall and 2m temperature	120
6.3.4 Trends in runoff and snowmelt	123
6.3.5 Trends in evaporation, precipitation minus evaporation and soil moisture.....	126
6.3.6 Trends in observed river discharge	127
6.3.7 Trend error comparison ERA5 vs ERA5-Land.....	129
6.3.8 ERA5 and ERA5-Land trends in 1981-2003 and 2004-2018.....	130
6.4 Conclusions.....	132
Chapter 7 Using ensemble reforecasts to generate flood thresholds for improved global flood forecasting.....	137
7.1 Introduction.....	138
7.2 System description, datasets and methods	141
7.2.1 GloFAS.....	141
7.2.2 River discharge reanalysis.....	141
7.2.3 Ensemble river discharge reforecasts.....	142

7.2.4 Flood thresholds.....	142
7.2.5 River catchments.....	143
7.2.6 Analysis methods	143
7.2.7 Experimental set-up	143
7.3 Results	145
7.3.1 How similar are the flood thresholds?	145
7.3.2 How reliable are the forecast probabilities?.....	148
7.3.3 What is the impact on forecast reliability and skill?	149
7.4 Discussion.....	151
7.4.1 Ensemble member independence	151
7.4.2 Best performing thresholds	152
7.4.3 Biases in the forecasts.....	152
7.4.4 Forecast post-processing	153
7.4.5 Modelling system independence	153
7.4.6 Practical recommendations for flood applications	153
7.5 Conclusions	154
7.6 Supplementary figures.....	156
Chapter 8 Summary of co-authored PhD publications	167
8.1 Exploring hydrological reanalysis data sets	167
8.1.1 Cao et al. (2022)	167
8.1.2 Winkelbauer et al. (2022).....	167
8.1.3 Ficchi et al. (2021)	167
8.1.4 Titley et al. (2021)	168
8.1.5 Muñoz-Sabater et al. (2021)	168
8.1.6 Harrigan et al. (2020b)	168
8.1.7 Alfieri et al. (2020).....	168
8.1.8 Towner et al. (2019)	169
8.1.9 Hersbach et al. (2018)	169
8.1.10 Hirpa et al. (2018b)	169
8.1.11 Emerton et al. (2017)	170
8.2 Exploring hydrological forecast data sets	170
8.2.1 Harrigan, et al. (2023)	170
8.2.2 Bischiniotis et al. (2020)	170
8.2.3 Passerotti et al. (2020)	170
8.2.4 Bischiniotis et al. (2019)	171
8.2.5 Emerton et al. (2018)	171

8.3 Exploring meteorological forcing data sets	171
8.3.1 Lavers et al. (2018)	171
8.3.2 Coughlan de Perez et al. (2018).....	172
8.3.3 Lavers et al. (2017)	172
8.3.4 Lavers et al. (2016)	172
8.4 Exploring global flood prediction methodologies	173
8.4.1 Baugh et al. (2020).....	173
8.4.2 Alfieri et al. (2019)	173
8.4.3 Santoro et al. (2018)	173
8.4.4 Hirpa et al. (2018a)	173
8.4.5 Alfieri et al. (2018)	174
8.4.6 Zsoter et al. (2016).....	174
Chapter 9 Discussion	175
9.1 Key results and discussion points	176
9.2 Next steps	182
9.3 Limitations of the thesis	185
Chapter 10 Conclusions	189
References	193
Appendix.....	217
A1: How well do operational numerical weather prediction setups represent hydrology?	219
A2: Hydrological impact of the new ECMWF multi-layer snow scheme	241
A3: Trends in the GloFAS-ERA5 river discharge reanalysis.....	267
A4: Using ensemble reforecasts to generate flood thresholds for improved global flood forecasting.....	343
A5: Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0	359
A6: GloFAS-ERA5 operational global river discharge reanalysis 1979-present.....	381
A7: Daily ensemble discharge reforecasts and real-time forecasts from the operational Global Flood Awareness System.....	401
A8: Additional figures	423

List of acronyms

4D-Var	Four-dimensional variational data assimilation (4D-Var)
AEP	Annual exceedance probability
AMSU	Advanced Microwave Sounding Unit
API	Application Programming Interface
ASCAT	Advanced Scatterometer
CaLDAS	Canadian land data assimilation system
ECMWF	European Centre for Medium-range Weather Forecasting
C3S	Copernicus Climate Change Service
CDS	Copernicus Climate Data Store
CEH	Centre for Ecology and Hydrology
CEMS	Copernicus Emergency Management Service
CMIP5	Coupled Model Intercomparison Project phase 5
CMORPH	Climate Prediction Center Morphing Technique precipitation analysis
CNRM	National Center for Meteorological Research
CPU	Central Processing Unit
DHS	Data handling system
DIS	River discharge
E	Evaporation
ECLand	ECMWF's land-surface modelling platform
ECFS	ECMWF's File Storage System
EFAS	European Flood Awareness System
EFI	Extreme Forecast Index
ENS	ECMWF ensemble forecasts
ENSO	El Niño Southern Oscillation
ERA5	ECMWF reanalysis version 5
ERA5T	Fast release version of ERA5
ERA-Interim	ECMWF ERA-Interim reanalysis, replaced by ERA5
ESM	Earth System Model
EU	European Union
EVD	Extreme Value Distribution
EWS	Early Warning System
FbF	Forecast-based Financing
FFIR	Flooding From Intense Rainfall

GEOSS	Global Earth Observation System of Systems
GFM	Global Flood Monitoring
GFP	Global Flood Partnership
GLDAS	Global Land Data Assimilation System
GloFAS	Global Flood Awareness System
GLWD	Global Lakes and Wetlands Database
GPCC	Global Precipitation Climatology Centre
GPCP	Global Precipitation Climate Project
GRAND	Global Reservoir and Dam Database
GRDC	Global Runoff Data Centre
HTESSEL	Hydrology Tiled ECWMF Scheme for Surface Exchange over Land
HPCF	High Performance Computer Facility
HYPE	Hydrological Predictions for the Environment model
IFS	Integrated Forecasting System
IMS	Interactive Multisensor Snow and Ice Mapping System
IPCC	Intergovernmental Panel on Climate Change
IPSL	Institut Pierre Simon Laplace
ISBA	Interactions Soil-Biosphere-Atmosphere land surface scheme
IVT	Integrated water Vapour Transport
JMA	Japan Meteorological Agency
JRA	Japanese reanalysis by JMA
JRC	Joint Research Centre of the European Commission
JULES	Joint UK Land Environment Simulator
KGE	Kling–Gupta Efficiency
LDAS	Land Data Assimilation System
Lisflood	Spatially distributed water resources model, developed by the Joint Research Centre of the European Commission
LSDA	Land Surface Data Assimilation
LSM	Land-Surface Model
MARS	Meteorological Archival and Retrieval System
NAO	North Atlantic Oscillation
NASA	National Aeronautics and Space Administration
MAE	Mean Absolute Error
ME	Mean Error

MERRA	Modern-Era Retrospective analysis for Research and Applications by NASA
MLS	Multi-Layer Snow scheme
NCAR	National Centre for Atmospheric Research
NCEP	National Centres for Environmental Predictions
NESDIS	National Environmental Satellite, Data and Information Service
NOAH	Community land-surface model, developed in the USA
NOAA	National Oceanic and Atmospheric Administration
NSE	Nash–Sutcliffe model Efficiency
NWP	Numerical Weather Prediction
OBS	Observations
ORCHIDEE	Organising Carbon and Hydrology In Dynamic Ecosystems land surface model
P	Precipitation
PCORR	Pearson correlation coefficient
R	Pearson correlation coefficient
RECF	Reforecast-control-member-based reanalysis
RFC	Reforecast
RO	Runoff
SD	Snow depth
SEKF	Simplified Extended Kalman Filter
SF	Snowfall
SHEAR	Science for Humanitarian Emergencies and Resilience
SM-MW	Microwave-based multisatellite surface soil moisture dataset
SMLT	Snowmelt
SLS	Single-Layer Snow scheme
SRFC	Seasonal reforecast
SRO	Surface runoff
SSA	Stand-alone Surface Analysis
SSRO	Sub-surface runoff
SWE	Snow water equivalent
SWV	Soil water volume
SYNOP	Surface synoptic weather observations
T2	2m temperature
TIGGE	THORPEX Interactive Grand Global Ensemble

TRMM	Tropical Rainfall Measuring Mission
UNDRR	United Nations Office for Disaster Risk Reduction
UoR	University of Reading
USA	United states of America
USD	United States dollar
WMO	World Meteorological Organization

My personal story of the PhD

The road to the PhD thesis, with the achievement of all the results that could be presented about global flood hazard climatologies, has to start with who I am and what is my personal PhD experience.

My name is Ervin Zsoter and I had the privilege to have written this PhD thesis. The start in September 2016, seems more like a lifetime ago, maybe it was not even in this life. But with all the work and stress that it has come with, I have enjoyed every second of it.

The fact that I have started this PhD, I must owe quite simply to Hannah Cloke and Florian Pappenberger, who presented the opportunity to me in 2016, including the possibility of the Wilkie Calvert co-supported PhD studentships (which I was awarded) and convinced me to start this journey. Just as I would not have pursued it, if it was only on me, as I knew how difficult and how much work it was meant to be.

In fact, I started another PhD many years ago in Hungary, but even though I did all the necessary studies, I quite simply could not find the time and energy to write the thesis. At the beginning of that PhD, I was working at the Hungarian Meteorological Service as a weather forecaster and forecast product developer, after graduating as a meteorologist. However, later I had the opportunity to join ECMWF in 2005, where I worked on weather product development for few years. Doing a PhD remotely from another country was way too much, so there was only one outcome of that PhD, an inevitable failure.

However, as a twist of fate, after spending few years back in Hungary, I again had the chance to join ECMWF in 2012. This time, I did end up doing hydrological modelling using the HTESSEL and CaMa-Flood models as part of a European project. This project has brought me the opportunity to join the flood group at ECMWF at the time, led by Florian Pappenberger. I am ever so grateful for Florian for inviting me into this fantastic group and helping me to set out on this fantastic journey in the world of hydrology, which culminated into my PhD studies a few years later.

During my PhD, I have worked in an extensive collaborative environment, most importantly at ECMWF, as I have been involved in the ever-expanding flood group in the Evaluation Section of the Forecast Department. When I joined the group in 2014, it had only maybe 5-6 people, while in 2022 at least 20 people worked in the extended group, spreading across several sections of the Forecast and Research Departments. The external collaborative work, on developing the EFAS and GloFAS systems, was centred around the flood related group at the JRC in Ispra, Italy. This included collaboration with the JRC, who are the managing body of the CEMS service and also the extensive

scientific collaboration by sharing development activities for both EFAS and GloFAS (including amongst others the development of the Lisflood hydrological model or the handling of the river discharge observations) between ECMWF and the JRC.

My work in the flood group at ECMWF has been centred around GloFAS and global hydrological issues in general. The early years of FbF, which GloFAS has been an integral part from its very beginning, coincided with my own first years in the field of hydrology from about 2014. This is how I got to know Hannah Cloke, Elisabeth Stephens and the water@Reading people in general, who the ECMWF flood group have also heavily collaborated.



Figure 0. My beloved helper team, the Zsoter-Kancsar family on vacation in Mallorca.

During the years of the PhD, I had the privilege to work with many fantastic scientists in the field of hydrology, not just internally at ECMWF and JRC, but also externally with PhD students and other colleagues at the University of Reading and other institutions. I have often helped them with GloFAS know-how, or provided them with GloFAS data, which had to be manually extracted, processed and transferred before the introduction of the GloFAS data sets in the CDS in 2019. These collaborations often resulted in a co-authored publication about various aspects of global hydrology (described in Chapter 8). The majority of these were related to GloFAS, with different aspects of the hydrological quality or the applicability of the model in different areas of the globe. In addition, true to my

meteorological heritage, I have also been involved in studies that addressed meteorological aspects of flood forecasting, such as extreme precipitation or temperature conditions.

Doing this PhD was a natural progression for me and a huge opportunity to deepen my hydrological knowledge, whilst continuing the work on GloFAS in my day job. The fact, that I could share the scientific topics of my PhD with my work, was not just a huge advantage but probably the only way I could do this PhD.

As a part-time PhD student, I had 4 to 6 years to submit my thesis, of which 5 years were funded by the studentship. At the beginning in 2016, I have planned for 4 years (as that was the minimum time that had to pass), as probably most of PhD students would have done. In fact, I was actually worried that my PhD findings could just become outdated even after 4 years.

How wrong have I been with my predictions. Just as 6 years does not seem that long at all now. And to my surprise, the field of hydrological science has not progressed that fast either, so I believe my findings actually have stood the test of time much better than I originally anticipated.

However, 6 years is still a long time and the PhD work plan has been adjusted several times during this period. The two studies, that I remember planning from the start, are the land data assimilation and the flood threshold generation impact studies, which I both saw causing issues in the GloFAS system during my work. The other two main studies of the GloFAS-ERA5 trend analysis and the multi-layer snow scheme hydrological impact study both were born later, from issues and ideas that we have found with my supervisors.

In summary, I am super grateful to my supervisors, Hannah Cloke, Elisabeth Stephens and Christel Prudhomme, who always managed to keep me on track, even if the track was not necessarily clear, at least not to me. I also thank Florian Pappenberger for the never-ending support from ECMWF management.

Finally, I am ultimately grateful to my family, my lovely wife and three daughters (Figure 0). My wife is a real hero, who never stopped believing in me throughout these 6 long years, regardless of my continuous complaining about not having enough time, always being late with deadlines and hardly ever being fully happy with my progress. But hopefully, none of that matters now, just as the only due date that truly mattered was the PhD submission date, which I am extremely pleased to have met.

Chapter 1 Introduction

1.1 Motivation and aims

Flooding is undoubtedly amongst the Earth's most common and destructive natural hazards. In the 20-year period from 2000 to 2019, floods have accounted for 44% of all global disaster events, affecting 1.6 billion people worldwide and causing an estimated \$651 billion (USD) in flood damages (UNDRR, 2019). The anticipation and forecasting of flood events are key components of managing flood risk that can help mitigate the impacts of these devastating disasters. An essential tool in this is the use of flood forecasting systems that provide early flood warnings on time scales up to several weeks ahead at local, regional and global scales (Emerton et al., 2016). Guidance on future flood events provides invaluable information for organisations that work globally across a range of sectors from agriculture to humanitarian aid, such as the Red Cross or the World Food Programme.

Global-scale hydrological forecasting has become possible during the last decade, due to improvements in meteorological and hydrological modelling capabilities and the increasing available computational resources (Harrigan et al., 2023).

The methodology behind the global-scale hydrological systems extends from the application of traditional hydrological models to the use of Earth system models. Hydrological models are generally input-to-output models that directly simulate (primarily) the evolution of water fluxes, such as traditional rainfall-streamflow models (Horton et al., 2021). On the other hand, Earth system models seek to simulate all relevant aspects of the Earth system, including interactions between the atmosphere, oceans and land as well as the biosphere and human activities (Schneider et al., 2017).

Operational numerical weather prediction (NWP) centres, such as the European Centre for Medium-Range Weather Forecasts (ECWMF), have already adopted the Earth system approach. Producing hydrological predictions at NWP centres has a clear advantage of working with unique access to resources and expertise. This includes, for example, the exploitation of Earth observations (either available currently, or expected in the future) and the related use in the data assimilation, for variables such as snow, soil moisture, evapotranspiration, groundwater or river discharge, especially with the help of remote sensing (Harrigan et al., 2020a).

On the other hand, the land-surface components of these Earth system models are traditionally designed to provide lower-boundary conditions to the atmosphere and not so much to simulate the hydrological cycle (Prentice et al., 2015). Therefore, even with state-of-the-art models, there can be limitations in the representation of the hydrological cycle by these land-surface models

(LSM), and some important processes could still be inadequately modelled or even neglected for streamflow simulation (Fisher and Koven, 2020).

The success of any global Earth-system-model-based hydrological prediction system will be reliant on the quality of the models involved, i.e. the land-surface model to produce runoff and the river routing model to generate river discharge, after coupling it to the LSM's runoff output. Another major contributing factor is the quality of the meteorological and hydrological data sets that are used to generate the initial conditions that the model will run from. In addition, the quality of the meteorological forcing for the forecasts, that comes from NWP meteorological predictions, will also heavily contribute in determining the quality of the forecast simulations.

The predictability of the forecast will depend on all these modelling, forcing and initial condition errors. Model errors can come from LSM hydrological process representation issues, simplifications in the routing component or NWP model errors in generating the meteorological forcings. Initial condition errors, on the other hand, are related to the use and quality of the different observations and to the way they are assimilated into the reality of the model. These different error sources all contribute to the evolution of biases in the forecasts, which can be quite large for global systems (Harrigan et al., 2020b).

The handling of these errors can be supported by the use of probabilistic predictions in the form of ensemble forecasts. Ensemble forecasts are often generated by the same modelling system but run from perturbed initial conditions to represent the error growth in the future. The state-of-the-art systems in use today, such as the Global Flood Awareness System of the European Commission's Copernicus Emergency Management Service (GloFAS, Harrigan et al., 2020b), provide an ensemble of equally likely solutions for flood events to occur (Cloke & Pappenberger, 2009; Wu et al., 2020). With the use of ensemble forecasts, the predicted events can be expressed in terms of a distribution of possible events. This distribution will show the likelihood of each future outcome, or in other words, the uncertainty that is associated with all these potential events.

In addition, ensemble forecasts can be used to express the flood events as probabilities, by comparing the predictions to predefined flood thresholds. These thresholds are certain climatological properties, such as selected return periods or quantiles, which describe the likelihood of different flood magnitudes to occur locally.

The use of thresholds and the related probabilities offer a way to account for some of the errors and biases, as these errors are expected to feature in a similar way in both the climatology (i.e. the thresholds) and the actual forecasts. This way, systematic errors (e.g. a generic positive bias) will

be removed by transforming the flood signals into the evolution of probabilities exceeding the threshold values.

Traditionally, these thresholds are derived from either observations or model simulations, often forced with meteorological reanalysis datasets (Alfieri et al., 2015), preferably based on a long enough period of at least 30 years (World Meteorological Organisation [WMO], 2017). However, due to the sporadic availability of river discharge observations on the global scale (Lavers et al., 2019; Rodda et al., 1993) and the lack of remote sensing observational alternatives for river discharge to date (Biancamaria et al., 2016), the use of model-simulation-based climatologies (generated from river flow reanalyses) is the only viable option for any hydrological application over the whole global domain.

The quality of the model climatologies is of crucial importance. Firstly, they can demonstrate the hydrological model performance, through the quality of the 30+ year simulation that is used for the model climatology. This will give a major indication about the quality of the forecasts, as the same model system is used to produce both the climatology and the forecasts (Harrigan et al. 2020).

Another very important aspect of the modelling and forecast chain is the consistency of flood event representation between the climatology-derived thresholds and the forecasts. All flow events, especially the extreme ones such as floods and droughts, should inherently be represented in the same way in both the climatology and the forecasts. If this is not the case, the reliability of the forecasts would be negatively impacted (Hirpa et al., 2016; Zsoter et al., 2014). For example, the flood thresholds that are used to determine the severity of the forecasted flood signal, should ideally represent extreme events in the same way, with the same overall frequencies; as they occur in the forecasts. If this is not the case and the different biases make an event of the same magnitude occur with a different frequency in the climatological data set (used to compute the thresholds) than in the forecasts (e.g., the 5% AEP flood magnitude happens more often in the forecasts than the expected 5% probability in a given year), then the flood forecast probabilities could become unreliable (e.g., leading to flood signals that often overestimate the flood severity).

An example when the consistency between climatologies and forecasts will inevitably decrease is the presence of large trends in the data that is used to produce the thresholds (Faulkner et al., 2019; Jiang and Kang, 2019). The non-stationary time series with trends or other discontinuities in them will lead to thresholds (e.g. return periods) which will either be too high (in case of a negative trend) or too low (in case of a positive trend), when compared with the floods that represent the forecasts (assuming the forecasts will behave similarly to the latter part of the climatological data time series).

The derivation of these model climatologies is not straightforward and the quality of the flood predictions largely depends on them (Hirpa et al., 2016). Uncertainties and limitations in these climatologies may play a major role and can significantly influence the forecast skill and ultimately the decision-making process by the users (Coughlan de Perez et al., 2016). Improving the model climatologies can increase the skill of the hydrological forecasts and therefore the usability of such forecasting systems.

The hydrological diagnostic of the components of the modelling chain, from the LSM, the routing model, the NWP modelling creating the forcing, all the way to the use of climatology, is crucial in order to understand the behaviour of the forecasting system (Towner et al., 2019). Identifying the limitations and the contributing errors coming from the model climatology can help improving the quality of the hydrological simulations and ultimately the forecasts.

A systematic analysis of the hydrological impacts of the components, focussing on the climatological simulations, has not yet been done on the global scale. There are many potential areas which could be considered for such an analysis and which promise to bring improved knowledge for better hydrological predictions.

The aim of this research is to understand global flood hazard climatologies, to analyse how well they represent the Earth system and to assess their relevance for application in flood forecasting. This thesis evaluates some of the crucial characteristics of the data sets used to produce the climatologies and aims to identify areas of limitations, where the quality of the applied modelling or input data sets can be improved. In addition, it seeks to develop innovative ways of generating and using the global flood hazard climatologies, which show better consistency with forecasts and are capable of delivering higher skill. This is done through the following specific objectives, concentrating on some specific areas:

1. Analyse how well the land-surface modelling approach in Earth systems is able to support hydrological applications, in particular focussing on the impact of land-data assimilation of snow and soil moisture on the hydrological cycle in reanalysis simulations.
2. Evaluate the hydrological impact of the complexity of the snow scheme in the land-surface models in reanalysis simulations, with special focus on cold climate areas in permafrost.
3. Evaluate the relevant trends in hydrological reanalyses for river discharge and other related land-surface variables and analyse how the interactions amongst these variables contribute to the trends.

4. Assess the impact of innovative ways of generating flood thresholds on the skill of global flood forecasts, using hydrological ensemble reforecasts.

The results presented in this thesis will provide a hydrologically relevant, global-scale analysis of flood hazard climatologies and will identify areas of limitations where the quality of these data sets can be improved. In addition, they will introduce new methodologies to generate higher quality flood climatologies and the related thresholds used in global flood prediction systems. The information presented in this thesis can ultimately contribute to increasing confidence in the flood forecasts, produced with the improved flood hazard climatologies, and should result in better flood preparedness for humanitarian and civil protection partners, thus potentially reducing flood-related damages and casualties world-wide.

1.2 Structure of the thesis

This thesis is structured around four main papers published during the PhD (Figure 1-1). To begin with, Chapter 2 introduces a literature review to substantiate the motivation for carrying out the study of exploring the flood hazard climatologies in this PhD.

Chapter 3 introduces the wealth of methodologies and data sets featuring in the papers published as part of this PhD, summarising the modelling systems used with their respective hydrological models and meteorological forcing data sets.

Chapter 4, the first paper presented in this thesis, provides an analysis of how the land-atmosphere coupling and the land-data assimilation of snow depth and soil moisture impact on the hydrological cycle in ECMWF's Earth system model, in particular the closure of the water budget.

Chapter 5 follows on from the findings of the first paper and explores a potential route to improve the snow processes and their hydrological response in ECMWF's modelling system by introducing the new multi-layer snow scheme, with special focus on cold areas in permafrost.

Chapter 6 addresses an important aspect of the hydrological reanalysis time series and analyses the relevant trends in them. Analysing trends is important as they can significantly impact on the quality of the climatologies and the flood thresholds generated from the reanalysis, potentially affecting the hydrological forecast skill.

Chapter 7, the last major paper of the PhD, explores how the use of ensemble reforecasts can help generating higher quality flood thresholds and delivering improved global flood forecasts, compared with using the traditional reanalysis-based method.

Chapter 8 lists the co-authored papers that were published during the PhD and gives a short summary of the findings and how the papers are connected to the PhD.

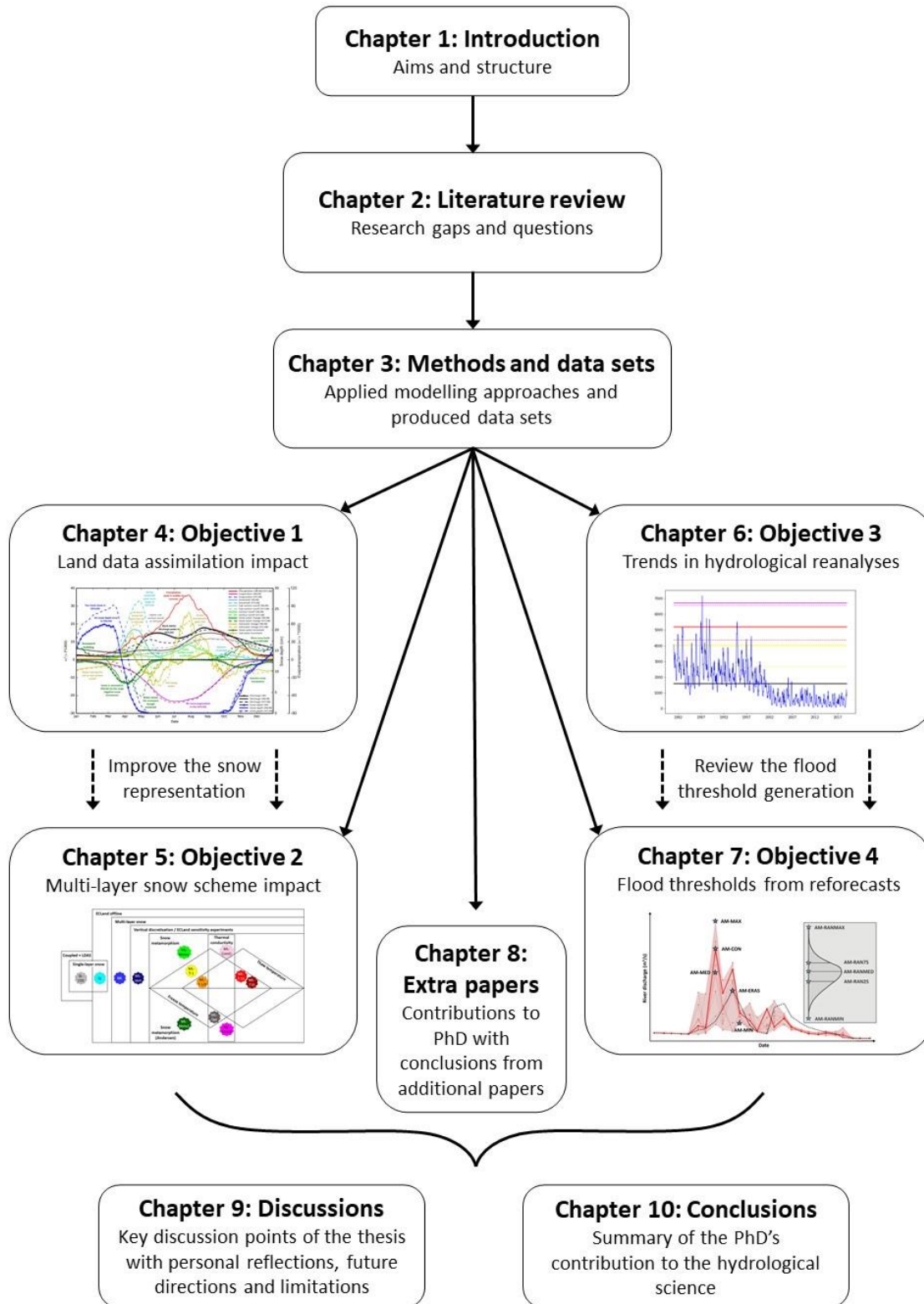


Figure 1-1. Schematic structure of the PhD.

Chapter 9 summarises the findings and the wider contribution of this thesis. It discusses the importance of land data assimilation by potentially opening the water budget and thus contributing to large negative biases in river discharge in the snow-dominant Northern Hemisphere. It also

discusses potential improvements to river discharge by using a physically more complex snow-model. This included a sensitivity study element to the very cold permafrost areas, where improvements were only achievable through additional land-surface parametrisation changes. In addition, this chapter discusses the widespread non-stationary signals found in a state-of-the art hydrological reanalysis for several land surface variables and summarises the new methodology of generating range-dependent flood thresholds from ensemble reforecasts instead of reanalysis data sets, that are able to deliver much improved flood forecast reliability and skill. Moreover, the chapter outlines the scope for future directions, such as ideas for analysing other aspects of land-surface modelling for hydrological contribution and list some limitations, for example the constraints in the availability of the river discharge observations used in the thesis or the extent to which the studies presented here can be generalised.

Finally, Chapter 10 will provide conclusions to the thesis, highlighting the most relevant aspects of the hydrological limitations and possible improvements in reanalysis data sets used for flood threshold generation and also summarising the potential new approach of using ensemble-reforecast for the generation of thresholds. Finally, this chapter also presents the reference list and the appendix with the most important papers of the PhD, attached as published.

Chapter 2 Literature review

2.1 Advances in Earth system modelling and the impact of land data assimilation on hydrology

Earth system models (ESM) are designed to simulate all relevant aspects of the Earth system. They include not only physical processes, such as their predecessor global climate models that primarily worked as coupled atmosphere-ocean models, but also chemical and biological processes or human activities (Flato, 2011; Schneider et al., 2017). Accurately modelling the complex interactions amongst the atmosphere, ocean, sea ice, land surface (including water surfaces) and biosphere (Bonan, 2008), will lead to improved and consistent predictions across variables of the various systems, including also precipitation and river discharge and other hydrological variables (Harrigan et al., 2020a; de Boisseson et al., 2021).

One of the key modelling components of Earth system models are land surface schemes (LSM), which describe the exchange of primarily water, energy and carbon fluxes at the surface-atmosphere interface of the Earth (Pitman, 2003). These models are traditionally used to provide physical boundary conditions to climate and Earth system models supporting atmospheric/climate modelling and forecasting activities. These are needed in terms of energy partitioning, surface roughness, and albedo, which represent the land influence on meteorological processes (Blyth et al., 2021).

As described in Abramowitz et al. (2008), LSMs are typically provided with meteorological conditions as inputs (from an atmospheric model) and produce various outputs, such as latent and sensible heat fluxes, CO₂ fluxes, solar and longwave radiation, surface and subsurface runoff. In addition, they have typical internal state variables such as soil moisture, temperature, vegetation and soil carbon pools and also snow and ice volume and density.

The role of land surface models is not only to provide lower boundary conditions to the atmosphere, but they are also used as the terrestrial component for several other research and operational applications, depending on the represented processes and their degree of complexity. Applications of LSMs can cover various scales from local to global or hourly to climate and various topics, such as climate change, floods and droughts, also agriculture or CO₂ monitoring (Boussetta et al., 2021). Their use in hydrological applications was facilitated by the expansion of the LSMs in scope and complexity through integrating and improving important hydrologically-relevant processes during the last few decades, for example subsurface lateral water movement, river flow routing and water management modules (Archfield et al., 2015; Clark et al., 2015).

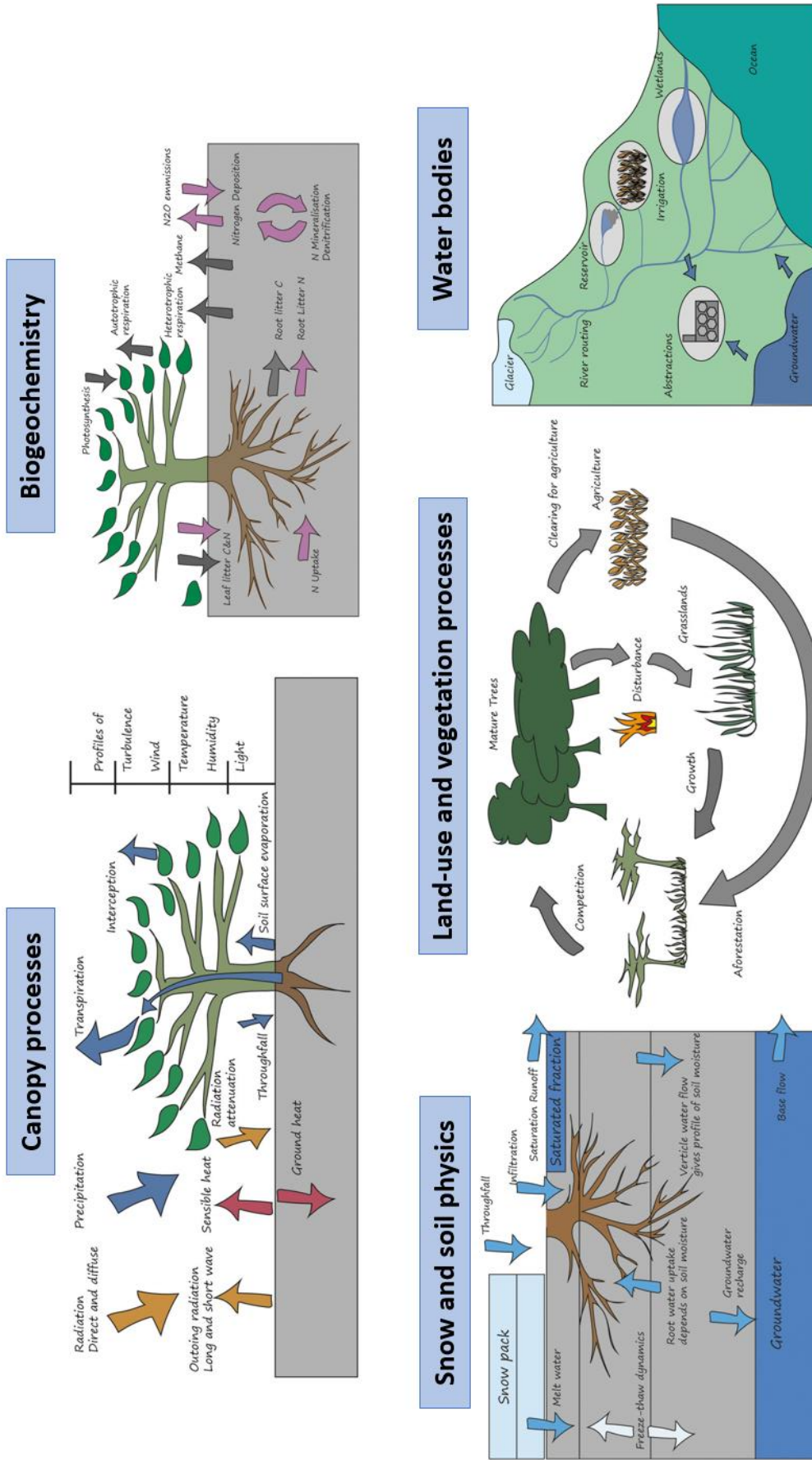


Figure 2-1. Processes in land surface models. After Blyth et al. (2021).

Figure 2-1 highlights the schematic of the wealth of processes that the state of the art LSMs include (based on Blyth et al., 2021). This PhD thesis focuses primarily on the snow and soil physics and the water bodies area of the land surface processes in Figure 2-1.

Earth system models, with LSMs as their land modelling components, are equipped with advanced data assimilation algorithms, which combine the large amount of available Earth observations with model information to produce the best possible estimate of the current state of the Earth system. These initial conditions provide the starting point for the run of the Earth system model to predict the likely evolution of the atmosphere and all related Earth system components (Bouttier and Courtier, 1999; Rabier, 2005).

State-of-the-art NWP centres, such as the European Centre for Medium-Range Weather Forecasts (ECMWF), include both an LSM and land data assimilation systems (LDAS). The LDAS's objective is to combine the LSM state with the available land surface observations to initialize the LSM's prognostic variables of the forecasting system (Bélair et al., 2003a). For example, the current ECMWF LDAS analyses soil moisture, soil temperature, snow mass, density, and temperature (de Rosnay et al., 2014), while other NWP centres all have their own LDAS assimilating, at a minimum, snow and soil moisture, such as the Global Land Data Assimilation System (GLDAS) by NASA/NOAA (Rodell et al., 2004), the Land Surface Data Assimilation (LSDA) by the Met Office (Pullen et al., 2011; Gómez et al., 2020) or the Canadian Land Data Assimilation System (CaLDAS) by Environment Canada (Carrera et al., 2015).

While the role of land data assimilation in LSMs is primarily to improve the atmospheric forecasts (e.g. de Rosnay et al., 2013; Drusch and Viterbo, 2007; Beljaars et al., 1996), LSMs have also been extensively shown to provide improvements for various aspects of hydrological modelling (e.g. Lievens et al., 2015; Clark et al., 2006). However, the related increments of snow or soil moisture will remove or add water, which will potentially open the water budget. In this sense, LSMs are in contrast with traditional hydrological models, which directly concentrate on solving the water balance, while LSMs will commonly tolerate errors in runoff as long as the evaporation and energy fluxes are in balance (Archfield et al., 2015).

By opening the water budget in LSMs, LDAS can potentially have a negative impact on the hydrological cycle (Zaitchik and Rodell, 2009; Arsenault et al., 2013; Andreadis and Lettenmaier, 2006a; De Lannoy et al., 2012; Pan and Wood, 2006; Kauffeldt et al., 2015.). On the contrary, systems with no LDAS could result in the accumulation of various atmospheric forcing and land surface modelling errors, which can also negatively impact on atmospheric or hydrological simulations.

Generally, data assimilation is shown to work best for unbiased systems (Dee and Silva, 1998; Pathiraja et al., 2018). In the presence of systematic biases, e.g. a model bias such as a systematically too slow snowmelt, it is not only that the efficiency of the data assimilation can decrease, but the bias can enhance the impact of the open water budget. In the above example, the too slow snowmelt in the LSM will force the LDAS to remove snow to correct for the excessive snow amounts. This removed water will likely cause a water deficit downstream in the land system, potentially causing an incorrect rate of runoff and ultimately river discharge.

This suggests that by opening the water budget, land data assimilation can cause problems for hydrological forecasting applications, which use runoff produced by LSMs to drive the system. A prime example is GloFAS (Alfieri et al., 2013; Harrigan et al., 2020b; Harrigan et al., 2023), which had used the land surface component (HTESSEL; Balsamo et al., 2009) of the ECLand platform (Boussetta et al., 2021) to provide runoff for all the hydrological simulations in its operational model versions until May 2021 (Alfieri et al., 2020). In fact, Harrigan et al. (2020b) reported large biases in the GloFAS-ERA5 reanalysis (which uses ECLand with land data assimilation) over many global rivers, suggesting that data assimilation could potentially be responsible for some of the biases.

Although the impact of land data assimilation on the water budget has been highlighted in scientific publications, comprehensive analysis of the wider global hydrological impact in a state-of-the-art land surface model has not previously been undertaken.

It would be important to investigate how an LSM with LDAS can support the combined task of traditional weather forecasting and hydrological simulations at the same time. Analysing the impact of the data assimilation on the hydrological cycle and the nature of the water budget issues on the global scale could bring invaluable knowledge in helping the development of the future land-surface schemes in Earth system models.

One way to test the LDAS impact could be with reanalysis simulations that can include land data assimilation. We explore this in Chapter 4 by using the state-of-the-art ECLand land surface modelling platform of ECMWF, together with the meteorological forcing data set of ERA5 (Hersbach et al., 2020). For this, two hydrological simulations are generated, one which assimilates land-surface observations (i.e. at least snow and soil moisture) and another control simulation which does not. By comparing these two simulations (i.e. verifying using available river discharge observations), the hydrological representation of NWP configurations with LSMs in their core will be explored and areas where the coupled Earth system modelling with LDAS does not yet work effectively for flood simulations highlighted.

2.2 Importance of snow scheme complexity on the quality of global hydrological simulations

As systematic biases can negatively impact on the land data assimilation's performance, the exploration of the land surface process limitations, which have the potential to cause biases, is of crucial importance in order to improve the quality of land-surface simulations.

The variability of snow extent and depth is one of the few crucial aspects of land-surface modelling in Earth system models, that can potentially contribute to developing model biases (Pitman, 2003). Snow is important, as it has a major modulating effect of the energy and water fluxes in the land-atmosphere system (Armstrong and Brun, 2008). See Figure 2-2 for the graphical representation of the wealth of snow-related physical processes. Snow impacts on the radiative-energy balance, through the surface albedo (Riihelä et al., 2021), resulting in large temperature changes in certain conditions (Betts et al., 2014). Snow also acts as a water reservoir, releasing snowmelt (mostly in spring) and influencing the entire hydrological cycle through runoff, soil moisture, evaporation and indirectly also precipitation (Adam et al., 2009; Slater et al., 2001; López-Moreno and García-Ruiz, 2004; Griessinger et al., 2016).

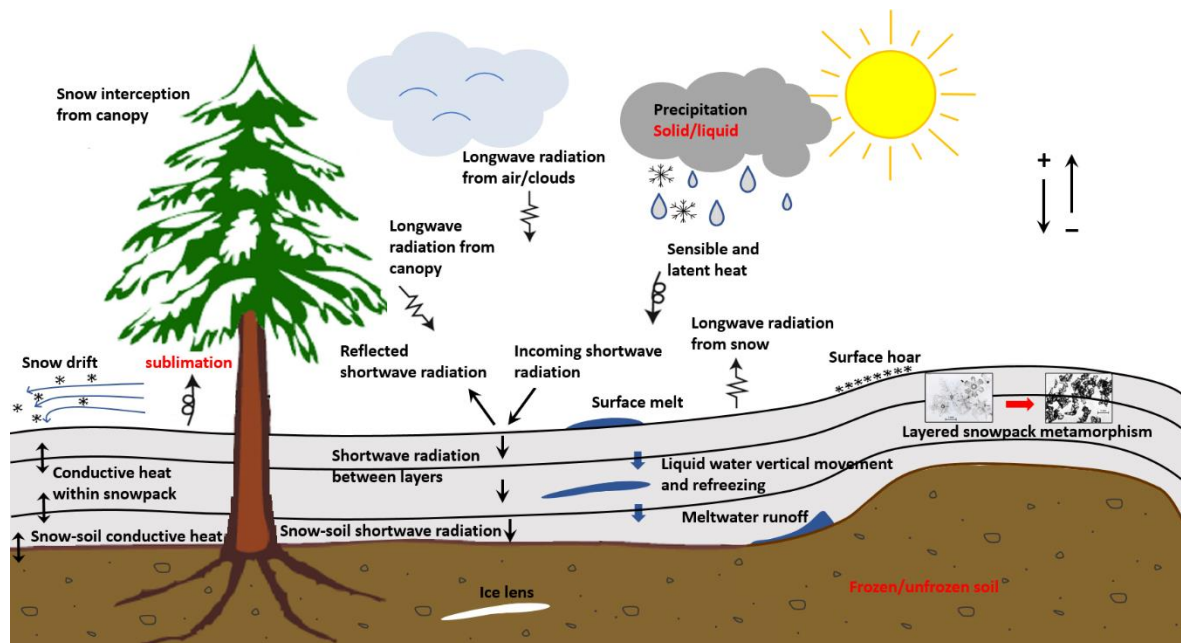


Figure 2-2. Schematic of principal snow-related physical processes. Adapted from Koczo et al. (2005), Arduini et al. (2019) and the SNOWPACK model description (<https://www.slf.ch/en/services-and-products/snowpack.html>).

Snow can be represented by snow schemes with different complexities in the land surface models, which differ in the snow parameterizations in terms of variables and processes considered. Boone and Etchevers (2001) have categorised snow schemes into three general classifications based on their complexities. These include the simplest models which use a composite snow–soil–vegetation energy budget and a single snow layer; the ‘intermediate complexity’ schemes that are based on

more complex physical parameterization with multiple layers (usually 2-10); and finally the detailed internal-snow-process schemes with detailed physical parameterization, usually used e.g. in avalanche predictions (see e.g. Vionnet et al., 2012). Largeron et al. (2020) summarised these model categories, including also the simplest degree-day schemes widely used in hydrological models, according to their physical processes and included state variables (Figure 2-3).

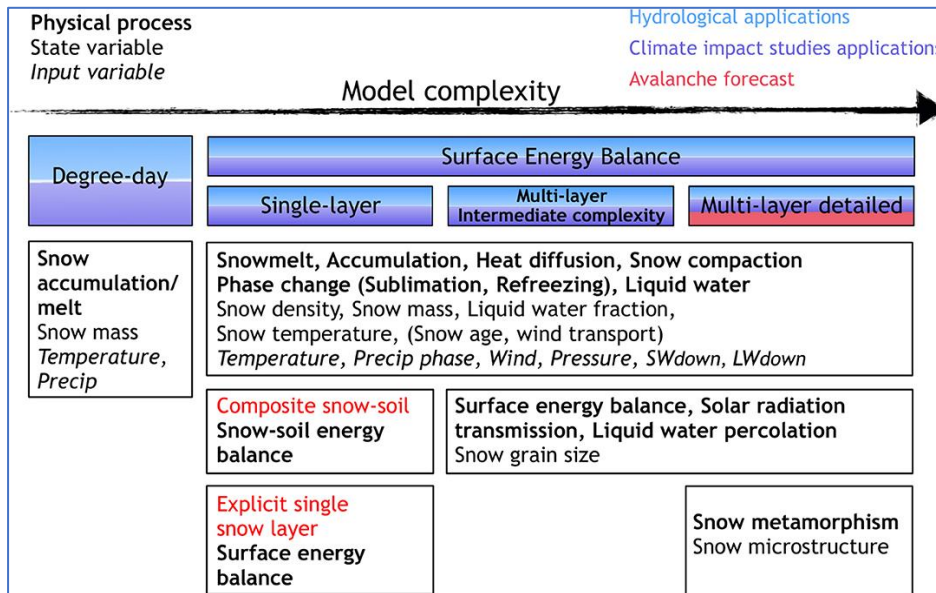


Figure 2-3. Description of snow physical processes (**bold**), state variables (*regular*) and input variables (*italic*) required per category of snow models complexity for offline applications. Taken from Largeron et al. (2020).

The intermediate complexity snow schemes, that are characterised by multiple layers and state variables such as snow density, snow thickness, temperature and liquid water content, allowing the calculation of vertical gradient of temperature and density, offer significant improvement on the single-layer schemes with better handling of the snow processes (e.g. Dutra et al., 2012; Burke et al., 2013). These schemes are widely used in land surface components of Earth system models, such as the land surface models of ECLand developed by ECMWF (Dutra et al., 2012); NOAH developed by NCAR, NCEP, NASA and university groups in the USA (Saha et al., 2017); JULES developed by researchers in the United Kingdom, coordinated by the Met Office and the UK Centre for Ecology and Hydrology (UKCEH) (Wiltshire et al., 2020; Walters et al., 2019); ISBA developed by the National Center for Meteorological Research (CNRM) in France (Decharme et al., 2016) and ORCHIDEE developed by the Institut Pierre Simon Laplace (IPSL) in France (d'Orgeval et al., 2008; Wang et al., 2013).

The higher complexity snow schemes, with the improving realism of representing the snow processes, have been shown to improve various aspects of the land-atmosphere processes, such as decreasing the biases in snow depth, melting timing or temperature conditions (e.g. Walters et al., 2019; Arduini et al., 2019; Day et al., 2020; Wang et al., 2013; Saha et al., 2017; Decharme et al.,

2016; Dutra et al., 2012). For example, Arduini et al. (2019) have shown that the new multi-layer snow scheme, implemented at ECMWF, improved the description of snow density in thick and cold snow packs and also improved the representation of sporadic melting episodes due to the introduction of a thin top snow layer with low thermal inertia. Moreover, snow depth improved at all lead times in coupled forecasts, also generally improving the 2m temperature forecasts by refining the diurnal cycle amplitude in snow covered regions, especially in clear sky conditions.

The more realistic snowpack representation, with better quality snow depth and snowmelt, will naturally be expected to contribute to an improved hydrological cycle and thus better river flow simulation. Although there have been studies evaluating the hydrological impact of the more complex multi-layer snow schemes, these are mostly evaluations in localised settings (i.e. using study sites such as Col de Porte in France) and often combined with other aspects of the land-surface modelling (e.g. snow and other surface variables) (Dutra et al., 2012; Magnusson et al., 2015; Wang et al., 2013). However, a hydrological impact analysis with river flow as the main focus, has not been done at regional or global scales.

Evaluating modelling improvements using river discharge can be advantageous for different reasons. Importantly, river discharge is a unique variable as it aggregates excess water originating upstream, resulting in an observation that is representative of various processes not only at the measurement location, but of a much larger area (Fekete et al., 2012). Moreover, river discharge is an accurately measured element of the water and energy cycle, with global in situ monitoring in place (Depetris, 2021), even if certain gaps exist (Lavers et al., 2019; Rodda et al., 1993).

The evaluation of snow scheme changes for hydrological process representation, with river discharge as a diagnostic tool, promises to help identifying existing limitations for most parts of Northern Extratropics. Moreover, a relatively large fraction of these areas is permafrost, where the ground remains below 0 °C for two or more consecutive years (Romanovsky et al., 2002; Zhang et al., 2008). In permafrost, the soil temperature conditions (through the frozen soil) play a key role. Soil freezing is important not only for hydrology, by preventing vertical water flow, but also for the change in the carbon cycle with a powerful greenhouse gas feedback as a consequence of the expected climate change and the thawing of these areas (Koven et al., 2013; Andresen et al., 2020; Yokohata et al., 2020).

As snow has a strong control on the surface energy budget, by insulating the soil (Gouttevin et al., 2012), understanding the hydrological impact of the snow scheme complexity, such as on runoff, infiltration and soil temperature, is of crucial importance especially in relation to permafrost. Undertaking this in the global context, using river discharge as a powerful diagnostic tool, would

provide a novel approach, with a broad viewpoint on the hydrological processes of different areas of the globe.

The hydrological impact of the snow scheme complexity is explored in Chapter 5 for reanalysis simulations, which allows the use of larger number of catchments with available river discharge observations over a longer period in the past (Lavers et al., 2019). This is done by comparing the new multi-layer snow scheme, implemented at ECMWF (Arduini et al., 2019), with the current operationally used single-layer snow scheme (Dutra et al., 2010), using the ECLand land surface modelling platform and ERA5 meteorological forcing. Reanalysis simulations of river discharge, produced with using either the single-layer or the multi-layer snow models, are compared globally over the snow impacted areas. Moreover, the hydrological sensitivity is also investigated with special emphasis on permafrost, by including some variations in the snow and soil freezing parametrisations. The comparison of these simulations, after verifying them with river discharge observations, will highlight areas of the globe, where the land surface processes and variables in the Earth system model could benefit most from the introduction of a more complex snow scheme.

2.3 Trends in hydrological reanalysis data sets

Global climate reanalysis data sets, such as ECMWF's latest 4th generation ERA5 reanalysis (produced by the Copernicus Climate Change Service of the European Commission (C3S)), or NASA's MERRA-2 and JMA's (Japan Meteorological Agency) JRA-55 as 3rd generation data sets (Keller and Wahl 2020), are important products of Earth system models. These data sets combine observations from the past with short-range forecasts using the data assimilation process to generate consistent time series over several decades at least, for multiple climate variables of the Earth system, such as air temperature, pressure or surface parameters like precipitation and soil moisture, or other variables like sea temperature or river discharge (Keller and Wahl, 2020; Mahto and Mishra, 2019). The main advantage of reanalysis data is the continuity in space and time, since they are produced at global scale and do not have geographical gaps or missing periods, and as such are ideal to study the Earth's hydroclimatic system (Hodges et al., 2011).

Global reanalysis data sets are widely used for hydrological and climate applications. Amongst other uses, they are the basis for generating model climatologies, which contain reference information on the local climate characteristics (e.g. Alfieri et al., 2020; Harrigan et al., 2020b; Arnone et al., 2020). For example, as described in Harrigan et al. (2020b), the hydrological reanalysis is used for two main tasks in GloFAS. Firstly, to produce the flood thresholds at 2-, 5-, and 20-year return periods, which are used to determine the severity of the ensemble river discharge forecasts, by comparing them to these predefined flood thresholds. Secondly, it provides the basis to derive

initial hydrometeorological conditions (i.e. soil moisture, groundwater, snow cover and water in the rivers) for the system, which are also the basis of evaluating the quality of the hydrological simulations, used as best estimate of historical hydrological conditions.

As the forecasts are compared to the thresholds to derive severity (and warnings), the consistency between the reanalysis time series, used to derive the thresholds, and forecasts is crucially important. The frequency as they represent different severity of hydrological events (e.g floods or droughts) should ideally be the same (Alfieri et al. 2019).

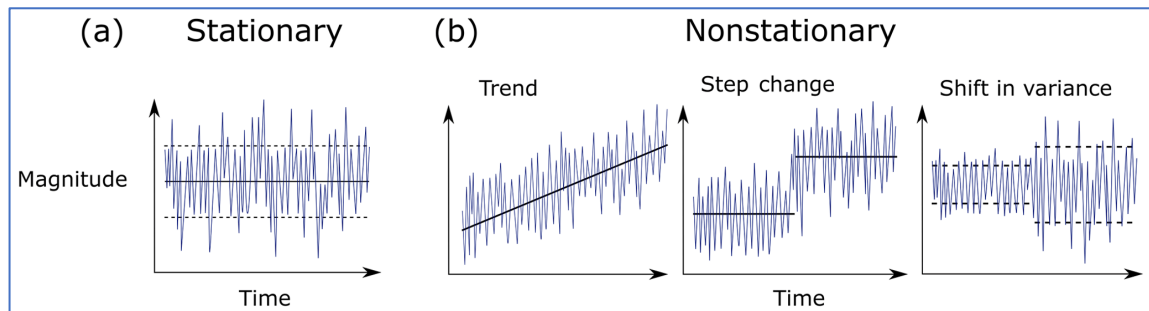


Figure 2-4. What is nonstationary? Examples of (a) a stationary time series with constant mean and variance and (b) three nonstationary time series in the form of a shift in mean (trend and step change) and a shift in variance. Solid and dashed black lines represent the mean and the variance of the time series, respectively. From Slater et al. (2021).

The forecasts are expected to be essentially from the same climatological period as the end of the reanalysis. Therefore, inconsistencies between reanalysis and the forecasts can arise from nonstationary behaviour in the long reanalysis time series of usually 30 years or more (WMO, 2017). As described by Slater et al. (2021), stationary time series would not exhibit any shift in the mean, variance or shape, while nonstationary examples are continuous trends, step changes or shifts in the variance (Figure 2-4).

For example, in the case of a major negative trend in the river discharge reanalysis time series, the flood thresholds will represent mainly the earlier period, where the magnitude of the floods will be much larger (due to the negative trend). However, the forecasts will be expected to behave similarly to the latter period of the reanalysis, where the flood events have lower magnitude (due to the negative trend), making the threshold non-representative of the forecast behaviour and ultimately decrease the reliability of the warnings.

There is an extensive literature of trend analysis for different variables of the hydrological cycle in regional or global settings. Example studies include the use of in situ or gridded satellite estimate observations (or a combination of those) and/or reanalysis data sets for river discharge (Do et al., 2017; Su et al., 2018; Dai et al., 2009; Archfield et al., 2016; Winkelbauer et al., 2022; Feng et al., 2021) or the contributing land-surface variables such as precipitation (Sun et al., 2020; Westra et

al., 2013; Donat et al., 2016; Nguyen et al., 2018), soil moisture (Pan et al., 2019; Feng and Zhang, 2016; Dorigo et al., 2012; Albergel et al., 2013), evaporation (Zhang et al., 2016; Anabalón and Sharma, 2017), snow conditions (Kunkel et al., 2016; Connolly et al., 2019) or temperature (Hansen et al., 2006; Hawkins et al., 2017).

The above-mentioned studies focus on a single variable and examine either global or in some cases regional trends (i.e. as in Winkelbauer et al., 2022 and Feng et al., 2021), while other studies consider different land surface variables at the same time. For example, Folton et al. (2019) looked at trends in both precipitation and river discharge in the Mediterranean area, Andreadis and Lettenmaier (2006b) analysed trends in soil moisture, runoff and drought over the US, Spinoni et al. (2017) and Javadian et al. (2020) evaluated trends in precipitation and evaporation over Europe and globally, Rosmann et al. (2016) analysed trends in mean and extreme temperature, precipitation and river discharge globally and Knowles (2015) looked at trends in snow depth, precipitation, snowfall and temperature data in the US.

However, comprehensive diagnostic studies with global focus, which analyse river discharge trends (both observed and reanalysed) together with the contributing trends from the different components of the land-surface system, such as precipitation, evaporation, snowmelt or soil moisture, have not been previously undertaken.

A potential candidate data set for such study is the river discharge reanalysis data set of GloFAS-ERA5 of CEMS (Harrigan et al., 2020b), which uses the ECLand land surface modelling platform forced by the latest ERA5 climate reanalysis of ECMWF. It is one of the most widely used river discharge reanalysis data sets, and by using ERA5, it provides a consistent representation across all land-surface variables, including not just river discharge but precipitation, soil moisture, evaporation, etc. Moreover, as ERA5 is also available in near real time (with usually only about 1-2-day delay), it allows GloFAS-ERA5 to be used in the forecast initialisation, which helps in achieving better consistency for extreme event representation between the GloFAS-ERA5-based flood thresholds and the forecasts.

Global climate reanalysis data sets, such as ERA5, are not immune from the impact of the changes that occur in their observing system over time. Their long-term temporal consistency is potentially impacted by the increase of observation input, that might enhance the quality of a reanalysis, and by the evolving observational system (e.g. the introduction of the Advanced Microwave Sounding Unit (AMSU-A) in 1998), that could also create potential inhomogeneities if either the observations or the model is biased (Thorne and Vose, 2010; Ferguson and Villarini, 2012).

The trend analysis of ERA5 and the related GloFAS-ERA5 river discharge data set would provide invaluable information on the nature of inhomogeneities and changes in the time series of land-surface variables in a state-of-the-art Earth system model, highlighting areas for improvements in future climate reanalyses. Moreover, this analysis could also help diagnosing the problematic geographical areas, where the GloFAS-ERA5 river discharge time series is not stationary and therefore the quality of the global hydrological predictions of GloFAS can potentially be impacted by the inconsistencies in flood event representation between climatology and forecasts.

A possible way to test the non-stationarity in the land-surface system of ERA5 with GloFAS-ERA5, could be to concentrate on the linear trends in annual maxima time series, which is explored in Chapter 6. This directly corresponds to the way as the river discharge return period flood thresholds are computed (using the annual maxima; Alfieri et al., 2019). The linear trend analysis is done using the global catchments, after transferring all land-surface variables into catchment-representing values. This allows a direct comparison of the land-surface trends with the simulated and observed river discharge trends and provides an ideal ground to disentangle the different contributions coming from the various land-surface processes.

2.4 Innovative ways of generating global flood thresholds

Hydrological forecasting systems, on regional or global scales, generally rely on reanalysis model simulations to generate flood thresholds and with this to associate a severity information to the predicted flood events and ultimately to help in delivering flood warning information. Examples include the European EFAS (European Flood Awareness system; Thielen et al., 2009) and E-HYPE (Donnelly et al., 2016) and the global GloFAS (Harrigan et al., 2020b; see Figure 2-5 as an example of the river discharge forecast and thresholds presented in GloFAS) and World-Wide HYPE (WWH; Arheimer et al., 2020), which all rely on reanalysis information to generate climatology for the whole river network.

The reasons why model-simulation-based thresholds are used are multifold. For example, Thielen et al. (2009) highlight some of the limitations associated with observations that make them less suitable to derive thresholds for operational flood forecasting systems. They mention lakes and reservoirs, which are not represented in the same way in observations and the model: this can create inconsistencies when the observation-based thresholds are compared with the model-simulated river discharge and e.g. the model does not have the influence of a reservoir while the observations have. In addition, there are differences between the behaviour of simulations and observations, i.e. the existence of model biases, which they showed to be significant even on the smaller European scale. This again can cause potential inconsistencies in the intended and

perceived interpretation of flood warnings, as e.g. the thresholds could be too high and not-representative of the model simulation behaviour. And finally, the fact that river discharge observation time series are only available at selected locations, which will only represent those river sections, leaving larger gaps on the river network especially on the global scale (Lavers et al., 2019; Rodda et al., 1993).

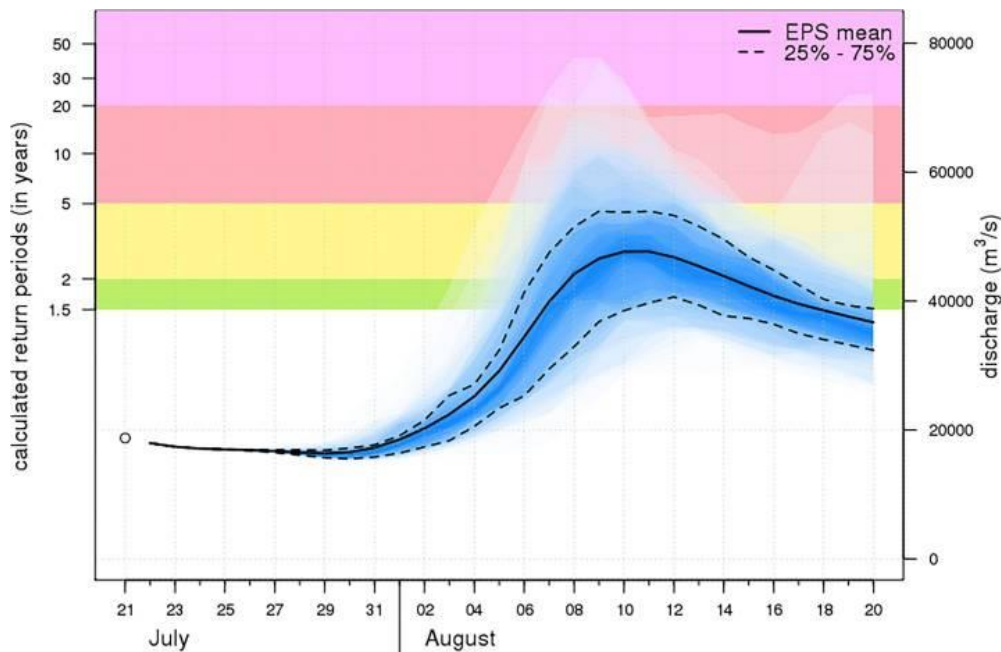


Figure 2-5. 30-day GloFAS ensemble forecast issued on 21 July 2018 for the Ganges river at Hardinge Bridge (Bangladesh) with the flood warning thresholds (coloured background areas) for return periods of 1.5-, 2-, 5- and 20-year (source: www.globalfloods.eu). Taken from Alfieri et al. (2019).

Weeink (2010) argued that the choice of thresholds is an essential element of any hydrological forecasting system. The strengths and limitations of the underlying modelling system has to be taken into consideration when designing the flood thresholds for the system. Having optimally behaving thresholds should provide the most reliable flood warnings with the highest rate of flood event detection and minimum number of false alerts. Hirpa et al. (2016) have also shown that the choice of data set to produce the reference climatology and thresholds has a major impact on the threshold magnitudes and the quality of the flood warnings in their evaluated GloFAS system.

The use of climatological simulations to generate flood thresholds will help in accounting for various biases in the forecasts, as these errors are expected to feature in a generally similar way in both the climatology (i.e. the thresholds) and the actual forecasts. This way, systematic errors (e.g. a generic positive bias) can largely be removed by transforming the flood signals into the evolution of information on threshold exceedance (Harrigan et al., 2020b; Alfieri et al., 2013). This is especially important on the global scale, where the availability of quality observations can be very limited and

thus the biases in global systems can be quite substantial, as demonstrated by e.g. Harrigan et al. (2020b) for GloFAS.

Biases in the forecasts can be nonstationary and will likely change with the increasing forecast lead time. As shown by Alfieri et al. (2019), this can make the use of the same nonvarying threshold (i.e. based on the ERA5 reanalysis) more and more suboptimal as it is used for increasing lead times out to several weeks.

The use of range-dependent thresholds, which vary across forecast lead times, promise to provide more optimal flood warnings. However, in order to generate them, sufficient number of forecasts are needed over a long enough period. The use of historical forecasts (produced in real time in the past) is an option, but these are created with many different model versions over a longer period, therefore will never be optimal. Reforecasts can offer a solution, which are forecasts re-generated for past dates. They are a better alternative over historical forecasts, as are produced using the same model version, consistently for the whole period.

In addition, many of the hydrological forecasting systems in use today have evolved into probabilistic systems and are based on ensemble forecast information to drive their predictions (Cloke and Pappanberger, 2009; Demeritt et al., 2010). These ensemble forecasts include a number of ensemble members by perturbing the initial conditions for the model runs, to reflect the full range of possible future outcomes (Palmer, 2019). These probabilistic systems can also include reforecasts, the so-called ensemble reforecasts, which are the probabilistic equivalent of the single-member reforecasts, containing a number of ensemble members (Harrigan et al., 2023). They are based on meteorological reforecasts produced by global NWP centres, most notably by ECMWF, who generate them dynamically in real time twice a week for the past 20 years (Vitart, 2014). As an example, Figure 2-6 highlights the schematic of the reforecasts taken from a whole calendar year.

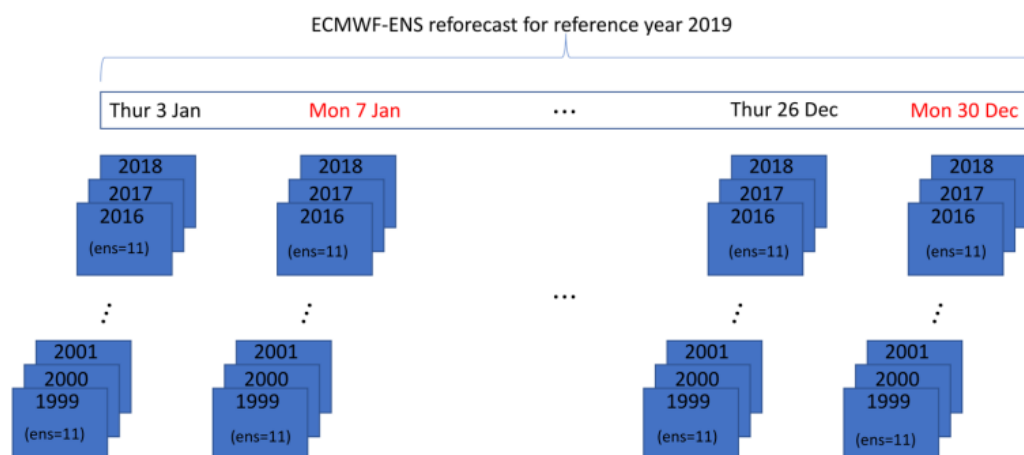


Figure 2-6. ECMWF reforecast schematic for the reference period of January to December 2019, including 104 run dates (all Monday or Thursday in 2019). Taken from Harrigan et al. (2023).

Despite the recent advancement of the ensemble-based forecast systems, ensemble information has not been yet considered anywhere in flood threshold generation. However, as ensembles can show different biases to single-member (deterministic) forecasts (e.g. Leutbecher et al., 2017), this suggests that using ensemble forecast information in generating range-dependent thresholds could be beneficial.

The use of reforecasts, and more specifically ensemble reforecasts, in generating the model climatologies have some examples in the scientific literature. For hydrology, Alfieri et al. (2019) showed that range-dependent, reforecast-based flood thresholds were substantially different from unique reanalysis-based thresholds in two thirds of the global rivers. This study, however, only considered the control member of the ensemble reforecasts (i.e. the member that runs from the lower resolution analysis without perturbation). Emerton et al. (2018), on the other hand, used hydrological ensemble reforecasts on the seasonal scale to generate the low- and high-flow thresholds for the seasonal hydrological predictions, specific for each weekly lead time. Other examples exist in meteorology, which use ensemble reforecasts to provide range-dependent climatologies in identifying extreme events in NWP meteorological forecasts (e.g. Tsonevsky et al., 2018).

The use of ensemble reforecast data promises to deliver not only range-dependent flood thresholds, but also thresholds that benefit from the representation of ensemble behaviour. These ensemble-forecast-based thresholds will be expected to help overcoming the problem of extreme event representation inconsistencies due to nonstationary forecast biases and will ultimately be expected to lead to higher forecast reliability and better skill in predicting global flood events.

The potential benefits of using ensemble reforecasts to define flood thresholds is analysed globally in the context of the GloFAS system in Chapter 7, using the river discharge reforecasts available for a calendar year period from the Copernicus Climate Data Store (CDS). The impact of the data source choice (ensemble reforecasts vs. ERA5 reanalysis) and the sampling strategies to calculate the annual maximum values (from the reforecasts, i.e. choosing one member randomly) for the return period threshold computation (as described e.g. in Harrigan et al. (2020b) for GloFAS) is evaluated. The flood threshold magnitudes are compared and the resulting forecast reliability and skill benefits (after applying the respective thresholds to compute exceedance probabilities) explored over thousands of global river catchments. This way, the benefit of replacing the non-varying reanalysis-based thresholds with range-dependent and ensemble-based thresholds can directly be quantified.

Chapter 3 Methods and data sets

This thesis explores several global hydrological data sets, produced by different modelling systems. In this section the models, the data sets and the methods used are described.

3.1 Hydrological and land-surface models

The hydrological simulations produced and explored in this PhD relied on the following land-surface and hydrological models, listed below.

3.1.1 *ECLand*

ECLand is the land surface modelling platform (Boussetta et al., 2021) that includes the HTESSEL (The Hydrology-Tiled ECMWF Scheme for Surface Exchange over Land; Balsamo et al., 2009; Balsamo et al. 2011) land surface scheme, the offline surface modelling driver, the river discharge routing scheme of CaMa-Flood (Yamazaki et al., 2011), the urban tile scheme (McNorton et al., 2021) and the representation of the carbon cycle (Boussetta et al., 2013). *ECLand* is expected to become open source in 2023 (Sleigh et al., 2022). It represents the vertical transfer of energy, water and carbon, and where appropriate, corresponding sub-surface quantities.

Each grid-box at the interface between the surface and the atmosphere is fractioned into tiles. There are currently seven tiles over land (high and low vegetation, bare ground, shaded and exposed snow, intercepted water and lake) and up to two tiles over sea and freshwater bodies (open and frozen water). Each tile is characterised with its own properties identifying separate water and heat fluxes and allow solving an energy balance equation for the tiles skin temperatures and taking into account physical processes that limit evaporation of vegetated areas and bareground.

In *ECLand*, the soil is divided into four layers with fixed layer depths, while runoff is generated as fast (surface) and slow (subsurface) components at each grid point (Balsamo et al., 2009). As there is no horizontal flow component between grid points in *ECLand*, it needs to be coupled to a routing model with a river network to produce river discharge.

The snowpack is characterised by considering snow density, thermal insulation properties interception of rain and the evolution of the albedo due to metamorphism aging processes. Further description of the land surface parametrization is available in Boussetta et al. (2021). *ECLand* is part of the Integrated Forecasting System (IFS) at ECMWF and used in coupled land-atmosphere simulations at various spatial resolutions, from short- to seasonal range, including the production of the ECMWF reanalysis data sets.

3.1.2 *Offline simulation methodology*

ECLand can be used in a stand-alone mode, when the model runs uncoupled from the atmosphere, forced with near-surface meteorological input data of temperature, specific humidity, wind speed, surface pressure, radiative fluxes (downward solar and thermal radiation), and water fluxes (liquid and solid precipitation), without land data assimilation.

This offline research methodology provides an affordable way of testing land surface improvements and has been used in various applications (e.g. Agustí-Panareda et al., 2010; Haddeland et al., 2011; Zsoter et al., 2019; Arduini et al., 2019). The ERA5-Land dataset is a prime example of this methodology, which was produced as an offline ECLand simulation (Muñoz-Sabater et al., 2021).

3.1.3 *Lisflood*

Lisflood is an open source (<https://ec-jrc.github.io/lisflood-model/>) distributed semi-physically based model, primarily developed to simulate major hydrological processes in large catchments at the Joint Research Centre (JRC) of the European Commission (van der Knijff et al., 2010). Processes simulated by Lisflood include soil freezing, snowmelt, surface runoff, lakes and reservoirs, water abstraction, infiltration, preferential flow, redistribution of soil moisture within the soil profile, drainage to the groundwater system, groundwater storage, and base flow.

Surface runoff is produced at every grid cell and then first routed to the nearest downstream river channel cell, then the water in the channel is routed through the river network using the kinematic wave approach (Chow et al., 1988). Lisflood also simulates the groundwater storage, groundwater flow, and flow routing into and through river channels. Groundwater storage and transport are represented using two interconnected groundwater zones each consisting of a linear reservoir (Burek et al., 2013), which subsequently transport water to the river channel with a time delay. The current Lisflood version includes 463 lakes and 687 reservoirs (Zajac et al., 2017), selected from the world's largest lakes and reservoirs.

3.1.4 *Lisflood-routing*

The simplified version of the Lisflood hydrological model is used specifically for flow routing, simulation of groundwater processes, human water use and lakes and reservoirs only. This version of Lisflood (called Lisflood-routing hereafter) takes surface and subsurface runoff as input variables. The sub-surface runoff is used as input for the groundwater module, while surface runoff is used as input for the river channel routing module. It includes the same lakes and reservoirs as Lisflood (463 lakes and 687 reservoirs).

3.1.5 *CaMa-Flood*

CaMa-Flood (Yamazaki et al., 2011) is a global river-routing model, which is part of ECLand since IFS cycle 47r1 (Boussetta et al., 2021). The model can be used to simulate the hydrodynamics and produce river discharge from the ECLand runoff outputs. CaMa-Flood routes runoff generated by land-surface models to oceans or inland seas. The model calculates river and floodplain water storages, discharge, water depth, as well as flood inundation. The CaMa-Flood version (v3) described and used in this PhD (and also earlier versions), does not include the representation of dams and permanent lakes and wetlands are only treated as part of the floodplain storages. CaMa-Flood is computationally cheap to run, even though it includes explicit representation of flood inundation. It has been used widely in global climatological research studies, such as Emerton et al. (2017), Dottori et al. (2018) and Zsoter et al. (2019).

3.2 Meteorological forcing data sets

In order to produce the hydrological simulations that were used to generate the climatologies in the studies of this PhD, different reanalysis meteorological forcings were applied. In addition, meteorological data was also used to force the forecast simulations for the medium-and seasonal ranges, which were analysed in different scientific studies, produced during the PhD. These meteorological data sets are all generated by ECWMF, listed below.

3.2.1 *ERA-Interim climate reanalysis*

ERA-Interim (ERA-I hereafter) is ECMWF's global atmospheric reanalysis from 1979 to present produced with an older (2006) version of the ECMWF Integrated Forecasting System, cycle 31r2 (Dee et al., 2011). The system includes four-dimensional variational data assimilation (4D-Var) with a 12-hour analysis window. The spatial resolution of the data set is approximately 80 km (T255 spectral) on 60 levels in the vertical from the surface up to 0.1 hPa. ERA-I is available from 1 January 1979 to 31 August 2019 and was superseded by the ERA5 reanalysis.

3.2.2 *ERA-Interim/Land climate reanalysis*

ERA-Interim/Land (Balsamo et al., 2015; hereafter ERA-I-Land) is a global reanalysis of land-surface parameters at 79 km spatial resolution, covering originally the period 1979-2010 which was later extended for subsequent years until 2018. It was produced with a single 'offline' simulation, uncoupled from the atmosphere, not including data assimilation and forced by atmospheric forcing from ERA-I, using a more recent version of the HTESSEL land-surface model (cycle 37r2 for the original ERA-I-Land data set). It also includes adjusted precipitation forcing, based on the Global Precipitation Climate Project (GPCP; Huffman et al., 2009), where the ERA-I 3-hourly precipitation values are rescaled to match the monthly accumulated precipitation provided by the GPCP v2.1

product (for more details please consult Balsamo et al., 2010). ERAI-Land preserves closure of the water balance and includes a number of parameterisations improvements in the land surface scheme with respect to the original ERAI dataset, which makes it more suitable for climate applications than ERAI, where the land data assimilation can add or remove water (Zsoter et al., 2019).

3.2.3 ERA-20CM climate reanalysis

ERA-20CM is an atmospheric model reconstruction of the 20th century, produced by ECMWF (Herschbach et al., 2015). It includes a 10-member ensemble estimate of the climate conditions over the period of 1899–2010 on horizontal resolution of about 125 km, produced by the IFS cycle 37r3.

3.2.4 ERA5 climate reanalysis

ERA5 is the latest global climate reanalysis of ECMWF (Herschbach et al., 2020). It is a key contribution to the C3S and is open access and free to download for all uses (in the CDS; <https://cds.climate.copernicus.eu/>). ERA5 covers the period 1950 to present, with quality assured data available from 1959. ERA5, similarly to ERAI, is a land-atmosphere-ocean coupled application which includes the assimilation of conventional in-situ and satellite observations for the analysis of soil moisture, soil temperature and snow fields (de Rosnay et al., 2014). ERA5 is based on the IFS cycle 41r2, which was operational at ECMWF in 2016. ERA5 thus benefits from 10 years of modelling improvements (physics, dynamics and data assimilation) relative to ERAI. In addition to a significantly increased horizontal resolution (31 km compared with 79 km for ERAI), ERA5 has also include hourly outputs and an uncertainty estimate based on a 10-member ensemble of data assimilations with 3-hourly output and 63 km resolution. ERA5 also provides a number of output parameters, e.g. including a 100 m wind product. The move from ERAI to ERA5 represents a step change in overall quality and level of detail.

3.2.5 ERA5-Land climate reanalysis

ERA5-Land is an enhanced global dataset for the land component of ERA5 within the C3S. It describes the evolution of the water and energy cycles over land in a consistent manner over the production period, which, among others, could be used to analyse trends and anomalies. ERA5-Land is a product of an offline ECLand simulation (with IFS cycle 45r1), forced by ERA5 atmospheric variables (e.g. air temperature or radiation). ERA5-Land is produced at 9 km spatial resolution using downscaled ERA5 atmospheric forcing and a vertical lapse rate correction. ERA5-Land shares with ERA5 most of the parameterizations that guarantees the use of the state-of-the-art land surface modelling applied to NWP models. There is no direct coupling or land data assimilation in ERA5-Land (there is only an indirect impact through the ERA5 forcing), which can have a large impact on

the hydrological cycle (Zsoter et al., 2019). The much higher resolution and the temperature lapse-rate correction in ERA5-Land are major differences to ERA5, which can further contribute to the impacts on the water budget, especially in mountainous areas, changing the snowpack and snowmelt through the temperature differences.

3.2.6 *ENS forecasts*

The ECMWF ensemble forecasts (ENS) includes 51 ensemble members operationally out to a lead time of 15 days, produced twice per day at 00:00 and 12:00 UTC. There is a single ‘control’ member which is generated from the most accurate estimate of current conditions and a remaining 50 members which have their initial conditions perturbed to provide a range of possible future weather scenarios. The ENS forecasts are extended to 46 days twice per week on Mondays and Thursdays. These extended-range forecasts are run at coarser horizontal resolution at ~36 km.

3.2.7 *SEAS5 forecasts*

SEAS5 is the 5th version of ECMWF’s long-range ensemble forecasting system made operational (using IFS cycle 43r1) in November 2017 (Johnson et al., 2018). Similarly to ENS, SEAS5 consists of 51 ensemble members, and has slightly lower horizontal resolution of ~36 km. The system produces forecasts each month with 7 months lead time. Seasonal forecasts provide predictions of how the average atmospheric, ocean and land surface conditions are likely to differ from the long-term average. The seasonal predictions have generally lower skill than the medium- or extended-range forecasts, but for some sectors such as agriculture or water management they can be useful in predicting potentially unusual conditions months in advance.

3.2.8 *ENS reforecasts*

The ECMWF ENS reforecasts are forecast integrations with 11 ensemble members and 46-day lead time, using the same IFS version and starting on the same day and month as the real time forecasts for each of the past 20 years (Vitart, 2014). The reforecasts are currently produced for Mondays and Thursdays (so twice weekly). For instance, if the starting date of the real-time forecast is 21 March 2022, then the corresponding reforecasts with 11-member ensemble will be produced starting on 21 March 2021, 21 March 2020, ..., 21 March 2002. The reforecasts are thus produced with 20 different starting dates from the past 20 years.

3.2.9 *SEAS5 reforecasts*

The SEAS5 reforecasts (also sometimes known as hindcasts or back integrations) start on the 1st of every month for the years 1981-2016, go out to 7 months and have 25 ensemble members. From 2017 the seasonal reforecasts include the same 51 ensemble members as in the real time forecasts.

The reforecasts are produced with the same model as the real time seasonal forecasts. They are available to users of the real-time forecast data to calibrate the forecast products.

3.3 Hydrological modelling systems

During the course of this PhD, three different hydrological modelling systems were used. The first two rely on the ECLand land-surface model as the core hydrological engine to produce runoff, either as a coupled system with land data assimilation or uncoupled from the atmosphere without assimilation, and the river routing component as either Lisflood-routing or CaMa-Flood.

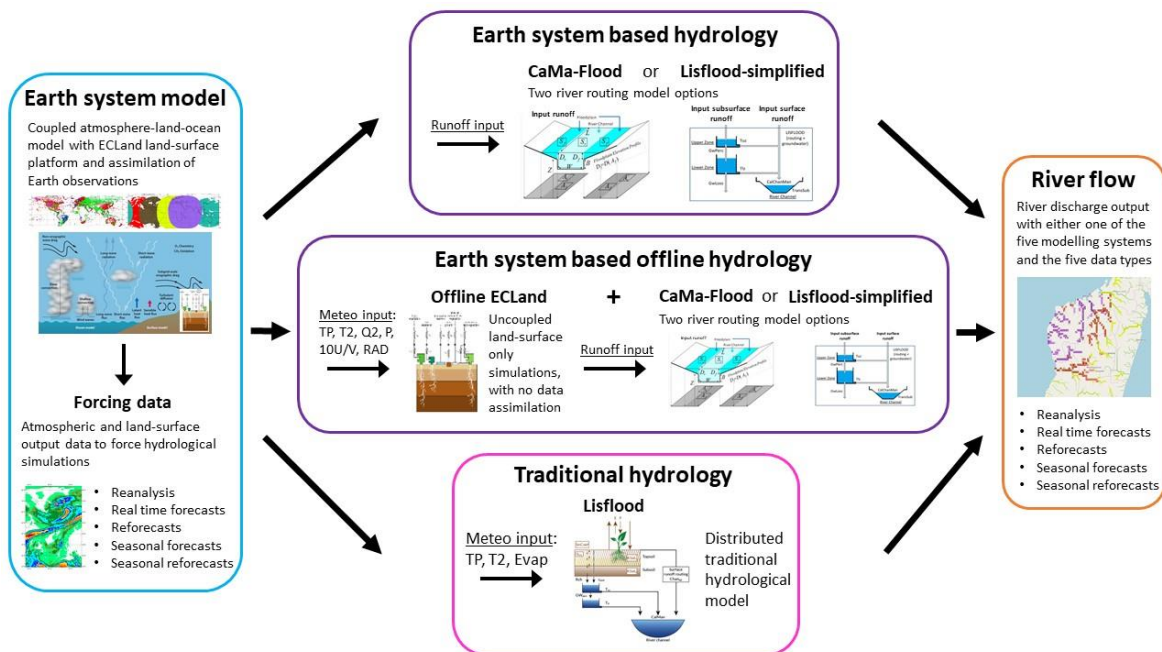


Figure 3-1. Schematic of the hydrological modelling systems used in the PhD with the related data flow.

In addition, the third system is fully based on Lisflood to produce river discharge. Figure 3-1 highlights the main components of the modelling systems with the applied forcing data and the relationships between them, while Figure 3-2 lists the major model versions with some of the main features and the timeline.

3.3.1 Global Flood Awareness System (GloFAS)

Many data sets used were produced in the Global Flood Awareness System, which is part of the Copernicus Emergency Management Service (CEMS), developed by JRC and ECMWF. GloFAS is one of the few global scale operational flood forecasting systems that exist (Harrigan et al., 2020b). GloFAS couples global NWP input data with hydrological modelling to produce river discharge on a global river network, currently at 0.1 degree resolution (van der Knijff et al., 2010; Alfieri et al., 2013).

GloFAS has a 30-day component with daily updates and a seasonal component with 4 months lead time and monthly updates. The 30-day component uses meteorological forcing from the medium- and extended-range ENS, while the seasonal system uses SEAS5, the latest version of ECMWF's long-range ensemble forecasting system. The GloFAS system also includes reforecasts for both the 30-day and seasonal components, which are hydrological forecasts produced for past dates to help with forecast evaluation and post-processing. These reforecasts use the ENS and SEAS5 reforecasts as forcing.

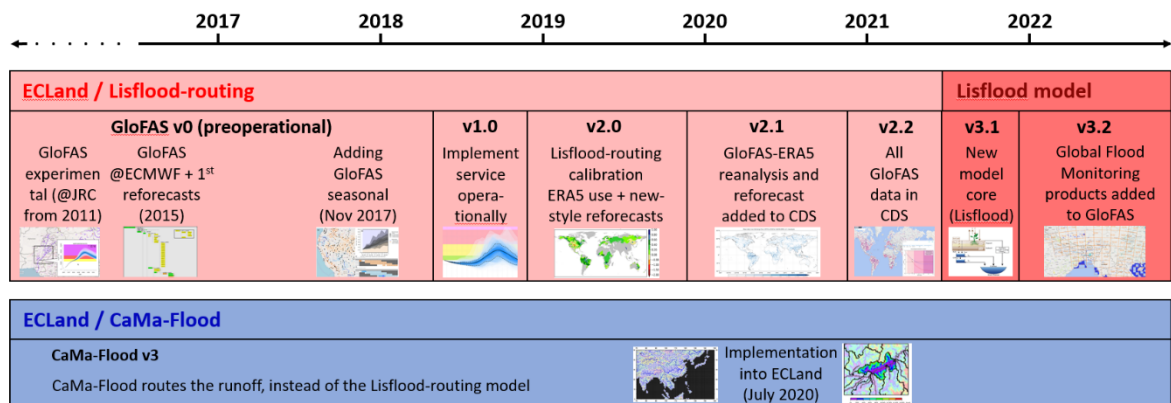


Figure 3-2. Time evolution of the hydrological modelling systems and the related versions used in the PhD. The three colours represent the three modelling systems used.

As part of the current GloFAS configuration, the real-time river discharge forecasts (30-day and seasonal) are compared to the hydrological thresholds which are derived from the model climatology. In the 30-day component, the flood thresholds are based on the river discharge reanalysis, a long simulation usually from 1979 to reconstruct historical river conditions. Whereas in the seasonal component, the low flow and high flow thresholds are based on the climatology produced from the seasonal reforecasts. All GloFAS data, the reanalysis, the 30-day and seasonal real time forecasts and reforecasts are freely available for all use from the CDS.

The climatologies play a crucial role in GloFAS. The consistency of the climatology-based thresholds and the real time forecasts are particularly important, as without it the hydrological warnings for floods in the 30-day component or high/low flow conditions in the seasonal would be unreliable. GloFAS was operationally implemented in April 2018, while it has been preoperationally providing forecasts already from 2011.

Two hydrological modelling systems have been part of GloFAS in the most recent few years. One with coupling ECLand to Lisflood-routing and another with using Lisflood (see Figure 3-1 and Figure 3-2. Below, a short description of the operational and major preoperational versions of the GloFAS modelling and data service are provided.

3.3.1.1 GloFAS version 0

GloFAS v0 was running preoperationally from 2011 on 0.1-degree resolution. It used Lisflood-routing to route the surface and sub-surface runoff outputs of the land component (HTESSEL) of the Integrated Forecast System to produce daily river discharge simulations. Lisflood-routing used expert-defined parameters which are uniform across the entire geographical domain, regardless of land surface characteristics (i.e. no calibration was performed). For the 30-day GloFAS version, meteorological forcing was used only in the first 15 days of the 30-day forecast horizon and the model climatology was generated based on the ERAI-Land-forced reanalysis data set. GloFAS v0 was associated with a reforecast set, containing regenerated forecasts for every day from 2008 to 2015 (then gradually extended with real time produced forecasts), using the 15-day ENS runoff forcing routed by Lisflood-routing and initialised by GloFAS-ERAI-Land-v0. GloFAS v0 was first run at the JRC, then adapted to run in ECMWF's environment in 2015.

3.3.1.2 GloFAS version 1.0

The first operational version of GloFAS was introduced in April 2018. It uses the same modelling configuration as GloFAS v0, with the uncalibrated Lisflood-routing coupled to ECLand surface and subsurface runoff outputs, ERAI-Land-forced reanalysis and ENS meteorological forcing in the first 15-day period of the 30-day forecasts. The 30-day reforecasts did not change either from GloFAS v0. In addition, the first seasonal version of GloFAS was introduced in November 2017 (Emerton et al., 2018). It used the same Lisflood-routing model with the surface and subsurface runoff outputs of ECLand in the SEAS5 forcing. The GloFAS seasonal forecasts (both real time forecasts and reforecasts) are initialised from the ERA5-forced reanalysis, which was first produced for periods (separate ERA5-forced simulations) when ERA5 was available (1990-1992, 2000-2007 and 2010-2016) at the time of setting up the system. The seasonal reforecasts only have 25 ensemble members. The reforecasts are used to compute the low/high flow thresholds used in the seasonal system, using the available years.

3.3.1.3 GloFAS version 2.0

This version of GloFAS v2.0 was implemented in November 2018. It uses the Lisflood-routing model, which was enhanced by the first calibration of the hydrological routing scheme (Hirpa et al., 2018b). It continues to couple the ECLand surface and subsurface runoff outputs to Lisflood-routing to produce river discharge. GloFAS v2.0 introduced the use of the ERA5 reanalysis (pre-release version of ERA5) in the 30-day model climatology and the forecast initialisation for both 30-day and seasonal, and a different method to generate the 30-day GloFAS reforecast data set was based on the 20-year ENS reforecasts of ECMWF with the ensemble runoff routed by Lisflood-routing. The 30-day reforecasts cover the period 1997-2016 (extended with real-time produced forecasts in 2017-2018). At the same time, the seasonal reforecasts were extended to include all years in 1981-

2017, initialised from the ERA5-reanalysis, and to include 2018 using the real-time-produced seasonal forecasts.

3.3.1.4 GloFAS version 2.1

GloFAS v2.1 was implemented in November 2019, using the same calibrated Lisflood-routing version as in GloFAS v2.0. It continues to couple the ECLand surface and subsurface runoff outputs to Lisflood-routing to produce river discharge. It includes a revised hydrological reanalysis based on the officially released ERA5 meteorological forcing from 1979 to 2019. Due to the updated reanalysis, the climatologies with the 30-day flood thresholds and the seasonal flow thresholds were also recomputed. The reforecasts were also regenerated, using the new reanalysis data as initialisation, covering the period of 1998-2017 (extended with real-time-produced forecasts in 2018-2019) for the 30-day and 1981-2019 for the seasonal. The GloFAS v2.1 upgrade introduced the reanalysis and real time 30-day forecasts in the CDS.

3.3.1.5 GloFAS version 2.2

The GloFAS v2.2 upgrade was introduced in December 2020 and included mainly changes in the service elements of GloFAS. Amongst others, it introduced the GloFAS reforecasts (both 30-day and seasonal) and the real time seasonal forecasts to the CDS. In this upgrade, the 30-day reforecasts were regenerated for the 2019 reference year, covering the period 1999-2018, while the seasonal reforecasts only got further extended into 2020 with the real-time-produced forecasts.

3.3.1.6 GloFAS v3.1

GloFAS v3.1 was implemented in May 2021 with a major change of the modelling system and related updates of all the GloFAS products and datasets; it was complemented by several new or updated web products. The GloFAS v3.1 modelling system is entirely based on the Lisflood model, which replaces the earlier setup of coupling the runoff output from ECLand to Lisflood-routing. Lisflood was calibrated over 1226 river catchments with a total drainage area of 51 million km² globally (Alfieri et al., 2020). Reanalysis and reforecast data sets were produced with the new modelling system. The reforecasts covered the period of 1999-2018, while the seasonal reforecasts covered 1981-2021.

3.3.1.7 GloFAS version 3.2

GloFAS v3.2 was implemented in October 2021 as a minor change to provide access to the Global Flood Monitoring (GFM) beta version through the GloFAS web interface. All hydrological data sets remained the same as in v3.1.

3.3.2 *ECLand/CaMa-Flood system*

Many of the hydrological data sets used in this PhD were produced with the Lisflood model, either as Lisflood-routing with coupling to ECLand or the full hydrological configuration version of Lisflood,

described earlier. In addition, several research studies, carried out at ECMWF either as part of this PhD or outside of it, used the CaMa-Flood hydrodynamic model, as the river routing component coupled to ECLand. Two early examples of using CaMa-Flood, are the flood inundation modelling in Pappenberger et al. 2012 and the flood forecast uncertainty analysis in Zsoter et al. (2016), which both explored ECLand runoff outputs routed with CaMa-Flood.

The representation of rivers and their role on the ocean and the land water cycle is very relevant in Earth system models, which motivated the integration of CaMa-Flood into the ECLand platform. Since IFS cycle 47r1, CaMa-Flood has been part of the IFS at ECMWF, introduced in June 2020 (Boussetta et al., 2021).

3.4 Hydrological data sets

In the following, the data sets produced during the PhD, either within GloFAS or as a research data set with the ECLand/CaMa-Flood coupled modelling system, are described. Most of them are variations of reanalysis simulations to study the model climatologies. In addition, real time forecasts and reforecast data sets have also been generated, which are used for forecast studies, many appearing in the four main papers and the co-authored publications produced during this PhD. Table 3-1 provides the name, data type, short description and indicates in which of the PhD publications these data sets appeared and finally in which section of the thesis to find them. The data sets are grouped according to the modelling system versions. For this PhD, all three modelling systems (see Figure 3-1 and Figure 3-2) were run with 24-hour coupling and output frequency. Therefore, all meteorological forcing data sets (such as ERA5) were used in 24-hour chunks for the simulations, and all hydrological output data sets in Table 3-1, many of which are analysed in this PhD, have 24-hour (daily) time steps, always following the 00-00 UTC convention.

Table 3-1. List of data sets produced and analysed in the PhD.

Name	Type	Description	PhD publications	Chapter
GloFAS version 0				
GloFAS-ERA5-Land-v0	Global reanalysis	Hydrological reanalysis with ERA5-Land runoff forcing (on ~80 km) routed with the uncalibrated Lisflood-routing on 0.1-degree (~11 km) river network. The data set was first produced for 1980-2010 and later gradually extended until 2017.	Bischiotti et al., 2019	Ch. 8

GloFAS- ERA1-Land- RFC-v0	Global ensemble reforecasts	Reforecasts produced for every day from 2008 with 30-day lead time. Real time, 15-day ENS runoff forcing (with 51 ensemble members; 2008-2009: 50 km; 2010-2015: 32/64 km; 2016-2018: 18/36 km) was routed by the Lisflood-routing model on 0.1-degree (~11 km) river network. The data set was produced for 2008-2015 in 2015 and regularly extended until 2018 by adding the real time forecasts produced pre-operationally. The initialisation was provided by GloFAS-ERA1-Land-v0.	Bischiniotis et al., 2019	Ch. 8
GloFAS version 1.0				
GloFAS- ERA1-Land- v1.0	Global reanalysis	Hydrological reanalysis simulation with ERA1-Land runoff forcing (on ~80 km) routed through the uncalibrated Lisflood-routing, produced for 1980 to 2017 on 0.1-degree (~ 11 km) river network. This version of ERA1-Land did not include GPCP precipitation correction anymore.	Towner et al., 2019 and Ficchi et al., 2021	Ch. 8
GloFAS- ERA5-v1.0	Global reanalysis	Hydrological reanalysis simulation with ERA5 runoff forcing (on ~32 km) routed through the uncalibrated Lisflood-routing, produced for 1980 to 2016 on 0.1-degree (11 km) river network.	Towner et al., 2019 and Passerotti et al., 2020	Ch. 8
GloFAS- ERA1-Land- RFC-v1.0	Global ensemble reforecasts	The reforecast data set in v1.0 is identical to the v0 reforecast. The only difference is the extended period covered with daily real time forecasts added in 2018. The initialisation was provided by GloFAS-ERA1-Land-v1.0.		

GloFAS-ERA5-SRFC-v1.0	Global seasonal ensemble reforecasts	Seasonal reforecast simulations with SEAS5 runoff forcing (on ~36 km) routed through the uncalibrated Lisflood-routing, produced for each month in the years of 1990-1992, 2000-2007 and 2010-2016 with 25 ensemble members on 0.1-degree (~11 km) river network. The initialisation was provided by GloFAS-ERA5-v1.0.	Emerton et al., 2018	Ch. 8
GloFAS version 2.0				
GloFAS-ERA5-Land-v2.0	Global reanalysis	Hydrological reanalysis with ERA5-Land runoff forcing (on ~80 km) routed through the calibrated Lisflood-routing on 0.1-degree (~ 11 km) river network. The data set was produced for 1980-2017.	Towner et al., 2019	Ch. 8
GloFAS-ERA5-v2.0	Global reanalysis	Hydrological reanalysis with pre-release version of ERA5 runoff forcing (on ~31 km) routed through the calibrated Lisflood-routing on 0.1-degree (~ 11 km) river network. This data set was produced from 1979 to 2018.	Towner et al., 2019	Ch. 8
GloFAS-ERA5-Land-v2.0	Global reanalysis	Hydrological reanalysis with the early release part of ERA5-Land surface and subsurface runoff forcing (9 km resolution) routed through the calibrated Lisflood-routing, on 0.1-degree (~11 km) river network. This data set covers 2001-2018.	Hersbach et al., 2018; Muñoz-Sabater et al., 2021 and Zsoter et al., 2020a	Ch. 7-8
GloFAS-RECF-v2.0	Global reanalysis	Hydrological reanalysis with reforecast ensemble control runoff forcing (the first 3- and 4-day periods of the twice-weekly reforecasts on ~18 km) routed	Towner et al., 2019 and Hirpa et al., 2018b	Ch. 8

		through the calibrated Lisflood-routing on 0.1-degree (~ 11 km) river network. This data set was produced for 1997-2016.		
GloFAS-ERA5-RFC-v2.0	Global ensemble reforecasts	The first version of GloFAS reforecasts with the 20-year ENS meteorological reforecast forcing. It was produced on 0.1-degree river network with 11 ensemble members. All available ENS reforecasts were used from the calendar year of 2017, which provided twice-weekly coverage of 1997-2016. The remaining period until June 2018 was covered by reforecasts using forcing of the 51-member real time ENS. Twice a week runs until Jun 2018 and daily from July 2018. The initialisation was provided by GloFAS-ERA5-v2.0.	Bischirotis et al., 2020; Passerotti et al., 2020 and Zsoter et al., 2020a	Ch. 7-8
GloFAS-ERA5-SRFC-v2.0	Global seasonal ensemble reforecasts	Seasonal reforecasts using SEAS5 runoff forcing (on ~36 km) routed through the calibrated Lisflood-routing for each month in the years of 1990-1992, 2000-2007 and 2010-2018 on 0.1-degree (~11 km) river network. The seasonal forecasts had 25 ensemble members until 2016 and 51 members thereafter. The initialisation was provided by GloFAS-ERA5-v2.0.		
GloFAS version 2.1				
GloFAS-ERA5-v2.1	Global reanalysis	A hydrological reanalysis forced with the officially released ERA5 data covering 1979-2019 on 0.1-degree river network.	Harrigan et al., 2020b; Zsoter et al., 2020b; Titley	Ch. 6 and 8

			et al., 2021 and Winkelbauer et al., 2022	
GloFAS- ERA5-RFC- v2.1	Global ensemble reforecasts	These reforecasts used the 20-year ENS reforecasts from the calendar year of 2018, providing twice-weekly coverage of 1998-2017 with 51 ensemble members on 0.1-degree river network (Harrigan et al. 2023). The remaining period was covered by reforecasts forced with 51-member real time ENS, twice a week runs until Jun 2019 and daily from July 2019. The initialisation was provided by GloFAS-ERA5-v2.1.		
GloFAS- ERA5- SRFC-v2.1	Global seasonal ensemble reforecasts	Seasonal reforecasts forced with SEAS5 covering 1981-2019 with 25 ensemble members until 2016 and 51 members thereafter on 0.1-degree river network. The initialisation was provided by GloFAS-ERA5-v2.1.		
GloFAS version 3.1				
GloFAS- ERA5-v3.1	Global reanalysis	A hydrological reanalysis forced with ERA5 covering 1979-2021 on 0.1-degree river network	Winkelbauer et al., 2022; Alfieri et al., 2019 and Alfieri et al., 2020 (with slightly earlier model version).	Ch. 8

GloFAS-ERA5-RFC-v3.1	Global ensemble reforecasts	In this version of the reforecasts, the calendar year of 2019 was used (same as in v2.2), which provides coverage for 1999-2018 with 51 ensemble members on 0.1-degree river network. The initialisation was provided by GloFAS-ERA5-v3.1.	Alfieri et al., 2019	Ch. 8
GloFAS-ERA5-SRFC-v3.1	Global seasonal ensemble reforecasts	Seasonal reforecasts forced with SEAS5 covering 1981-2021 with 25 ensemble members until 2016 and 51 members thereafter on 0.1-degree river network. The initialisation was provided by GloFAS-ERA5-v3.1.		
ECLand/CaMa-Flood				
CAMA-ERA20CM-R	Global reanalysis	Hydrological reanalysis forced with ERA-20CM runoff (125 km resolution), using CaMa-Flood river routing (0.5-degree, 55 km river network) and covering the period 1901-2010.	Emerton et al., 2017	Ch. 8
CAMA-ERA1-Land	Global reanalysis	Hydrological reanalysis forced with ERA1-Land runoff (meteorological forcing on ~80 km, runoff produced by offline ECLand simulation on ~32 km), using CaMa-Flood river routing and covering 1979-2015. CaMa-Flood run on a ~25 km river network	Towner et al., 2019	Ch. 8
CAMA-ERA5	Global reanalysis	Hydrological reanalysis forced with ERA5 runoff, using CaMa-Flood river routing and covering 1979-2018. The ERA5 runoff has a ~31 km resolution, while CaMa-Flood run on a ~25 km river network.	Zsoter et al., 2019 and Zsoter et al., 2022	Ch. 4-5

CAMA-ERA5-off	Global reanalysis	Hydrological reanalysis forced with runoff simulated offline by ECLand forced with ERA5, without land-atmosphere coupling and land data assimilation, using CaMa-Flood river routing and covering 1979-2018. The runoff input was produced on a ~31 km resolution grid (the ERA5 grid), while CaMa-Flood run on a ~25 km river network.	Zsoter et al., 2019 and Zsoter et al., 2022	Ch. 4-5
CAMA-ERA5-off-snow	Global reanalysis	Hydrological reanalysis experiments forced with runoff simulated offline by ECLand forced with ERA5, without land-atmosphere coupling and land data assimilation, using CaMa-Flood river routing and covering 1979-2018. These experiments differ to CAMA-ERA5-off by having modifications in the ECLand snow soil freezing schemes (i.e multi-layer parametrisation and other changes for the permafrost).	Zsoter et al., 2022	Ch. 5

3.4.1 Data production

To generate the 23 data set types, listed in Table 3-1, during the course of the PhD, in total hundreds of experiments were run at ECMWF. Each of these has produced data from few tens of GB to few TB, depending on the type of simulation and the list of variables outputted. In total, tens of TB of simulation data was produced during the 6 years of the PhD.

3.4.1.1 IT infrastructure

The simulations to produce the data sets were run primarily on the ECMWF High Performance Computer Facility (HPCF; Figure 3-3), which has two identical Cray XC40 clusters, each self-sufficient with their own storage to provide resilience to system failures and flexibility in performing maintenance and upgrades.



Figure 3-3. Picture of the Cray supercomputer at ECMWF.

The two clusters provide a total of about 260000 CPU (Central Processing Unit) cores, about 0.9 petabytes of memory and 8499 teraflops peak performance. Each of the XC40 systems has about 10 petabytes of storage and offers more than 350 gigabytes per second of I/O bandwidth. The HPCF processes more than 200,000 jobs per day on each cluster on average.

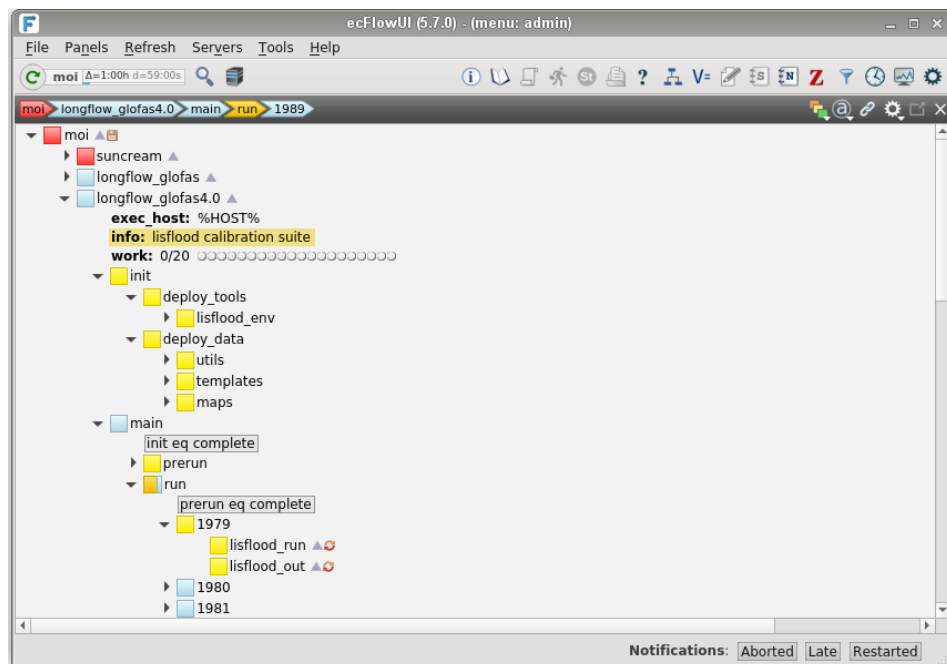


Figure 3-4. The ecFlow workflow manager software of ECMWF.

The majority of the experiments were run using ECMWF's in-house open-source workflow manager software, ecFlow (Figure 3-4). ecFlow enables users to run a long list of tasks, with dependencies on each other and on time, in a controlled environment. It provides tolerance, for hardware and software failures, combined with good restart capabilities. It is used at ECMWF to run all the operational and research suites (and produced the experiments) across a range of platforms. In

total, ECMWF run over 750,000 jobs per day on the supercomputer and other systems in a controlled manner.

3.4.1.2 Data archival

There is a large-scale data handling system (DHS) in place at ECMWF that help users to store and manage data related to weather and climate simulations, either for any required input data or for any model outputs. ECMWF's data archival currently contains hundreds of petabytes of operational and research data. The data is stored on a hierarchy of disks and tapes to store the files and their associated metadata. There are two data archival applications through which users can access the data:

- MARS (Meteorological Archival and Retrieval System) provides access to a powerful abstraction engine that allows users and applications to access data sets collected or generated at ECMWF for more than 30 years. MARS stores files in GRIB and BUFR formats only.
- ECFS (ECMWF's File Storage System) provides users with the possibility of archiving any data that is not suitable for storing in MARS. UNIX-like commands enable users to copy files to and from any of ECMWF's computing platforms. Unlike MARS, data on ECFS is not curated or indexed. It is designed to store unstructured data.

MARS data represents about 75% of the volume of data stored in the DHS, but only about 4% of the number of files. ECFS data represents almost all of the remaining 25% of the data, corresponding to 96% of the files. The DHS provided access to over 360 PB of primary and about 350 million files in ECFS and over 13 million in MARS in August 2020. All data related to simulations carried out in this PhD are archived on ECFS as a primary location.

All data produced during the PhD was archived on ECFS, while the more recent river discharge data sets were also stored in the MARS archive.

3.4.1.3 Data accessibility and the Copernicus Climate Data Store

All the official GloFAS river discharge data sets (reanalysis, real time forecasts, reforecast, seasonal real time forecasts and seasonal reforecasts) are currently archived in MARS for the two latest versions, 2.1/2.2 and v3.1 (operational). These data sets are accessible either directly through MARS (from ECMWF computers or the WEB API), or the Copernicus Climate Data Store. All other data sets, produced and analysed in this PhD, are archived on ECFS and available upon request.

The Climate Data Store is the cornerstone infrastructure which supports the implementation of the Copernicus Climate Change Service. It enables the provision of various climate reanalyses, projections and indicators at temporal and spatial scales. The CDS offers seamless web-based and

API-based search and retrieve facilities to access climate data and information. It is an easy way for users to access climate-related data. In addition, the CDS also provides a generic software toolbox that allows users to develop web-based applications that make use of the datasets available in the CDS. All data in the CDS, including the GloFAS data sets, are free and open, subject to the user agreeing to the relevant dataset licence(s). The GloFAS data sets are available via the CDS website at <https://cds.climate.copernicus.eu/cdsapp#!/search?text=glofas>.

3.5 Limitations of the modelling systems and data sets used

The models and data sets used in this PhD have all some generic uncertainties and limitations that need to be considered when interpreting the results of the PhD. The global nature of the modelling, with the very high computational cost, means that some meteorological and hydrological processes are either simplified or neglected (Harrigan et al., 2020). It has to be acknowledged, that regardless of the vast amount of Earth observations assimilated in the production of the meteorological forcing data sets listed in Section 3.2 (<https://www.ecmwf.int/en/research/data-assimilation/observations>), there can still be large uncertainties in the accuracy of the meteorological data, especially for surface variables such as precipitation. For example, precipitation is known to be less predictable in the tropical areas, where convection is dominant and NWP models are known to be struggling to predict convection (Lavers et al., 2022). In addition, surface observations, such as precipitation, are not currently used in these global hydrological modelling systems for the initialisation of the river conditions. This means, the initial uncertainty of river discharge can be large, especially in areas where the meteorological forcing data is of lower quality, for example the tropics. This can have a large negative impact on the quality of the flood forecasts even several days into the forecast horizon (Zsoter et al., 2016).

Chapter 4 How Well Do Operational Numerical Weather Prediction Configurations Represent Hydrology?

The first area that I wanted to explore in the PhD concentrated on the way in which LDAS impacts hydrology in reanalysis simulations. Data assimilation is one of the most important components of the Earth system models. The extent to which LDAS impacts on river discharge was particularly important to look at, as data assimilation is not always active in hydrological simulations, which can potentially create inconsistencies in the flood forecasting systems if LDAS is not used in each component consistently. The paper addressing this research question has been published in *Journal of Hydrometeorology (JoHM)* with the following reference (<https://centaur.reading.ac.uk/83453/>):

Zsoter, E., Cloke, H., Stephens, E., de Rosnay, P., Muñoz-Sabater, J., Prudhomme, C., and Pappenberger, F., 2019: How well do operational Numerical Weather Prediction setups represent hydrology?, *J. Hydrometeorol.*, 14, <https://doi.org/10.1175/JHM-D-18-0086.1>.

The contributions of the authors of this paper are as follows: E.Z. designed the experiment, produced the datasets, carried out the river discharge data analysis, and led the writing of the manuscript. H.C. and E.S. assisted with posing the research question and designing the analysis, P.deR. and J.M-S. helped with the scientific analysis of data assimilation and coupling issues and C.P. and F.P. helped design the research methodology. All authors assisted with writing the manuscript. Overall, 90% of the writing was undertaken by E.Z.

The published article can be found in the Appendix A1.

Abstract. Land surface models (LSMs) have traditionally been designed to focus on providing lower-boundary conditions to the atmosphere with less focus on hydrological processes. State-of-the-art application of LSMs includes a land data assimilation system (LDAS), which incorporates available land surface observations to provide an improved realism of surface conditions. While improved representations of the surface variables (such as soil moisture and snow depth) make LDAS an essential component of any numerical weather prediction (NWP) system, the related increments remove or add water, potentially having a negative impact on the simulated hydrological cycle by opening the water budget. This paper focuses on evaluating how well global NWP configurations are able to support hydrological applications, in addition to the traditional weather forecasting. River discharge simulations from two climatological reanalyses are compared: one “online” set, which includes land–atmosphere coupling and LDAS with an open water budget, and an “offline” set with a closed water budget and no LDAS. It was found that while the online version of the model largely improves temperature and snow depth conditions, it causes poorer representation of peak river flow, particularly in snowmelt-dominated areas in the high latitudes. Without addressing such

issues there will never be confidence in using LSMs for hydrological forecasting applications across the globe. This type of analysis should be used to diagnose where improvements need to be made; considering the whole Earth system in the data assimilation and coupling developments is critical for moving toward the goal of holistic Earth system approaches.

4.1 Introduction

Land surface models (LSMs) have traditionally been designed to focus on providing lower-boundary conditions to the atmosphere by describing the vertical fluxes of energy and water between the land surface and the atmosphere, with less focus on predicting runoff (Mengelkamp et al., 2001). LSMs therefore maximize the quality of the atmospheric forecast, but do not necessarily bring the same benefits in the representation of the hydrological cycle (Kauffeldt et al., 2015).

There is a wide literature on assessing the hydrological capabilities of LSMs and describing various improvements in the modelling of the hydrological cycle (e.g., Wang et al., 2016; Blyth et al., 2011; Wu et al., 2014). However, there are significant limitations in the representation of hydrological fluxes and storages in LSMs, largely due to the large-scale focus of LSM applications, which has led to the neglect of some important processes for runoff generation (Overgaard et al., 2006; Le Vine et al., 2016), including inadequate snowmelt processes (Dutra et al., 2012; Zaitchik and Rodell, 2009).

Data assimilation is an essential part of any numerical weather prediction (NWP) system (Rabier, 2005). It is designed to provide initial conditions for the Earth system by updating the model in all of the components: atmosphere, land, ocean, and sea ice. State-of-the-art NWP configurations, such as used at the European Centre for Medium-Range Weather Forecasts (ECMWF), include both an LSM and a land data assimilation system (LDAS). The objective of the data assimilation in this context is to combine the land surface model state with the available land surface observations to initialize the land surface model prognostic variables of the forecasting system (Bélair et al., 2003a). The current ECMWF LDAS analyses soil moisture, soil temperature, snow mass, density, and temperature (de Rosnay et al., 2014). Land data assimilation was shown to contribute significantly to more skilful atmospheric forecasts, with the soil moisture data assimilation also proven essential in countering a positive precipitation/evapotranspiration feedback which can cause large positive precipitation biases (e.g., de Rosnay et al., 2013; Drusch and Viterbo, 2007; Beljaars et al., 1996).

While the improved surface conditions make LDAS an essential component of the ECMWF NWP system, by design the related increments remove or add water which can potentially have a negative impact on the representation of the hydrological cycle by opening the water budget (Zaitchik and Rodell, 2009; Arsenault et al., 2013; Andreadis and Lettenmaier, 2006a; De Lannoy et

al., 2012; Pan and Wood, 2006). On the contrary, in a system without LDAS and coupling, the errors resulting from atmospheric forcing insufficiencies and imperfect land surface process representations are not corrected by the assimilation of land surface observations.

As an ideal configuration, an Earth system model should always maintain a closed water budget, where the amount of water in the system remains the same. By opening the water budget, river discharge biases could emerge in situations where the LSM has an energy balance bias that is not corrected by the assimilation but only by accurate precipitation and snow accumulation forcing. For example, if the snow in the LSM is melting too slowly, this forces the LDAS to remove water (through snow) artificially to correct for the excessive amount of snow on the surface. If the water that is removed with the snow (and thus could not melt) is not retained within the Earth system that could lead to soil water deficit downstream, potentially causing an incorrect rate of river discharge. In such cases, LDAS could lead to replace incorrect snowmelt timing issue with incorrect snowmelt runoff amount.

Thus, an open water budget could cause problems for associated hydrological forecasting applications, which uses runoff calculated from LSMs with LDAS, such as the Global Flood Awareness System (GloFAS; Alfieri et al., 2013). As global hydrological modelling is increasingly possible with the improved realism that the state-of-the-art LSMs can nowadays offer (Overgaard et al., 2006), it is important to investigate how an LSM with LDAS can support the combined task of traditional weather forecasting and hydrology at the same time. This investigation was undertaken with this dual focus in mind, by analysing the hydrological cycle and the open water budget issues that can help the Earth system model developments with highlighting areas where the coupled system with LDAS does not yet work effectively for flood simulations.

To understand how well an NWP configuration with LSM and LDAS represents hydrology, and in particular to interpret the influence of the LDAS on hydrological simulations from LSMs, in this paper river discharge simulations from two climatological reanalyses of GloFAS are compared: one operational set, which includes land–atmosphere coupling and LDAS with an open water budget, and also an “offline” set with a closed water budget and no LDAS. From these two datasets, a range of hydrological and atmospheric variables will be analysed globally.

4.2 System description, datasets, and methods

Two hydrological experiments, ONLINE (run in operational mode with active land–atmosphere coupling and LDAS) and OFFLINE (run in offline mode without coupling and LDAS) provide time series of various surface variables (e.g., 2-m temperature, snow depth, and runoff), and also discharge after routing the runoff. Figure 4-1 highlights the schematic of ONLINE and OFFLINE with

the main characteristics, components and data periods. In this section the two experiments with the model and data aspects, and the data analysis methods will be described in detail.

4.2.1 Land surface model HTESEL

The hydrological component of the analysed datasets is based on the Hydrology Tiled ECMWF Scheme of Surface Exchanges over Land (HTESEL) land surface model (Balsamo et al., 2009, 2011). HTESEL is part of the ECMWF NWP system and used in coupled land–atmosphere mode on time ranges from short-range to seasonal forecasts. It includes a snow parameterization based on a single-layer snowpack model (Dutra et al., 2010).

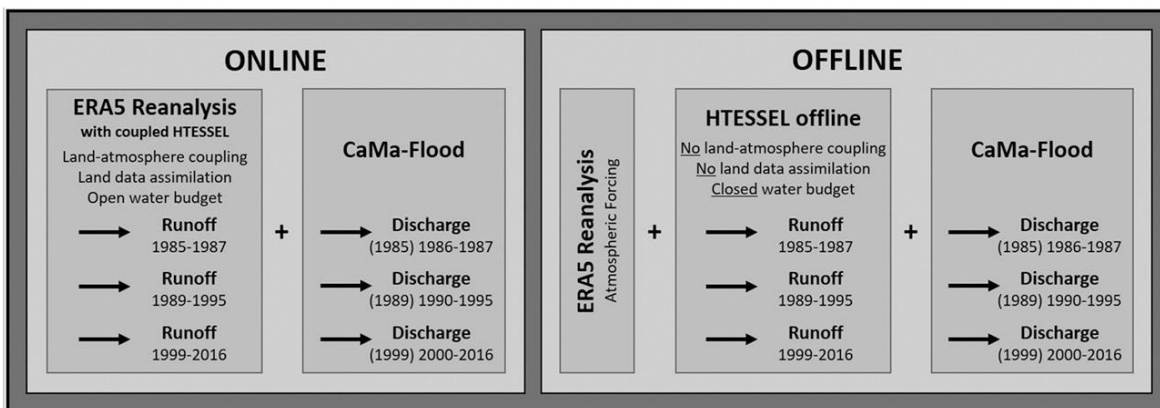


Figure 4-1. Schematic of the ONLINE and OFFLINE experiments that were carried out to produce the ERA5-D25 dataset. The years in parentheses for the discharge indicate the first spinup year in each period that was excluded from the analysis.

The soil vertical diffusion solves the Richards equation using a four-layer vertical discretization with layer depths at 7, 28, 100, and 289cm (Balsamo et al., 2009). HTESEL provides boundary conditions for the atmosphere (heat, moisture, and momentum) by simulating water and energy budgets on the surface and through the soil, snowpack, and vegetation interception. HTESEL generates surface (fast) and subsurface (slow) runoff components at each grid point (Balsamo et al., 2009). Surface runoff depends on the standard deviation of the orography, soil texture, and soil moisture, while subsurface runoff is determined by the soil water percolation.

4.2.2 Land data assimilation

The ECMWF LDAS is part of the ECMWF Integrated Forecasting System (IFS). It is coupled to the atmospheric four-dimensional variational data assimilation (4D-Var) scheme (Rabier et al., 2000), both using a 12-h assimilation window. The upper-air and land surface analyses are running separately and are used to initialize a coupled land–atmosphere short-term forecast, which provides the background for the next data assimilation window. The land data assimilation relies on advanced methods to optimally combine in situ and satellite observations with model background information. A schematic diagram of the ECMWF LDAS is provided in Figure 4-2. Initial

implementations of the ECMWF LDAS relied on simple assimilation methods for snow and soil moisture analyses (Drusch et al., 2004; Mahfouf et al., 2000), with air temperature and humidity measurements being the main input for the soil moisture analysis (Mahfouf et al., 2000; Drusch and Viterbo, 2007). The system has evolved in the past decade to use a more physically based approach and to combine satellite and in situ data in the soil analysis (de Rosnay et al., 2014; de Rosnay et al., 2013; Albergel et al., 2012).

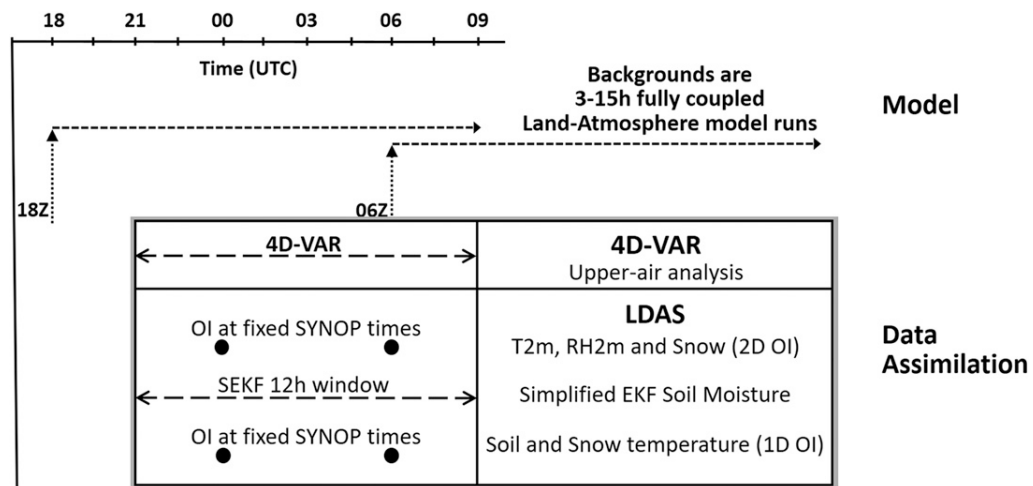


Figure 4-2. Schematic diagram of the land data assimilation system at ECMWF.

In the current LDAS, a simplified extended Kalman filter (SEKF) is used to analyse soil moisture. The approach combines analysed 2-m air temperature and humidity with satellite measurements from the Advanced Scatterometer (ASCAT) sensor on board of MetOp, as described in de Rosnay et al. (2013) and Albergel et al. (2012). For snow, a two-dimensional optimal interpolation (OI) is used to analyse snow mass and snow density following the method described in Brasnett (1999). In situ snow depth observations, available on the SYNOP network are used along with the 4-km resolution snow cover product from the NOAA National Environmental Satellite, Data, and Information Service (NOAA/NESDIS) Interactive Multisensor Snow and Ice Mapping System (IMS) product (Helfrich et al., 2007).

Even though it provides significant improvements to the atmospheric forecasts and independent in situ snow depth measurements (de Rosnay et al., 2015), the current ECMWF snow data assimilation follows a relatively basic method. Operational NWP configurations generally rely on simple approaches, compared to research environment, that are based on more sophisticated snow assimilation methods using in situ and remotely sensed observations (e.g., Helmert et al., 2018; De Lannoy et al., 2012; Pan and Wood, 2006; Slater and Clark, 2006).

The ECMWF LDAS and its performance is presented and discussed in de Rosnay et al. (2014) and de Rosnay et al. (2015). A full description of the technical implementation is provided in the IFS

documentation (<https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model/ifs-documentation>). The system used for this study is that used for the production of ERA5 (section 4.2.5), with IFS cycle 41r2 at a resolution of ~31 km.

4.2.3 CaMa-Flood river routing

The Catchment-based Macroscale Floodplain model (CaMa-Flood; Yamazaki et al., 2011) was applied in this study to simulate the hydrodynamics and produce river discharge from the HTESEL runoff outputs. CaMa-Flood is a distributed global river-routing model which uses a river network map and routes runoff to oceans or inland seas. The CaMa-Flood model was chosen for the routing component as it had already been used in several similar climatological research experiments such as Emerton et al. (2017).

4.2.4 GloFAS

GloFAS is one of the few global scale flood forecasting systems that currently exist (Emerton et al. 2016). It is part of the Copernicus Emergency Management Service (CEMS), developed by the Joint Research Centre of the European Commission and ECMWF. The HTESEL runoff output is coupled to the Lisflood hydrological model over a global river network to produce river discharge with a forecast horizon of 30 days across a global river network at 0.1 degree resolution (van der Knijff et al., 2010; Alfieri et al., 2013).

As part of the GloFAS configuration, the real-time river discharge forecasts are compared with climatological simulations (called reanalysis) to detect the likelihood of high-flow situations. These real-time and climatological datasets also present a unique opportunity for experimental analysis (Emerton et al., 2017; Stephens et al., 2015).

4.2.5 Offline land surface modelling

The current GloFAS operational setup uses a climatology based on the ERA-Interim/Land reanalysis of ECMWF (Balsamo et al., 2015). ERA-Interim/Land is an improved version of the ERA-Interim reanalysis (Dee et al., 2011) produced with an improved version of HTESEL, run offline, using a rescaling of monthly precipitation totals with GPCP v2.2 (Huffman et al., 2009; Balsamo et al., 2010).

Offline HTESEL simulations, such as the OFFLINE experiment in this study, are uncoupled from the atmosphere, without the LDAS and forced with near-surface meteorological input data such as temperature, specific humidity, wind speed, surface pressure, radiative fluxes, and water fluxes. Offline land-surface-only simulations are an affordable way of achieving land surface improvements, and this offline research methodology has been used in numerous studies with

HTESSEL in the last few decades (e.g., Agustí-Panareda et al., 2010; Dutra et al., 2010, 2011; Haddeland et al., 2011).

4.2.6 ERA5 reanalysis

The fifth-generation global climate reanalysis (succeeding ERA-Interim) at ECMWF is ERA5 (Hersbach and Dee, 2016). ERA5 is a key contribution to the EU-funded Copernicus Climate Change Service (C3S). ERA5 will cover the period from 1950 to present and is in production with 2008–17 already officially released. The release of the remaining period is foreseen by end of 2018. ERA5 will then continue running in (non-quality-assured mode) near-real time with only a few days' delay. The data are open access and free to download for all uses (<https://climate.copernicus.eu/>).

ERA5 uses the IFS cycle 41r2 and it relies on land surface model and assimilation configuration that are consistent with those used for operational NWP with coupled land–atmosphere simulations and the latest soil moisture and snow assimilation (see sections 4.2.1 and 4.2.2 above). ERA5 has a high-resolution component at ~31km which is used in this study (hereafter called ERA5-HRES). In ERA5-HRES, variables (analysis and short-range forecasts generated at 0600 and 1800 UTC) are available hourly. Variables that are valid for a period, for example, precipitation or runoff with an accumulation time, are provided as hourly forecasts.

At the time of writing, approximately 28 years of ERA5-HRES data were available in the ECMWF MARS data archive in three separate periods: 1985–87, 1989–95, and 1999–2016. The first years (1985, 1989, and 1999) were used as spinup years, so in total 25 years of daily river discharge and other surface data could be processed for the analysis (hereafter called ERA5-D25).

4.2.7 Experimental setup

In the ONLINE experiment, the operational ERA5-HRES reanalysis data were used directly from all three ERA5-HRES periods for land surface variables, including runoff, produced by coupled land–atmosphere model with LDAS and an open water budget (Figure 4-1). In the OFFLINE experiment, on the other hand, three stand-alone HTESSEL runs were set up, one for each of the periods, to reproduce the land surface variables in land surface only mode without the impact of coupling and LDAS, but with a closed water budget. As ERA5 has a recent model cycle (41r2), the same HTESSEL version could be used in the offline experiment as in the operational ERA5.

In the ECMWF NWP system, there is no option currently to run the land–atmosphere coupling and LDAS separately. Either both are active as in ONLINE, or neither of them as in OFFLINE. It would be interesting to separate the impact of these two contributing modelling options, but as they are too strongly interwoven the separation would require a very large effort, which is outside of the scope of this study.

In the OFFLINE experiment, the offline HTESSSEL model was forced with hourly ERA5-HRES atmospheric data, wherever it was possible on the lowest model level, with an hourly model time step. The model was run on the original horizontal resolution of ERA5-HRES (~31 km). For precipitation, temperature, specific humidity, wind speed, and surface pressure the hourly analysis fields were applied, while for radiation and precipitation fluxes the first 12-h period of the 0600 and 1800 UTC short-range forecasts were used to cover each 24-h periods.

The river discharge was generated by routing the runoff using CaMa-Flood for both the ONLINE and OFFLINE datasets over the ~25 km river network. CaMa-Flood was run with a 1-h time step and a 24-h output frequency to match the 24-h reporting frequency of the river discharge observations.

4.2.8 River discharge observations

In this study, daily river discharge observations used in the GloFAS system are selected. These are mostly from the Global Runoff Data Centre (GRDC) archive, an international depository of river discharge observations and associated metadata.

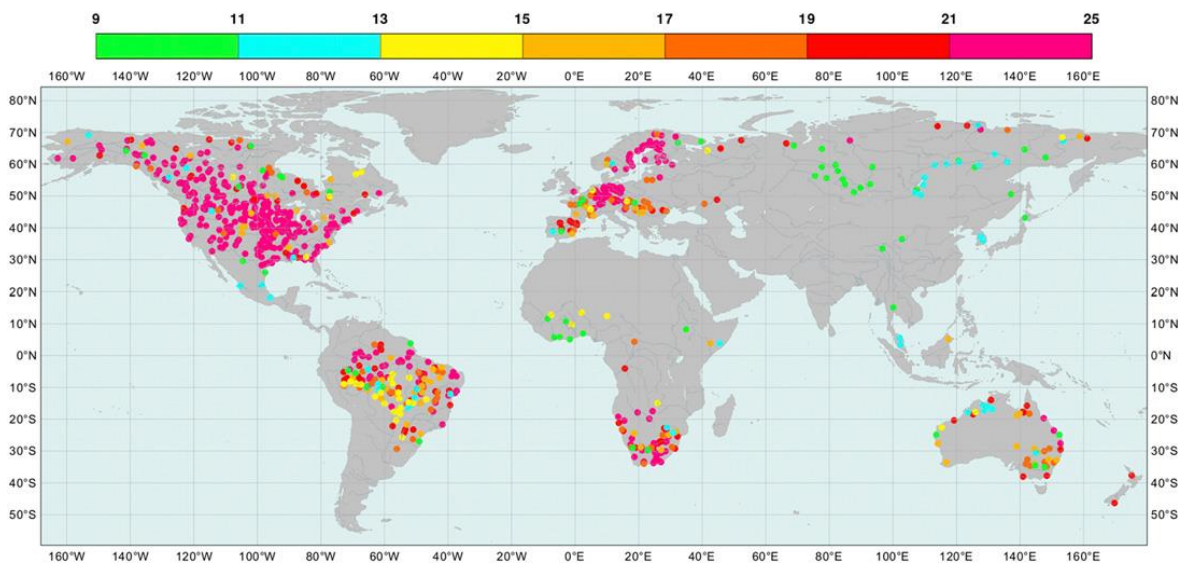


Figure 4-3. Geographical distribution of river discharge observations with sufficient record length selected for the analysis. Colours indicate the length of the available data in years (from 9 to 25).

The observations consist of a network of approximately 900 river gauging stations with upstream areas over 10 000km², selected from the catchments used in Zsoter et al. (2016). After visual inspection those catchments that showed a clear non-realistic behaviour and/or influence of dams were excluded. During the visual checks, the focus was on excluding stations that have obvious observation errors, such as constant values, linearly interpolated periods or unrealistic jumpiness, and stations with signatures of regularities that are unlikely to come from weather and suggests dam operation impact on the observed river discharge. In total, about 15% of the catchments were excluded for quality reasons. A minimum of 9 years, with at least 330 days in each of those calendar

years, was selected as criteria for the stations to be included in the river discharge analysis. This is quite a short period, but due to the limited availability in more recent years, it was accepted as a compromise. In total 590 stations could be processed globally leaving large blank areas mostly in Asia and Africa (Figure 4-3).

4.2.9 Annual peak river discharge

For the river discharge verification, the annual peak river discharges from the two ERA5-HRES simulations were determined in each calendar year as the highest value in the ± 30 -day window around the observed annual maximum river flow. The 30-day window was defined as a safeguard to avoid detecting high skill with similar peaks in observation and simulation of completely different flood waves at very different periods of the year.

4.2.10 Water budget increments

This study focuses on the impact of the water budget closure on river discharge. To analyse this, the daily (0000–0000 UTC) water budget error term dA was computed as

$$dA = P - E - R - dS, \quad (4-1)$$

where P is precipitation, E is evapotranspiration, and R is runoff, all taken as the sum of the hourly forecast values (24 in total) in the ONLINE experiment from the 0000–0000 UTC period, and dS is the change in the storage term (water content in the soil including all four layers and also in the snow cover) computed as the difference between the two subsequent 0000 UTC analysis values in ONLINE (representing the change in the water content during the 24-h period). Even though the water budget error is zero in OFFLINE (the water budget is closed), the contributing variables can help identifying the behaviour of the surface processes in both the ONLINE and OFFLINE simulations.

The imbalance in the amount of water that is not accounted for in the ONLINE water budget effectively comes from the snow depth and soil moisture increments in LDAS which remove or add water in the system. The daily increments (valid for a 0000–0000 UTC 24-h period) are computed as the sum of two increment values at 0600 and 1800 UTC (each day). Both of these increments are computed as the ERA5-HRES analysis value minus the corresponding 12-h ERA5-HRES forecast value (initialized 12 h earlier).

4.2.11 Daily 2-m temperature and snow depth

The in situ surface synoptic observations (SYNOP) were used to verify 2-m temperature and snow depth for both the OFFLINE and ONLINE experiments. The observing stations were filtered according to the station altitude difference to the model orography and only those were used which

had less than 150-m discrepancy, as orography has control on both variables and large differences would make the comparison unreliable. This maximum orography difference value was chosen in accordance with the general practice at ECMWF, where 100m is used to filter stations in the 2-m temperature verification. For our study, a less stringent compromise value was preferred in order to increase the sample size and still guarantee good match between model and real orography.

The 2-m temperature was verified for around local noon (Table 4-1), while for snow depth the first measurement of the calendar day was evaluated in case of subdaily records. In total, observations from about 4000 stations for 2-m temperature and 1500 stations for snow depth were available for verification. For each catchment, a representative daily observation was also determined for both variables.

Table 4-1. Criteria for selecting daytime 2-m temperature.

Longitude band	30°W–60°E	60°–150°E	150°–180°E	120°–180°W	30°–120°W
Approx. local noon	1200 UTC	0600 UTC	0000 UTC	0000 UTC	1800 UTC

For catchments with more than one SYNOP station available, these were calculated as the arithmetic average of the stations within the catchment. It has to be acknowledged that the observation network available was not dense enough to represent the full spatial variability of these surface variables, especially snow depth, which vary dramatically in space from one point to another (Molotch and Bales, 2005). However, for a global study on the hydrological impacts it is expected to be sufficient.

4.2.12 Climatologies

Daily climatologies were used for river discharge and other surface variables in this work for both observations and the two simulations. These datasets were produced with all potentially available 25 years of data in ERA5-D25, always matching the number of available nearly complete calendar years (with minimum 330 river discharge observations) for all the catchments. For each day of the year a 21-day window, centred over the day, was used, which provided a minimum of about 180 values in the climate sample (with the 9 years minimum criteria). The only exceptions are 2-m temperature and snow depth, where a fixed shorter period of 2000–07 was used without the criteria of nearly complete years. As the 2-m temperature and snow depth observation availability is much better in more recent periods and less prone to missing values than river discharge, a shorter fixed period (when ERA5-HRES was available) is sufficient.

Table 4-2. List of verification scores used in the analysis with a short description and the areas where they were applied.

Score	Description	Used for
ME	Mean error	Daily river discharge, snow depth, and 2-m temperature
MAE	Mean absolute error	Daily river discharge, snow depth, and 2-m temperature
NSE	Nash–Sutcliffe efficiency	Daily river discharge time series
R	Pearson correlation coefficient	Daily river discharge time series
PMnE	Percentage sample mean error	Whole river discharge sample
PMnAe	Percentage sample mean absolute error	Whole river discharge sample
PStE	Percentage sample standard deviation error	Whole river discharge sample
PStAe	Percentage sample standard deviation absolute error	Whole river discharge sample
PkTiMe	Peak timing mean error	Annual river discharge peaks
PkTiMae	Peak timing mean absolute error	Annual river discharge peaks
PPkMgMe	Percentage peak magnitude mean error	Annual river discharge peaks
PPkMgMae	Percentage peak magnitude mean absolute error	Annual river discharge peaks

4.2.13 Verification statistics

A number of statistics were applied to evaluate the overall performance of the two climatological simulations in ERA5-D25 (Table 4-2). Several scores were selected in order to give a more representative description of the general behaviour including the differences between the ONLINE and OFFLINE experiments. This is recommended, for example, by Legates and McCabe (1999) as different scores demonstrate different aspects of the model attributes ultimately providing a more complete picture.

The climatological daily time series were compared to the observed data using mean error (ME), mean absolute error (MAE), Nash–Sutcliffe model efficiency (NSE; Nash and Sutcliffe, 1970), and Pearson correlation coefficient R (Pearson, 1896) in order to measure the fit between model and observations. In addition, the mean and standard deviation of the observed and modelled values were analysed with four additional indices, the percentage sample mean error, the percentage sample mean absolute error, the percentage sample standard deviation error, and the percentage sample standard deviation absolute error.

Another very important aspect of hydrological model verification is the ability of the systems to correctly predict the extremes, as these events can cause the highest impact. To measure this, the timing and magnitude errors of the annual peaks were considered. Both the ME and MAE measures (mean of all years in the sample) were computed for the timing and for the percentage magnitude errors using the annual peaks over the 25 analysed years (for details on how the annual peaks were computed, see section 4.2.9). For the analysis of the data assimilation impact on 2-m temperature and snow depth the ME and MAE scores were used. In this study verification was conducted on homogeneous samples across all compared scores for all the verified surface variables.

4.3 Results

The river discharge behaviour provides a useful indication of the hydrological differences between the ONLINE and OFFLINE simulations. However, in order to understand the underlying processes better, the coupling and LDAS impact was also analysed globally and regionally based on the water budget and the related surface variables.

4.3.1 Snow depth and 2-m temperature impact

The LDAS is designed to provide adequate initial surface conditions to the NWP forecasts. The impact on the hydrology could be demonstrated on two important surface variables: 2-m temperature and snow depth (at least in snow impacted areas) which are relatively well observed variables and can be used to analyse the impact of the land–atmosphere coupling and LDAS on the surface globally in the two experiments. For details on how the observations were used please see section 4.2.11.

The picture for 2-m temperature is rather mixed geographically with an overall MAE improvement in ONLINE of around 0.3–0.4°C as a global average up to 1–2°C locally (not shown). This corresponds to about 20–30% decrease in MAE on average in ONLINE, with the impact of coupling and LDAS, compared to OFFLINE.

The improvement in the snow depth, which has much larger direct impact on the hydrology, is more pronounced, based on the stations used in this study. The errors in ONLINE are significantly reduced with most stations showing below ± 1 –2cm of ME (not shown), and decrease of MAE by as much as 10–20cm in some of the snow dominant locations in the 50–70° latitude band (Figure 4-4). This is a very large improvement in ONLINE by removing 70–80% (as global average) of the errors found in the OFFLINE experiment. Countries of Central America, including Mexico, Venezuela, and Columbia, tend to provide snow information in their SYNOP observations.

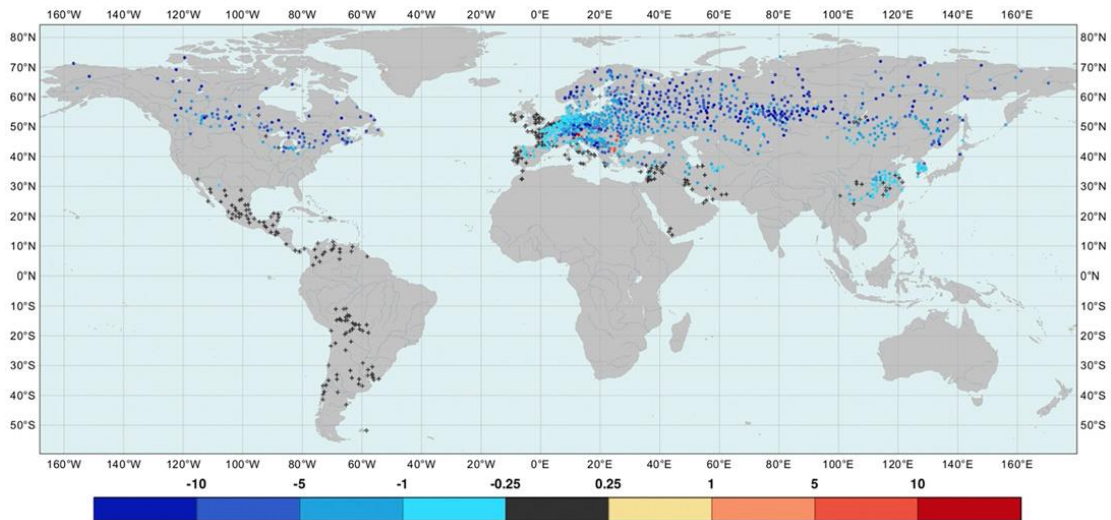


Figure 4-4. Difference in the snow depth mean absolute errors between *ONLINE* and *OFFLINE* for January based on observations in 2000–2007 (cm). Points are shown where observations are available. Blue colours indicate lower errors in the *ONLINE* experiment.

* An additional similar figure (4-4b) was added for the impact on 2m temperature in Appendix A8. Please be aware that this figure does not appear in the published version of this chapter.

In these regions both the model and the in situ stations mostly indicate snow free conditions, leading to very low MAE as shown in Figure 4-4. Although the improvements are large, this does not necessarily mean that the simulation is generally better. In situ snow observations are associated to potential representativeness issues, particularly in mountainous areas. When assimilating a nonrepresentative dataset at a coarse special scale, the results can potentially degrade, even though the match to the actual observations is better (Molotch and Bales, 2005). As the 2-m temperature and snow depth observations used in this study for verification were also assimilated in ERA5, the result will favour to some extent the *ONLINE* experiment.

4.3.2 Global water budget analysis

The water budget is closed in *OFFLINE* by design, while in *ONLINE* the LDAS increments can add or remove water, which could potentially lead to large errors in the budget over a long period. The first aspect that was important to check is the amount of water that is lost or gained in a day on average in the hydrological cycle.

Figure 4-5 shows the average daily water budget errors (Figure 4-5a) and the related snow water equivalent (Figure 4-5b) and soil water content (Figure 4-5c) increments (for the definition of these terms please see section 4.2.10). In Figure 4-5, negative values (red) indicate water removal by LDAS, while positive values (blue) show where water is added to the hydrological cycle.

The three figures highlight significant biases in the ONLINE experiment as these water budget errors represent generally $\pm 10\text{--}25\%$ of the total precipitation with locally even higher ratios (not shown).

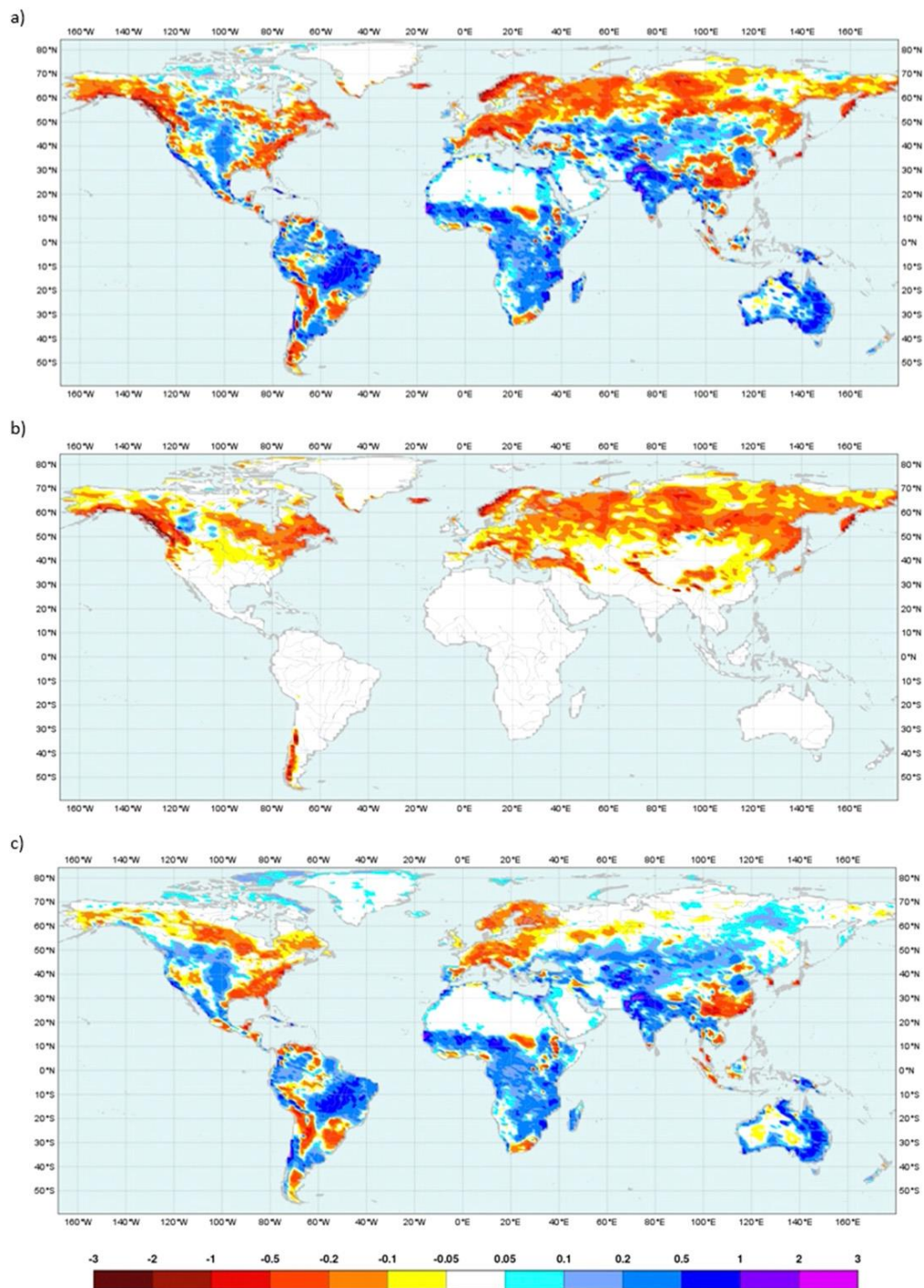


Figure 4-5. Average daily water budget analysis (mm/day) of the ONLINE experiment based on the ERA5-D25 dataset for (a) the total 24-h water budget errors, (b) the 24-h snow water equivalent increments, and (c) the 24-h soil water content increments. Negative values (red) indicate water removal by LDAS, while positive values (blue) show where water is added to the hydrological cycle.

In addition, at latitudes higher than 50°N the dominant pattern is a negative water budget error (Figure 4-5a). The major contributing factor to the clearly negative errors in this area is the correction of snowpack with LDAS removing snow to account for possible inaccuracies in the Chapter 4 How Well Do Operational Numerical Weather Prediction Configurations Represent Hydrology?

HTESSEL snow scheme (Figure 4-5b). On average snow water increments are negative almost everywhere where snow is present. The only notable exception is in Canada, where some central areas have positive water budget errors which could possibly come from a negative precipitation bias that needs to be compensated by LDAS.

Other areas of the world—the central United States, most of Amazonia, Africa, South Asia with India, and large parts of Australia—show positive errors in Figure 4-5a, where extra water is added by LDAS. However, the positive errors are not exclusive, as large parts of China, the southeastern United States, and areas in central South America experience negative water budget errors in these mostly warm climatic conditions. Most of these increments come from the soil moisture assimilation impact (Figure 4-5c). The soil moisture assimilation can generally compensate for precipitation or 2-m temperature biases. For example, if the 2-m temperature is too low, the assimilation will remove water, therefore reducing evaporative cooling which subsequently increase the temperature in general.

4.3.3 Catchment-level process examination

To demonstrate how HTESSEL handles the land surface processes with and without coupling and LDAS, an in-depth case study analysis of the annual water budget cycle was performed for an example catchment on the Amur River in east Russia (see Figure 4-6, catchment 13). This catchment is heavily snow impacted during winter and can demonstrate nicely the important aspects of the hydrological cycle behaviour with the LDAS in action.

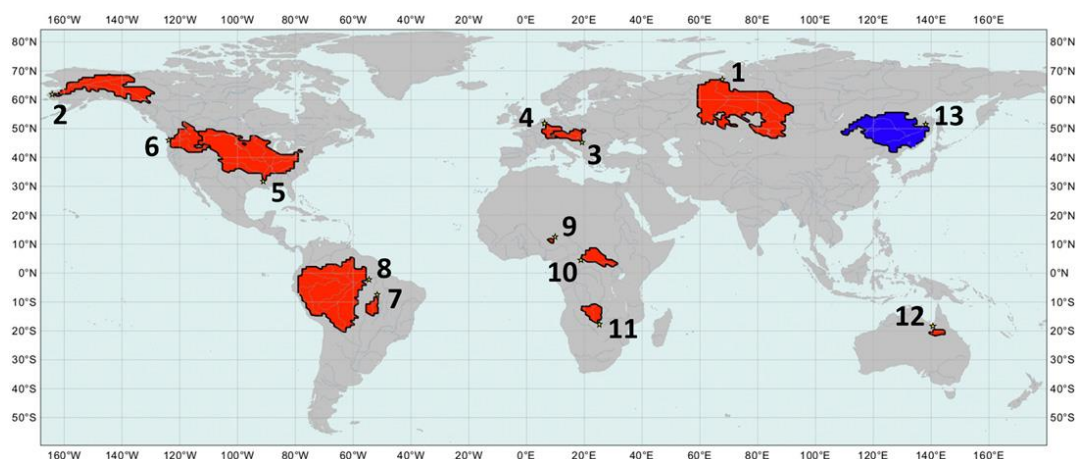


Figure 4-6. Map of the catchments analysed in section 4.3.3 (Figure 4-7), where the catchment-level process is examined over the Amur River (blue area, 13), and in section 4.3.4 (Figure 4-8), where the simplified representation of the annual water cycle is shown for some selected regional catchments of the world (red areas, 1-12). The catchment details are provided in Table 4-3.

In the HTESSEL hydrological cycle representation the input precipitation combined with the melted part of the snowpack (snowmelt) is distributed into evapotranspiration, runoff (as sum of surface

and subsurface runoffs), snow water storage (falling snow part of the precipitation) and soil water storage (soil moisture in the four soil layers). The daily water budget error, computed as in Eq. (4-1) (without the snowmelt separated), is zero in OFFLINE, while ONLINE can show errors due to the increments adding or removing water. Figure 4-7 summarizes the annual cycle of all the water budget contributing variables. The displayed variables are daily climatological means calculated as described in section 4.2.12. The following variables are shown in Figure 4-7: simulated precipitation (same for both experiments), evapotranspiration, runoff, soil water, and snow water storage terms [in Eq. (1)] for both ONLINE and OFFLINE; snow and soil water content increments for ONLINE; simulated snowmelt, snow depth, and river discharge for both the ONLINE and OFFLINE experiments; and finally the corresponding river discharge and snow depth observations.

Figure 4-7 shows that for the Amur the ONLINE simulation significantly improves the representation of snow depth, but as consequence, by the snow assimilation removing a lot of snow, it drastically reduces the river discharge peak seen during the snowmelt season. The explanation of this conclusion with detailed analysis of the evolution of the different surface variables in the different seasons is given in the following:

- Winter: During December–February there is relatively little activity. The little amount of precipitation falls mostly as snow, building the snowpack. Some snow is removed by the assimilation through the small negative snow increments. Water leaves the bottom of the soil as subsurface runoff with hardly any surface runoff. The OFFLINE simulation is generally similar to ONLINE, but snow depth bias shows increasingly positive values in OFFLINE due to the extra amount of water going into the snowpack in the OFFLINE experiment from snowfall (especially during first half of the winter).
- Spring: From March, there is a pronounced snowmelt period in the model, peaking at the end of April, lasting until the middle of June (with virtually zero snowpack in catchment average after middle of May). The increased precipitation in this spring period, with the large amount of snowmelt, increases the soil water content, and results in larger surface runoff output in both experiments. However, the snowmelt is much smaller in ONLINE during April–May as a direct consequence of the large negative snow increments (peaking early April) removing snow in the ONLINE experiment. Similarly, due to the smaller amount of available water in ONLINE, the surface runoff is also significantly smaller mainly in April/May. The snow depth errors peak in middle of March by about 5 cm in OFFLINE with no errors in ONLINE (as catchment average). The data assimilation rightly corrects this substantial positive snow bias, however, the removed snow will be missing from the water cycle, as highlighted by the unnoticeable spring peak river flow, which is higher in the OFFLINE simulation mainly due to the extra snowmelt.

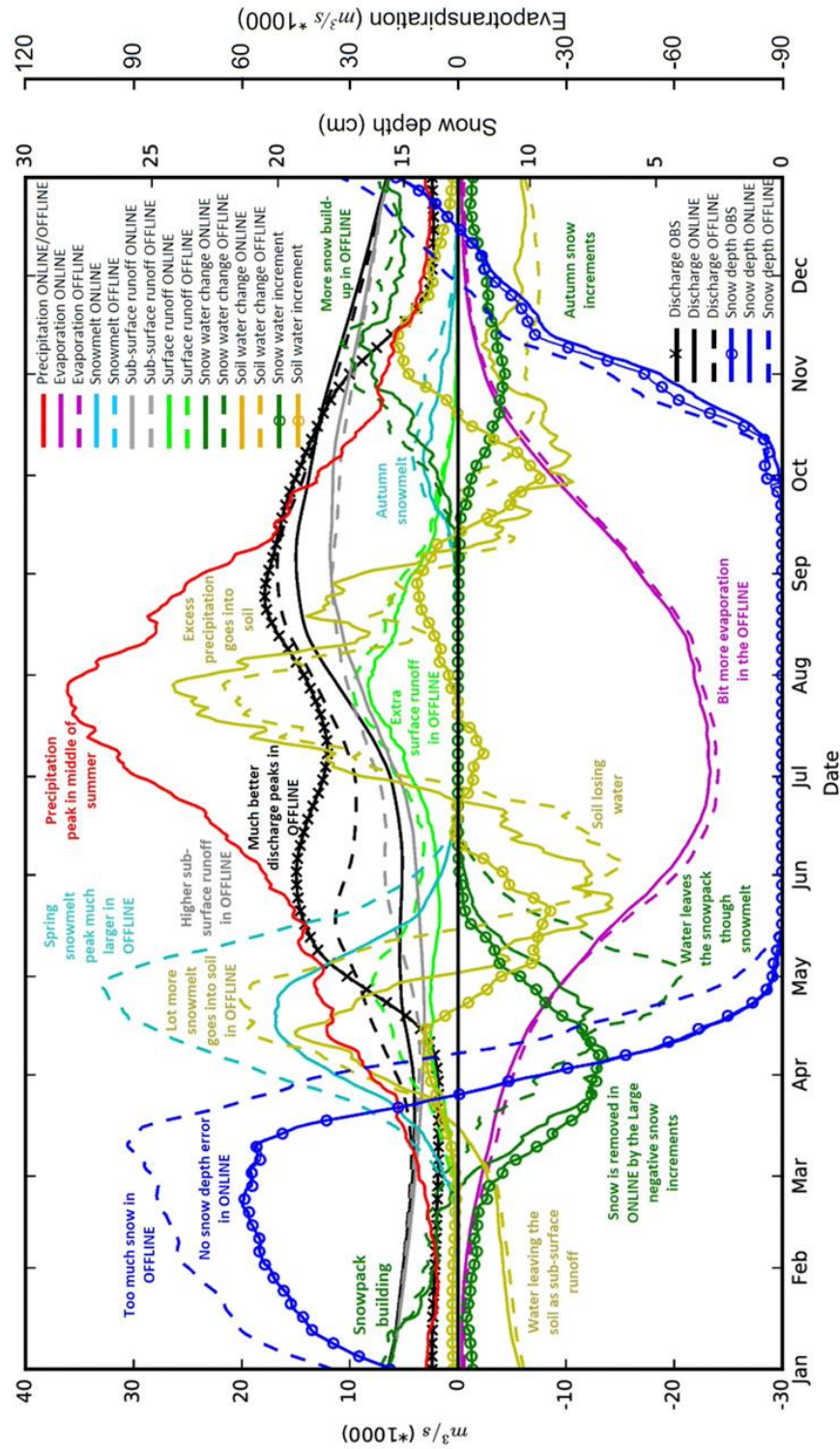


Figure 4-7. Average daily water budget cycle for a catchment on the Amur River in Russia at Komsomolsk. It includes the following parameters: precipitation (red line), snow (green line with markers), and soil (mustard line with markers) water content increments for the ONLINE simulation; surface runoff (light green), subsurface runoff (grey), evapotranspiration (magenta), snowmelt (cyan), and soil (mustard) and snow (green) water storage daily changes for both ONLINE (solid lines) and OFFLINE (dashed lines) experiments and observations (lines with markers). The snow depth values are based on 2000-2007, while all other displayed daily climatological means are based on the ERA5-D25 dataset (for more detail on the computation of these values, see sections 4.2.11 and 4.2.12).

- Snowmelt problem: This behaviour of HTESSSEL with LDAS is rather surprising, and at first it might sound like a contradiction. How can the correct snow conditions lead to such poor river discharge in the ONLINE experiment? A possible explanation could be the representativeness issue of some of the snow observations, which can potentially cause local degradation in some of the catchments. It can also be explained by the HTESSSEL's tendency to melt the snow too slowly (Dutra et al., 2012). In its simple, single layer snow scheme, too much snow accumulates into the snowpack and then that snow melts too slowly. For example, during a 20mm mixed snow/rain forecast event (10mm liquid and 10mm solid) the snow scheme will accumulate most of the 10mm solid (snow) part of the precipitation into the snowpack regardless of the temperature conditions and melt only a little of this 10mm. However, in reality a lot of that rain, sleet, or wet snow would not accumulate on the ground, and instead most of it would melt straightaway. It seems the OFFLINE simulation gets the river discharge right mainly for the wrong reasons. Although the snowpack is clearly more poorly represented, the better timing with the delayed snowmelt (through the too slow melting) and the extra water in the snowpack, the OFFLINE experiment gets the runoff peak more correct.
- Summer: The water budget is balanced between precipitation and evapotranspiration with some soil water increments. During early summer water is taken out of the soil to cover the higher evapotranspiration. In OFFLINE more water leaves the soil which increases the runoff and evapotranspiration. By August, however, the excess water from precipitation over evapotranspiration goes again into the soil, which is more pronounced in ONLINE where the soil is drier. The end of summer river discharge peak is present in both simulations, with the OFFLINE showing a better peak due to more water in the soil and subsequently higher surface and subsurface runoff during all summer. The OFFLINE river discharge exceeds the ONLINE values all summer and the two will level out by September, when the runoffs become similar in the two experiments.
- Autumn: From the middle of September there is another smaller snowmelt period starting with the falling temperatures and bringing some negative snow increments in the ONLINE simulation. The snow accumulates into the snowpack in both experiments, but again with a higher rate in OFFLINE, and with larger snowmelt amounts in OFFLINE.

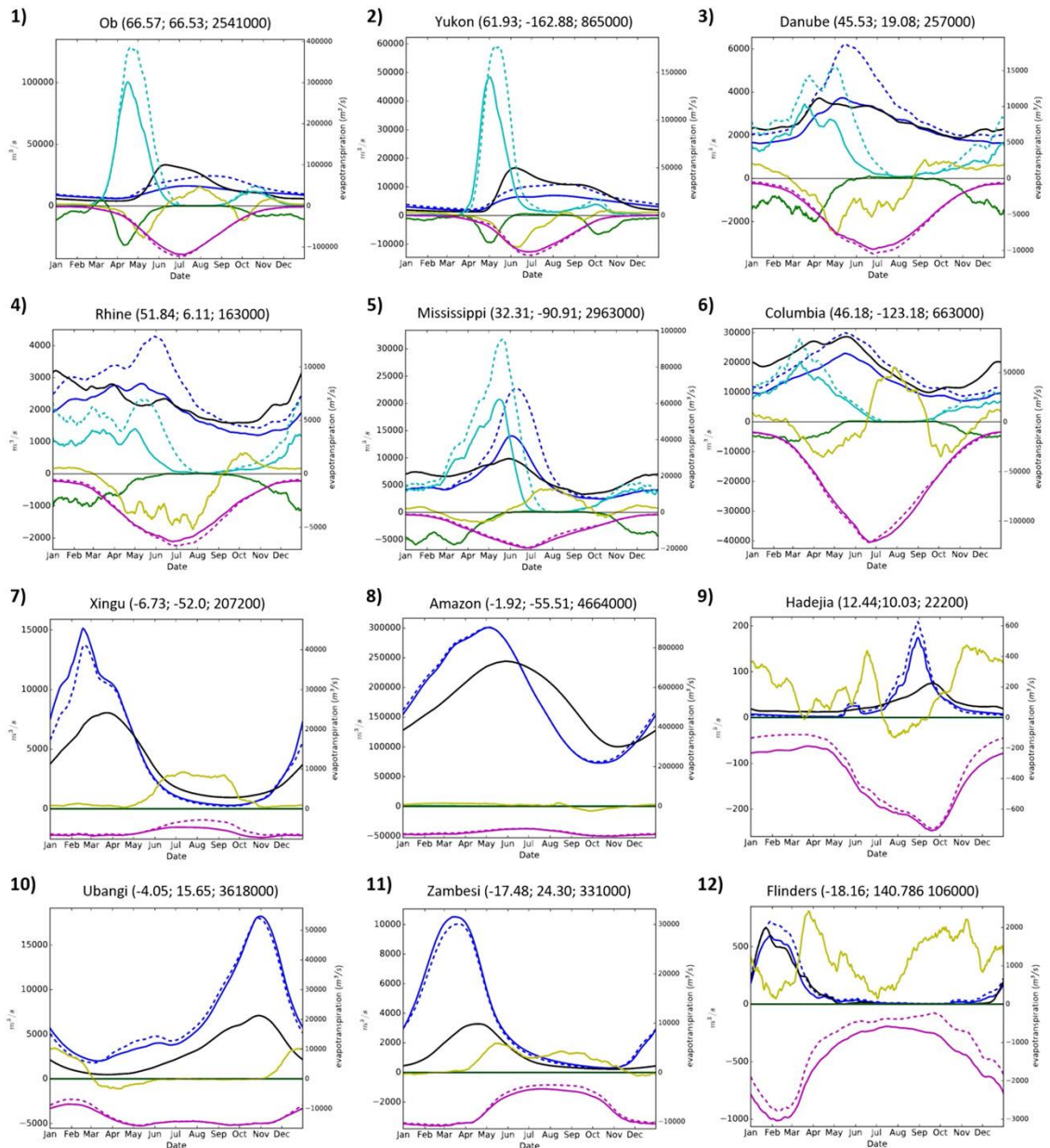


Figure 4-8. The annual cycle of water budget variables for a selection of catchments worldwide numbered from 1 to 12 (see Fig. 6). The displayed variables are the snowmelt (cyan), evapotranspiration (magenta), and river discharge (blue) for both the ONLINE (solid lines) and OFFLINE (dashed lines) experiments; the snow (green) and soil (mustard) increments for ONLINE; and the river discharge observations (black line). All values are daily climatological averages based on the ERA5-D25 dataset (for details on the computation of these values, see section 4.2.12). The river names, the gauge coordinates, and the upstream area values are displayed in the subplot titles. The catchment descriptions with the main verification score values for the ONLINE and OFFLINE simulations are provided in Table 4-3. In addition, the catchment area contours are provided in Figure 4-6. The evaporation scale is provided on the secondary vertical axis, while the scale for all other parameters is shown on the main vertical axis.

4.3.4 Regionally representative catchments

In the previous section the LDAS response was highlighted for an important weakness of HTESSEL with significant consequences on river discharge. In the following, the land–atmosphere coupling

and LDAS impact is now demonstrated with a simplified representation of the annual water cycle in different geographical areas and also various climatic conditions for a selection of the world's catchments in Figure 4-8. The displayed variables are simulated snowmelt, evapotranspiration, and river discharge in both the ONLINE and OFFLINE experiments, the snow and soil water increments for ONLINE, and finally the river discharge observations. All values are daily climatological mean values as in Figure 4-7. The location of the catchments is provided in Figure 4-6. In Figure 4-8, 12 catchments are selected to represent all main areas of the world where river discharge observations are available. Many of them are very large rivers, some of the catchments are dominated with mixed snow and soil moisture influence from the Northern Hemisphere, while others, mainly in the tropics, are only soil moisture impacted. In Table 4-3, the main catchment details are provided (following the numbering from Figure 4-6), complemented with the NSE and the percentage peak magnitude ME and MAE values for the catchments. Bold numbers denote the better score of ONLINE and OFFLINE.

Table 4-3. Details of the 13 catchments analysed in Figure 4-7 (13) and Figure 4-8 (1–12) with the NSE, percentage peak magnitude ME (PPkMgMe) and percentage peak magnitude MAE (PPkMgMae) score values for the ONLINE and OFFLINE experiments. Bold scores denote better performance. For further details on the scores see section 4.2.13.

Catchment No.	Station	River	Area *1000km ²	NSE		PPkMgMe (%)		PPkMgMae (%)	
				ONLINE	OFFLINE	ONLINE	OFFLINE	ONLINE	OFFLINE
1	Salekhard	Ob	2541	0.40	0.52	-55.0	-40.7	55.0	40.7
2	Pilot station	Yukon	865	0.31	0.64	-64.7	-50.7	64.7	50.7
3	Boogojevo	Danube	257	0.47	-0.43	-3.5	29.1	19.8	32.4
4	Lobith	Rhine	163	0.45	0.05	-39.1	-14.8	39.1	18.5
5	Viicksburg	Mississippi	2963	-0.02	-2.69	1.6	31.4	17.7	43.5
6	Quincy	Columbia	663	0.25	0.54	-24.0	-7.6	27.5	20.2
7	Boa Sorte	Xingu	207	-1.53	-0.85	159.0	147.9	159.0	147.9
8	Obidos-Linografo	Amazon	4664	-0.17	-0.21	26.6	26.9	26.6	26.9
9	Hadejia	Hadejia	22	-9.01	-11.85	297.1	436.1	297.1	436.1
10	Bangui	Ubangi	496	-5.72	-6.17	162.8	159.1	162.8	159.1
11	Katima Mulilo	Zambesi	331	-7.97	-6.70	196.6	183.0	196.6	183.0
12	Walkers bend	Flinders	106	0.66	0.62	224.5	211.4	46.9	45.9
13	Komsomolsk	Amur	1846	0.43	0.68	233.5	218.7	33.5	18.7

Figure 4-8 suggests that the decreased snowmelt is a general feature in ONLINE across the Northern Hemisphere as predicted already by Figure 4-5b. All displayed catchments have generally lower

river discharge in ONLINE, either concentrated over the high river discharge season [e.g., Ob (1) and Yukon (2)], or elongated over most of the year [e.g., Danube (3) and Rhine (4)]. The snowmelt is universally smaller in the ONLINE simulation, with the LDAS removing snow at different periods of the year, which seems to be the driving force behind the river discharge differences.

The decreased amount of water has a mixed river discharge skill impact. For some catchments [Ob (1), Yukon (2), Columbia (6), and the case study catchment on the Amur (13)] the change during the high river discharge season is disadvantageous in ONLINE, confirmed by mostly negatively impacted scores, such as the NSE and the percentage peak magnitude MAE values in Table 4-3. On the other hand, for the Mississippi (5), Danube (3), and Rhine (4), it is rather beneficial as the daily climatological mean river discharge is closer to the corresponding observations during the high season, accompanied with mainly positive skill changes in the ONLINE experiment as both NSE and percentage peak magnitude MAE improves (Table 4-3), except the Rhine catchment (4), where the percentage peak magnitude MAE deteriorates.

In the warm climate, however, where soil water dominates the land surface processes [Xingu, Amazon, Hadejia, Ubangi, Zambesi, and Flinders (7–12)], the land–atmosphere coupling and LDAS impact on river discharge seems to be smaller than for the snow-influenced catchments, and on evapotranspiration it tends to be larger. There are large biases over five of the six highlighted tropical catchments (the only exception being the Flinders River in Australia), where both the ONLINE and OFFLINE experiments show significant mismatch with the observed values for the total river discharge volume and for the annual peaks. For example, as displayed in Table 4-3, on the Hadejia River in Nigeria the percentage peak magnitude ME is 297% (the simulation is almost three time higher than the observation) in ONLINE, which is significantly better than OFFLINE (the improvement is 139% in the percentage peak magnitude MAE). This points to the fact that even though the river discharge differences are smaller in relative terms, it can still lead to noticeable change in the scores for some of these highlighted catchments (Table 4-3).

Even though there is no clear systematic difference between the exclusively soil moisture and mixed (snow and soil moisture) catchments in terms of river discharge skill impact, the snow clearly looks to carry a more direct influence on the river discharge volume and river discharge skill.

4.3.5 Global river discharge analysis

In the previous sections it could be shown that the water budget is out of balance in the ONLINE simulation over large parts of the world leading to significant impact on the river discharge for the analysed list of catchments. As an extreme example, it was demonstrated that the snowmelt-driven spring river discharge peak was almost completely missed in a large catchment in east Russia in

ONLINE. After the individual catchment examples, a systematic analysis of the river discharge quality in the ONLINE and OFFLINE experiments is provided based on all available catchments globally.

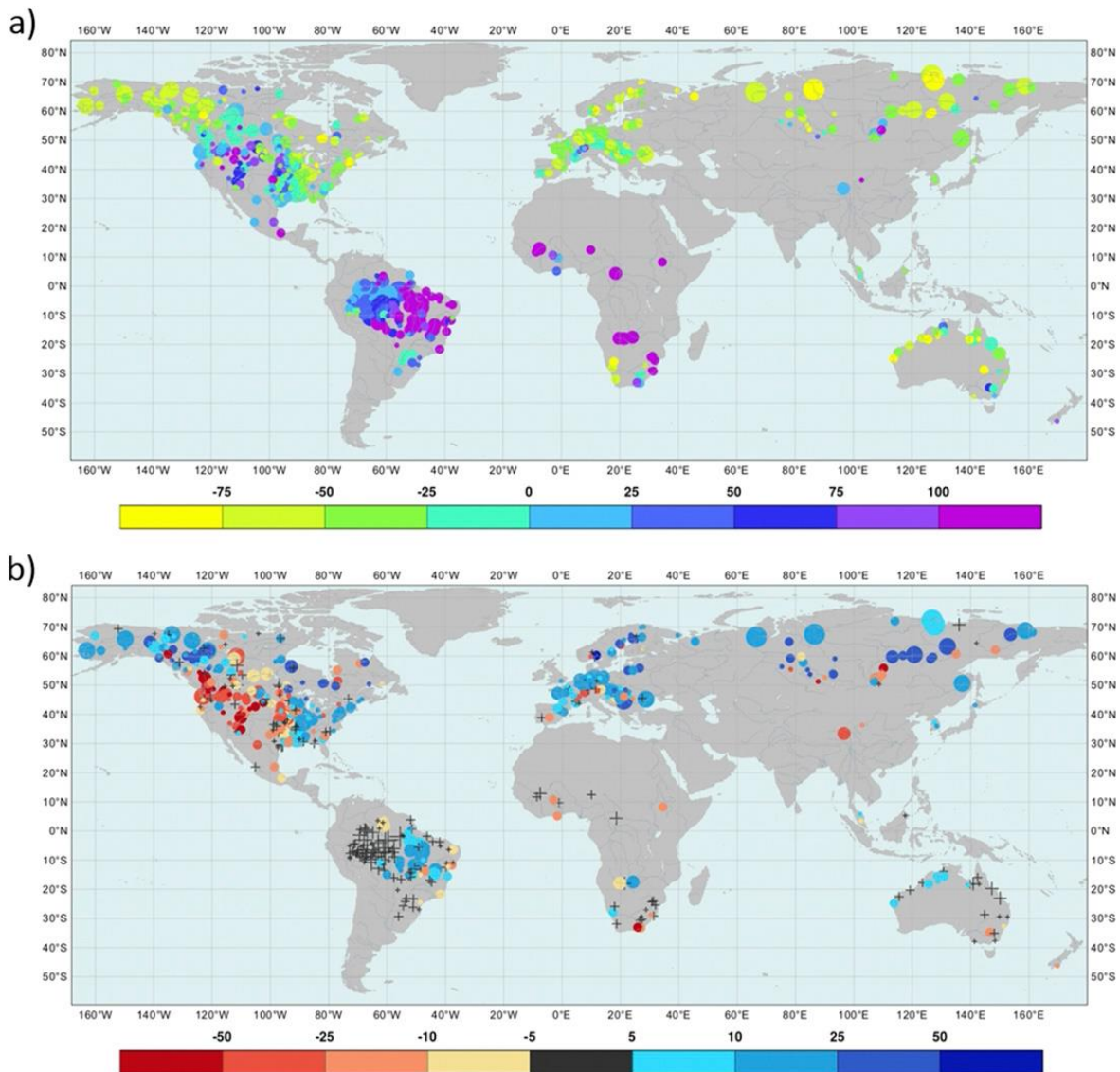


Figure 4-9. River discharge (a) percentage peak magnitude ME (%) of the ONLINE experiment and (b) change in the percentage peak magnitude MAE (%) between ONLINE and OFFLINE based on the ERA5-D25 dataset. Positive error differences in (b) indicate deterioration (blue) while negative changes show improvement (red) in the ONLINE simulation compared with OFFLINE. The catchments are displayed with different marker sizes representing the size of the catchment area. Near-zero differences are shown by black crosses, while all other categories are displayed by circles.

Although a large number of scores was computed in this study, this section will focus only on the annual peak flow scores. The timing and magnitude of the high river discharges are both crucial aspects of river discharge simulations in any flood prediction system such as GloFAS. The accurate simulation of the river discharge peaks is essential to get the best possible guidance for the potentially most damaging floods. The analysed performance of the annual peak river flows should give a good indication on the general ability of the two experiments to predict peaks.

Figure 4-9a highlights a large systematic percentage peak magnitude ME in the ONLINE simulation. Many catchments show over 50% error (either positive or negative) of the annual river discharge peaks on average. The majority of the Northern Hemispheric higher latitudes is overwhelmingly underpredicted, while Amazonia, the western United States, and many catchments in Africa are overpredicted in the ONLINE experiment. The geographical pattern in Figure 4-9a is rather similar to the one seen in Figure 4-5a. Most of the catchments with significant negative values over the Northern Hemisphere and positive ones mainly in lower latitudes do resemble well the water budget error pattern seen in Figure 4-5a.

The water budget imbalance, caused by the increments in LDAS, is only one of the many potential contributing factors to peak river flow errors (and in fact to general river discharge errors); atmospheric forcing biases, imperfect river routing, and observation errors could also lead to large inaccuracies (Zhao et al., 2017).

The impact of the land–atmosphere coupling and LDAS seems to decrease the amount of water overwhelmingly in the rivers (decreased sample mean river discharge, not shown). The sample average river discharge increased only in the southern half of Brazil, in the central part of Canada, and one or two catchments in Africa, East Asia, and South Australia (not shown). It is expected that the decreased average river discharge in ONLINE should generally also result in lower annual peak river flows over most of the globe. Figure 4-9b shows that this decreasing tendency of the annual peaks in the ONLINE experiment coincides with widespread, quite large deterioration in the percentage peak magnitude MAE score (increase of the annual peak magnitude errors) especially in Asia and Europe and the northwestern part of North America, where the majority of the catchments show significant negative bias in Figure 4-9a. On the other hand, quite a few catchments seem to benefit from the coupling and LDAS as the annual peak errors decrease, especially in the western parts in North America, where there is a large cluster of catchments with noticeably smaller percentage peak magnitude MAE.

The river discharge peak timing bias in the ONLINE simulation is dominantly positive (peaks are too late) in the Northern Hemisphere and mainly negative (peaks too early) in the tropics (not shown). However, the coupling and LDAS do not seem to have any systematic impact on this aspect of the peak river flows. There are noticeable differences, but they have no distinguishable geographical pattern (not shown). It seems the short time series (9–25 annual values only) were not sufficient to extract any representative timing differences between the two experiments.

Table 4-4. List of global average scores for the ONLINE and OFFLINE experiments based on the ERA5-D25 dataset. Each value is a mean of scores from 590 catchments (where a minimum of 9 years of river discharge observations was available) weighted by the square root of the catchment area sizes. For further details on the scores, see section 4.2.13. Bold numbers denote the better score of ONLINE and OFFLINE. The following scores are displayed: ME, MAE, NSE, R, percentage sample mean error (PMnE), percentage sample mean absolute error (PMnAe), percentage sample standard deviation error (PStE), percentage sample standard deviation absolute error (PStAe), peak timing ME (PkTiMe), peak timing MAE (PkTiMae), percentage peak magnitude ME (PPkMgMe), and percentage peak magnitude MAE (PPkMgMae).

Score	ME (m^3s^{-1})	MAE (m^3s^{-1})	NSE	R	PMnE (%)	PMnAe (%)	PStE (%)	PStAe (%)	PkTiMe (day)	PkTiMae (day)	PPkMgMe (%)	PPkMgMae (%)
ONLINE	-264	3017	-0.29	0.67	-2.6	29.0	9.6	48.3	-0.95	11.8	6.3	61.3
OFFLINE	236	2954	-0.53	0.70	16.9	27.2	34.2	52.1	-0.81	11.8	27.3	59.2

In addition to the analysis of the annual river discharge peak performance, the general fit between modelled and observed daily river discharge time series is also extensively measured by several scores. Table 4-4 shows a global summary giving an indication on the overall performance of the two experiments. The scores are calculated as global averages weighted by the square root of the catchment size. This way a more representative picture can be provided by giving more emphasis on the larger catchments.

The generally decreasing amount of water leads to larger differences for most of the volume-related bias scores. The percentage sample ME, the percentage sample standard deviation error, and the percentage peak magnitude ME scores all decrease significantly in the ONLINE simulation, bringing the global biases closer to zero. The only exception is the discharge ME score, which changes from a positive value to a negative one with similar magnitude. The better biases, however, do not necessarily help improve the river discharge skill globally; the scores presented in Table 4-4 provide a mixed picture, with some favouring the ONLINE while others favouring the OFFLINE simulation. This agrees with the mixed scores shown in Table 4-3 for the regional example catchments. In general, the MAE, R, the percentage sample MAE, and the percentage peak magnitude MAE values are all slightly better for OFFLINE, while the NSE and percentage sample standard deviation absolute error show improvement for ONLINE. And finally, the peak timing ME is slightly better for the OFFLINE experiment, while there is no difference in the global average peak timing MAE.

4.4 Discussion

In section 4.3, the land-atmosphere coupling and LDAS impact on hydrology, including river discharge and the related water budget variables, was analysed. The river discharge scores showed a mixed picture between the ONLINE and OFFLINE simulations with relatively similar global

performance. Larger differences could be highlighted in certain regions, such as many of the snow-dominant catchments in the Northern Hemisphere, where over many areas a large amount of water is missing from the hydrological cycle and causing downstream issues in river discharge especially during the snowmelt season in ONLINE.

The general decrease in the volume of water in the ONLINE experiment, mainly coming from the snow-dominated areas where the assimilation removes snow, seems to be the primary impact on the hydrology. In soil moisture-dominated areas the river discharge seems to be less impacted by the increments and the evapotranspiration rate holds a more important role.

Data assimilation is a very important component of any NWP system with a lot of effort and research concentrated on the use of observations to correct for random (day-to-day) errors. Data assimilation systems are not there to correct for systematic biases. The fact that LDAS produces consistent negative increments in snow covered areas in this study is pointing toward an apparent snow model bias. In contrast, a model affected by random errors only, would lead to data assimilation increments of both signs with close to zero annual mean values.

Other studies have also highlighted significant snow assimilation impacts on the water balance. For example, De Lannoy et al. (2012) showed that on a small catchment in Colorado (United States) the season averaged snowpack water content is largely decreased by the snow water equivalent assimilation in the Noah land surface model, and could only be overcome by scaling applied (to anomalies) to the observations prior to assimilation. Similarly, Arsenault et al. (2013) found that assimilating MODIS snow cover fraction observations into the CLM land surface model by a simple rule-based direct insertion and the one-dimensional ensemble Kalman filter methods, lead to substantial snowpack removal (without melting, thus causing negative bias in runoff), by both methods in Colorado and Washington.

In the ECMWF system, the snow increments are correcting for the systematic overestimation of the current HTESSEL snow scheme that melts the snow too slowly. Dutra et al. (2012) highlighted that although the current snow scheme provides a significant improvement over the previous one, it does not yet improve on the short-duration melting events during late winter and spring. They argued that the experimental multilayer snow scheme was able to reproduce, at least partially, those snowmelt episodes thanks to the top snow layer having a reduced thermal inertia.

The findings in this work are specific to the NWP configuration at ECMWF with the HTESSEL land surface model and the processes within. However, any LSM's ability to support hydrological simulations can be limited by inadequate handling of the processes, potentially causing a similar problem downstream in the hydrology. The areas highlighted here for ECMWF's HTESSEL in

supporting the flood forecasting activities can be improved by some potential developments in the future. Some of the areas where substantial improvements could be achieved are described:

A new multilayer snow scheme is currently being tested at ECMWF, which is similar to the one evaluated in Dutra et al. (2012). This improved snow scheme is expected to represent better the snowmelt processes and therefore reduce the snow increments that currently remove a significant amount of water from the hydrological cycle. The hydrological context developed in this study will be used to aid this development of the new scheme.

Another potential way of improving HTESEL performance for hydrological applications would be to modify the LDAS by special handling of the snow increments in order to retain the water in the hydrological cycle during the data assimilation. For example, Zaitchik and Rodell (2009) proposed an interesting approach using near-future, snow-covered area observations to adjust the air temperature and precipitation forcing data in order to preserve the local hydrological balance. In another study, Pan and Wood (2006) developed a constrained ensemble Kalman filter method to assure closure of the water balance when assimilating hydrological observations. These types of studies rely on uncoupled systems, and they would be difficult to implement in an operational, real-time environment. However, they provide some insight on water budget closure in data assimilation, and they should be further investigated and adapted to coupled land–atmosphere NWP systems. In the longer term, further coupling between NWP and hydrological forecasting systems will be considered, thereby opening the possibility for coupled land–hydrology data assimilation. In this context, joint assimilation of land surface and river discharge observations will consistently correct the different components of the Earth system.

In addition, the land surface development methodology including data assimilation techniques and process representation is continuously improved at ECMWF. The future inclusion of the LDAS scheme in the offline HTESEL is in development. It will create an environment where the offline research work, including the reanalysis improvements (e.g., ERA5), could be done in a consistent way with the real-time forecast generation. In parallel to these developments, addressing the water budget closure in land–atmosphere data assimilation systems should be a priority in the future to ensure consistent high-quality coupled NWP and hydrological forecasts.

GloFAS is one of the few existing flood forecasting systems that utilizes an LSM (HTESEL) for representing the hydrology (Emerton et al. 2016). Although we acknowledge that in some cases a simple routing model, initialized from observed upstream river levels (either from river gauges or satellite measurements), could be a simpler alternative to simulate downstream discharge on large rivers a few days in advance, for example, in Hossain et al. (2014); in other cases where forecasts

are required further in advance or where observations are unavailable or of too low quality, a more complex modelling configuration, which represents hydrological fluxes, becomes essential. Regardless of some limitations (e.g., the one highlighted in the ECMWF NWP configuration), these complex models play crucial roles in harnessing the available predictability in the land–atmosphere system.

4.5 Conclusions

Understanding the impacts of both the data assimilation and land surface process representation in land surface models on simulated hydrological variables is very important, not only for improving the weather and climate forecasts, but specifically for supporting flood forecasting and other hydrological applications such as drought forecasting, and for giving feedback about the Earth system. In this paper, the influence of land–atmosphere coupling and land data assimilation on global hydrological simulations from LSMs was evaluated. Two river discharge simulations from two climatological reanalyses (based on ERA5) were compared: one operational set which includes land–atmosphere coupling and LDAS with an open water budget, and an offline HTESSSEL set with a closed water budget and no LDAS.

It was found that while the ONLINE version of the model largely improves the 2-m temperature and snow depth conditions, it is causing poor representation of peak river flow in snowmelt-dominated areas, particularly in the high latitudes. However, there are also localized improvements to peak river flow, such as in the western United States. The LDAS increments remove or add water even on an annual average scale which inevitably leads to systematic water budget errors and subsequently contribute to significant errors in river discharge during times of peak flow downstream, something that is critical during times of flooding.

4.5.1 Implications for hydrological forecasting

This study has highlighted the impact of using land data assimilation in reanalysis products. Where data assimilation is adjusting snowpack in forecasting mode then there will also be important implications for hydrological predictions. Future studies should address how far ahead the impact of data assimilation propagates in hydrological forecasts. In addition, hydrological forecasting systems often use initial river conditions derived from climatology. In these circumstances using climatological products derived using data assimilation methodologies could lead to issues with the hydrological forecasts. There are also related issues for forecasting systems such as GloFAS that compare model output to climatology to provide early awareness of extreme events—consistency between operational and climatological configurations goes some way to bypass this problem, and

this conclusion has directly influenced the design of the new GloFAS-seasonal system (Emerton et al., 2018).

4.5.2 Implications for land surface modelling and data assimilation

Data assimilation is designed to compensate for noise errors and not systematic bias. In the case of the current HTESEL snow assimilation scheme it is doing the latter—compensating for system deficiencies such as the slow snowmelt process. This paper has discussed potential ways of addressing water budget deficiencies in land surface approaches, for example, including multiple layers within the HTESEL snow scheme or moving toward data assimilation that conserves the water budget.

Without addressing such issues there will never be confidence in using LSMs for hydrological forecasting applications across the globe. This type of analysis should be used to diagnose where improvements need to be made; considering the whole Earth system in data assimilation and coupling developments is critical for moving toward the goal of holistic Earth system approaches.

Acknowledgements. Ervin Zsoter’s PhD is supported by the Wilkie Calvert Co-Supported PhD Studentships at the University of Reading. Ervin Zsoter was supported by the Copernicus Emergency Management Service - Early Warning Systems [CEMS-EWS (EFAS)]. Hannah Cloke is supported by the TENDERLY project: Towards END-to End flood forecasting and a tool for Real-time catchment susceptibility U.K. NERC Flooding From Intense Rainfall (FFIR) programme, NE/K00896X/1. Elisabeth Stephens and Hannah Cloke are supported by the FATHUM project: Forecasts for Anticipatory HUMANitarian Action funded by U.K. NERC as part of their Science for Humanitarian Emergencies and Resilience (SHEAR) programme, NE/P000525/1. We are also grateful to The Global Runoff Data Centre, 56068 Koblenz, Germany, for providing the observation dataset for our river discharge analysis. Ervin Zsoter designed the experiment, produced the ERA5-D25 datasets, carried out the river discharge data analysis, and led the writing of the manuscript. Hannah Cloke and Liz Stephens assisted with posing the research question and designing the analysis, Patricia de Rosnay and Joaquin Muñoz-Sabater helped with the scientific analysis of data assimilation and coupling issues, and Christel Prudhomme and Florian Pappenberger helped design the research methodology. All authors assisted with writing the manuscript.

Contribution of this chapter to the thesis. This chapter addressed the objective: “Analyse how well the land-surface modelling approach in Earth systems is able to support hydrological applications, in particular focussing on the impact of land-data assimilation of snow and soil moisture on the hydrological cycle in reanalysis simulations.” This work demonstrated that land data assimilation can have a large impact on river discharge, especially in snow dominated areas, which potentially

can create problems in the forecasting system if data assimilation is not used in all components of the modelling chain. This chapter also highlighted some potential limitations of the handling of snow in the land-surface modelling as potential contributing factors to the documented impacts. The next chapter will follow up on this and analyse the hydrological impact of the snow model complexity.

Chapter 5 Hydrological Impact of the New ECMWF Multi-Layer Snow Scheme

Chapter 4 demonstrated, using Earth system reanalysis simulations, that the inclusion of LDAS can create large differences in river discharge, especially in the snow dominated areas. One of the highlighted potential causes was a limitation of the evaluated land-surface modelling system that the simple single layer snow scheme melts snow too slowly. This chapter addresses this area and evaluates how the snow model complexity, through using a new multi-layer snow scheme, can improve the river discharge simulations. This chapter has been published in *Atmosphere* (MDPI) with the following reference (<https://centaur.reading.ac.uk/105034/>):

Zsoter, E., G. Arduini, C. Prudhomme, E Stephens and H. Cloke, 2022: Hydrological Impact of the New ECMWF Multi-Layer Snow Scheme. *Atmosphere*, 13, 727. <https://doi.org/10.3390/atmos13050727>.

The contributions of the authors of this paper are as follows: E.Z. designed the research methodology with the help of H.C., E.S. and C.P. G.A. has designed and produced the land-surface simulations while E.Z. run the CaMa-Flood simulations. E.Z. carried out the data analysis and lead the writing of the manuscript. All authors assisted with writing the manuscript. Overall, 90% of the writing was undertaken by E.Zs.

The published article can be found in the Appendix A2.

Abstract. The representation of snow is a crucial aspect of land-surface modelling, as it has a strong influence on energy and water balances. Snow schemes with multiple layers have been shown to better describe the snowpack evolution and bring improvements to soil freezing and some hydrological processes. In this paper, the wider hydrological impact of the multi-layer snow scheme, implemented in the ECLand model, was analyzed globally on hundreds of catchments. ERA5-forced reanalysis simulations of ECLand were coupled to CaMa-Flood, as the hydrodynamic model to produce river discharge. Different sensitivity experiments were conducted to evaluate the impact of the ECLand snow and soil freezing scheme changes on the terrestrial hydrological processes, with particular focus on permafrost. It was found that the default multi-layer snow scheme can generally improve the river discharge simulation, with the exception of permafrost catchments, where snowmelt-driven floods are largely underestimated, due to the lack of surface runoff. It was also found that appropriate changes in the snow vertical discretization, destructive metamorphism, snow-soil thermal conductivity and soil freeze temperature could lead to large river discharge improvements in permafrost by adjusting the evolution of soil temperature, infiltration and the partitioning between surface and subsurface runoff.

5.1 Introduction

Land-surface models (LSMs) are vital tools for simulating water and energy fluxes at the land–atmosphere interface of the Earth (Fisher and Koven, 2020). Although LSMs were originally designed to provide lower-boundary conditions to the atmosphere (Mengelkamp et al., 2001), with the improving realism of these models they are increasingly used for simulating the hydrological cycle (Prentice et al., 2015) and supporting hydrological applications (e.g. Bouilloud et al., 2010; Le Vine et al., 2016; Harrigan et al., 2020b).

However, there are still significant limitations in the representation of the hydrological cycle in LSMs, as important processes can still be inadequately modelled or even neglected for runoff generation, for instance groundwater simulation, snow-vegetation interactions, representation of frozen soil and lateral flow between adjacent grid cells among others (Fisher and Koven, 2020; Le Vine et al., 2016; Overgaard et al., 2006; Dutra et al., 2012; Kauffeldt et al., 2015; Koren et al., 2014; Krogh et al., 2017).

Simulating the extent and variability of the snow cover is a crucial aspect of land surface modelling, as it strongly influences the energy and water balances (López-Moreno and García-Ruiz, 2004; Griessinger et al., 2016). Snow schemes have various complexities in the representation of the snow physics (Boone et al., 2001; Best et al., 2011; Decharme et al., 2016; Dutra et al., 2010; Wang et al., 2013), differing largely in their handling of the snowpack and the creation of snowmelt, which in turn impacts runoff generation and river flow in snow dominated areas (Dutra et al., 2012; Slater et al., 2011).

The snow scheme currently used operationally in the ECLand land-surface model at the European Centre for Medium-Range Weather Forecasts (ECMWF) (Balsamo et al., 2009; Boussetta et al., 2021) and which is used in the production of the ERA5 (Hersbach et al., 2020) and ERA5-Land (Muñoz-Sabater et al., 2021) reanalysis datasets, is a single-layer snow scheme (SLS hereafter) with an additional snow layer on top of the soil (Dutra et al., 2010). The use of only one layer limits the handling of the temporal evolution of the snow, as changes on multiple time scales (i.e., diurnal to seasonal) cannot be accurately represented. This has a significant impact on the quality of the snow depth, mainly during periods of accumulation and ablation, which then impacts the soil freezing, the snowmelt and ultimately the hydrological cycle (Dutra et al., 2012; Saha et al., 2017).

Snow schemes using multiple layers to represent the snowpack offer significant improvement on the single-layer schemes with better handling of the snow processes. For hydrological and climate applications, so-called “intermediate complexity” snow schemes are generally used, following the terminology introduced by Boone and Etchevers (2001). Such schemes include a description of

snowpack properties, such as density and temperature, using a limited number of vertical layers, as opposed to detailed snow physics models, which aim at simulating the microstructure properties of the snowpack as well (see for instance Vionnet et al. (2012)). Such intermediate complexity schemes have been implemented in various LSMs used in Earth system modelling during the last decade. Examples are ECLand (Dutra et al., 2012; Arduini et al., 2019), Noah (Saha et al., 2017), JULES (Walters et al., 2019), ISBA (Boone and Etchevers, 2001; Decharme et al., 2016) and the ORCHIDEE (Wang et al., 2013) land-surface models.

The multi-layer snow scheme (MLS hereafter), introduced experimentally in ECLand, is an intermediate complexity scheme representing the vertical structure and evolution of snow temperature, density, liquid water content and surface snow albedo with a maximum of five layers (Arduini et al., 2019). It has been shown to increase the realism of snow representation, including decreasing snow depth and snowmelt timing errors, and has been shown to largely improve 2-metre temperature in coupled forecasts, especially in clear sky conditions (Arduini et al., 2019).

The more realistic snowpack representation, with better snow water equivalent, snow depth and snowmelt, is expected to improve the hydrological cycle and thus have a positive impact on river flow simulations (Magnusson et al., 2015). Preliminary studies have demonstrated improvements in localized settings (Dutra et al., 2012; Wang et al., 2013) or combined with other land-surface improvements (Decharme et al., 2019). However, the hydrological impact analysis of MLSs, focusing on river flow, has not been done at regional or global scales.

Areas in which snow plays an important role are predominantly found in the Northern Hemisphere over higher latitudes. A large fraction of these areas is permafrost (Romanovsky et al., 2002), where soil freezing/thawing conditions play a major role in controlling the hydrological processes (Koven et al., 2013; Andresen et al., 2020). Representing permafrost in LSMs is important for better understanding of the hydrological variability and the impacts of climate change (Andresen et al., 2020; Yokohata et al., 2020; Gouttevin et al., 2012).

In this study, the hydrological impact of the MLS implemented in ECLand is analysed on more than 400 catchments globally, with over a third located partially or entirely in permafrost areas. To achieve this, ECLand experiments forced with ERA5 over the period 1979–2018 are coupled to the Catchment-based Macro-scale Floodplain model (CaMa-Flood; Yamazaki et al., 2011) to generate river discharge, allowing direct comparison with gauged observations. Different sensitivity experiments are conducted to evaluate the impact of the more physically complex snow scheme on the terrestrial hydrological processes, with particular focus on permafrost, where complicated error dynamics arise from different land-surface processes. Two main questions are posed:

- How does the MLS impact the simulated hydrological processes and river discharge, especially in the snowmelt-driven flood season?
- How sensitive is the hydrological representation of permafrost to the snow and soil parametrization?

5.2 Materials and methods

In this section, the data set, models and methods used will be described.

5.2.1 ECLand land-surface model and offline methodology

The hydrological core of the analysed data sets was provided by the ECLand land surface model, formerly known as HTESSSEL (The Hydrology-Tiled ECMWF Scheme for Surface Exchange over Land (Balsamo et al., 2009; Boussetta et al., 2021; Balsamo et al., 2011)). ECLand is part of the Integrated Forecasting System (IFS) at ECMWF and used in coupled land-atmosphere simulations for describing the evolution of soil, vegetation and snow conditions over land, at various spatial resolutions, from short- to seasonal-range.

In ECLand, up to six tiles are present over land (bare ground, low and high vegetation, intercepted water, shaded and exposed snow) and three over water (inland, open and frozen water) that provide the interface between the atmosphere and the one-dimensional soil column, with all tiles having their separate energy and water balances.

The snowfall (the solid fraction of precipitation) is collected in the snowpack, which overlays the soil (Balsamo et al., 2009). The fraction of the soil that is covered by snow (snow cover fraction) is parametrized as a linear function of snow depth (Dutra et al., 2010). This assumes that a model grid-box is fully covered for snow depth greater than 10 cm. The same parametrization is used for exposed and shaded snow (i.e., snow under high vegetation) tiles in ECLand. The soil is divided into four layers with fixed layer depths (0–7, 7–28, 28–100 and 100–289 cm). Runoff is generated as fast (surface) and slow (subsurface) components at each grid point (Balsamo et al., 2009; Boussetta et al., 2021).

Snowmelt occurs when the temperature of the snow is high enough, contributing to surface runoff, soil infiltration and evaporation. Some part of rain and snowmelt will be removed as surface runoff. This surface runoff fraction depends on the standard deviation of the sub-grid scale orography (a measure of unresolved orographic features), the soil texture and the soil water content. Subsurface runoff is the water leaving the soil column at the bottom. It depends on the infiltration and surface evaporation as top boundary conditions, while water can be extracted by roots in each soil layer where vegetation is present (Balsamo et al., 2009; Boussetta et al., 2021).

ECLand can be used in a stand-alone mode, when the model runs uncoupled from the atmosphere, usually with hourly time step, forced with near-surface meteorological input data of temperature, specific humidity, wind speed, surface pressure, radiative fluxes (downward solar and thermal radiation) and water fluxes (liquid and solid precipitation), without land data assimilation. This offline research methodology provides an affordable way of testing land-surface improvements and has been used in various applications (e.g. Arduini et al., 2019; Agustí-Panareda et al., 2010; Haddeland et al., 2011; Zsoter et al., 2019). The ERA5-Land dataset is a prime example of this methodology, which was produced as an offline ECLand simulation with downscaled ERA5 meteorological forcing on higher resolution, including an elevation correction for the thermodynamic near-surface state (Muñoz-Sabater et al., 2021).

5.2.2 ERA5 reanalysis

The meteorological forcing for the offline ECLand simulations was taken from ERA5, the latest global climate reanalysis of ECMWF (Hersbach et al., 2020). ERA5 is a key contribution to the EU funded Copernicus Climate Change Service (C3S) and is open access and free to download for all users (<https://cds.climate.copernicus.eu/> (accessed on 24 October 2021)). It covers the period 1979 to present, with a preliminary version also available from 1950. It includes a high-resolution component (~31 km) and a lower resolution (~62 km) ensemble component with 10 members. In this study, the high-resolution component (hereafter referred to as ERA5) was used from 1979 with ~31km horizontal resolution and hourly output frequency.

5.2.3 CaMa-Flood river-routing

The hydrodynamics to produce river discharge from the ECLand runoff output were simulated by CaMa-Flood (Yamazaki et al., 2011), a global river-routing model, which is part of ECLand since IFS cycle 47r1 (Boussetta et al., 2021). CaMa-Flood routes runoff generated by land-surface models to oceans or inland seas. The model calculates river and floodplain water storages, discharge, water depth, as well as flood inundation. CaMa-Flood does not currently include the representation of dams and permanent lakes and wetlands are only treated as part of the floodplain storages. CaMa-Flood is computationally cheap to run and has been used widely in global climatological research studies, such as Emerton et al. (2017), Dottori et al. (2018) and Zsoter et al. (2019).

5.2.4 ECLand snow and soil freezing schemes

The current SLS, used operationally in ECLand, is a basic energy balance model describing the temporal evolution of the heat and mass contents of the snowpack (Dutra et al., 2010). The MLS, used experimentally in ECLand, is an intermediate complexity snow scheme (Boone and Etchevers, 2001), which represents the vertical structure and time evolution of snow temperature, density,

liquid water content and surface snow albedo with up to five active snow layers. In this section, the important model features are described that are necessary to understand the scheme variations tested in this paper (described later in Section 5.2.8). Further details of MLS, including a detailed comparison of snowpack properties to SLS, are given in Arduini et al. (2019).

5.2.4.1 Snow vertical discretization

The number of active snow layers and their associated thicknesses are defined diagnostically at each time step before the updating of the other snow variables. The number of active layers (N) is dependent on the snow depth D_{sn} and varies from one layer to a maximum of five (N_{max}). The topmost snow layer, in contact with the atmosphere, is assumed to be the first one, whereas layer N is the one in contact with the soil. N is determined as the lowest number that satisfies the following inequality for $N = 1, \dots, N_{max} - 1$:

$$\sum_{j=1}^{N+1} D_{min,j} > D_{sn}, \quad (5-1)$$

where $D_{min,j}$ is the minimum snow depth allowed for layer j . It is by default set to 0.05 m for all layers (denoted hereafter as D_{min} for all layers). The depth of the first layer is defined as:

$$D_{sn,1} = \begin{cases} D_{sn}, & \text{if } D_{sn} < 2D_{min} \\ D_{min}, & \text{if } D_{sn} > 2D_{min} \end{cases}. \quad (5-2)$$

Note that with this choice, MLS has only one active snow layer for $D_{sn} < 0.1$ m. For the remaining layers, the vertical discretization is defined as:

$$D_{sn,i=2,\dots,N_{max}} = \begin{cases} 0, & \text{if } D_{sn} < \sum_{j=1}^{j=i} D_{min,j} \\ \min \left[\frac{D_{sn} - D_{sn,1}}{N - 1}, D_{max,i} \right], & \text{if } D_{sn} > \sum_{j=1}^{j=i} D_{min,j} \end{cases}, \quad (5-3)$$

where $D_{max,i}$ is the maximum snow depth allowed for the i -th active snow layer. This effectively means, the snow is evenly divided into the remaining layers, as long as the maximum layer depths allow it. By default, these maximum values are 0.05, 0.10, 0.20, ∞ and 0.15 for layers $1 - N_{max}$, respectively. This definition of the maximum layer depths means that when all N_{max} snow layers are active, the $N_{max} - 1$ layer is used as the accumulation layer for thick snowpacks. This layering allows a relatively high vertical resolution both at the interfaces to the atmosphere above and to the soil underneath the snowpack. Take $D_{sn} = 1.25$ m as an example, for this depth the snowpack is discretized into 5 layers with thicknesses of 0.05, 0.1, 0.20, 0.75 and 0.15 m from top to bottom.

5.2.4.2 Destructive metamorphism of the snow

The density of freshly fallen snow can vary rapidly with time due to metamorphic processes, that is, the change in shape and size of snow grains once they settle in the snowpack. The rate of change

of snow density due to destructive metamorphic processes of the snow (ξ) is parametrized in both SLS and MLS using the formulation introduced by Anderson (1976):

$$\frac{1}{\rho_{sn,i}} \left[\frac{\partial \rho_{sn,i}}{\partial t} \right]_{\xi} = a_{\xi} \exp[-b_{\xi}(T_f - T_{sn,i}) - c_{\xi} \max(0, \rho_{sn,i} - \rho_{\xi})], \quad (5-4)$$

where $T_{sn,i}$ and $\rho_{sn,i}$ are the snow temperature and snow density for each snow layer (for SLS, $i = 1$) and $a_{\xi} = 2.8 * 10^{-6} s^{-1}$, $b_{\xi} = 4.2 * 10^{-2} K^{-1}$, $c_{\xi} = 460 m^3 kg^{-1}$, $T_f = 273.16 K$ and $\rho_{\xi} = 150 kg m^{-3}$.

5.2.4.3 Snow-soil thermal conductivity

In SLS and MLS, the thermal coupling between the snowpack and the soil underneath is described using a thermal conductance between the two media (λ_b), thus the heat flux (G_b) can be written as:

$$G_b = \lambda_b (T_{sn,N} - T_{so}), \quad (5-5)$$

where $T_{sn,N}$ is the snow temperature of the bottom active snow layer ($N = 1$ for SLS) and T_{so} is the temperature of the topmost soil layer. Given that the heat resistances are in series, the sum of the inverse of the conductance of each medium yields the total conductance of the snow-soil system, that is:

$$\lambda_b^{-1} = \frac{l_b \Delta z_{sn,N}}{\lambda_{sn,N}} + \frac{l_b \Delta z_{so}}{\lambda_{so}}, \quad (5-6)$$

where $\Delta z_{sn,N}$ and $\lambda_{sn,N}$ are the thickness and conductivity, respectively, of the bottom active snow layer, and Δz_{so} and λ_{so} are the thickness and conductivity of the topmost soil layer. The parameter l_b is set as 0.5 in SLS, whereas it was changed to 1 in MLS, as described in Arduini et al. (2019), to account for the additional insulation effect due to organic material or vegetation between the snow layer and the soil.

5.2.4.4 Soil freezing scheme and relationship to runoff generation

Frozen soil is characterized by very different thermal and hydrological properties compared to unfrozen (wet) soil. For instance, precipitation infiltrates less into a frozen soil, thus more water goes into runoff than into the soil. In ECLand, a simplified representation of unfrozen soil was introduced by Viterbo et al. (1999) to account for the latent heat release/absorption of soil water around 0 °C, reducing a pronounced cold bias in 2-metre temperature in the ECMWF forecasts. In this simplified approach, the frozen water fraction in each soil layer ($\theta_{ice,i}$, $i = 1, \dots, 4$) is given as a function of the soil temperature ($T_{so,i}$) as:

$$\theta_{ice,i} = \begin{cases} 0 & \text{for } T_{so,i} > T_{th} \\ 0.5 \left(1 - \sin \left(\frac{\pi (T_{so,i} - 0.5 (T_{Th} + T_{Fr}))}{T_{Th} - T_{Fr}} \right) \right) & \text{for } T_{Fr} \leq T_{so,i} \leq T_{Th} \\ 1 & \text{for } T_{so,i} < T_{Fr} \end{cases} \quad (5-7)$$

where $T_{Th} = +1$ °C (thaw temperature) and $T_{Fr} = -3$ °C (freeze temperature) are the soil temperature thresholds for which soil water is totally unfrozen or frozen, respectively, and $[T_{Th}, T_{Fr}]$ is the temperature interval for which phase change can occur. Both SLS and MLS use this soil freezing scheme.

The runoff generation in EC-Land follows the formulation described in Balsamo et al. (2009), which includes a dependency of surface runoff on soil textures as well as subgrid scale orography features not resolved at the resolution used in the simulation. The surface runoff can be as large as 50% of the available precipitation and snowmelt, for large standard deviation of the sub-grid scale orography and finer soil textures.

When the soil is partially frozen, the surface runoff is enhanced, as less water infiltrates and percolates within the soil column. This soil freezing mechanism is represented by reducing the soil hydraulic conductivity and diffusivity. This is done by computing the soil hydraulic conductivity and diffusivity as a weighted average of the values for the unfrozen soil water fraction and for the frozen water fraction (Balsamo et al., 2009).

As previously stated, the frozen water fraction parametrization described in Equation (5-7) has been developed to address temperature errors in weather forecasting applications. Such simplified approach can have limitations for hydrological applications in cold regions, where the interaction between frozen soil and runoff is key in modulating river streamflow (see e.g. Niu and Yang 2006).

5.2.5 River catchment selection

The river catchments were selected for this study only if they experience regular snowfall and they have adequate river discharge observations available. Observations are selected from the Global Runoff Data Centre (GRDC; <https://www.bafg.de/GRDC/> (accessed on 20 November 2019)) supplemented by additional data collected by the Copernicus Emergency Management Service for floods in the Global flood Awareness System (GloFAS), as described in Harrigan et al. (2020b). The catchments and associated river discharge observations are selected with the following set of criteria:

- Minimum 8 years of river discharge observations in 1980–2018. Gaps are not considered a problem, as long as the climatological mean can be computed for each day of the year (see Section 5.2.7);

- Stations in snow impacted climate, defined by the percentage ratio of ERA5 snowfall and total precipitation being at least 10%, based on the 1979–2018 mean for each catchment;
- Catchment area of at least 5000 km² (e.g., minimum of 8 river pixels);
- Good general quality. After visual inspection of the river discharge time series, the catchments that showed observation errors, problems with station metadata (wrong or uncertain location, etc.) or visible influence of dams and lakes were excluded. To help with identifying reservoir and lake influence, the Global Reservoir and Dam Database (GRAND; Lehner et al., 2011) and the Global Lakes and Wetlands Database (GLWD; Lehner and Döll, 2004)) were used as visual tools.

Out of the 2119 stations in the Copernicus Emergency Management Service GloFAS station database with at least 1 year of observations, 1913 met the criteria of at least 8 years of data, 889 catchments had also at least 10% snowfall ratio and 849 were additionally over 5000 km² area.

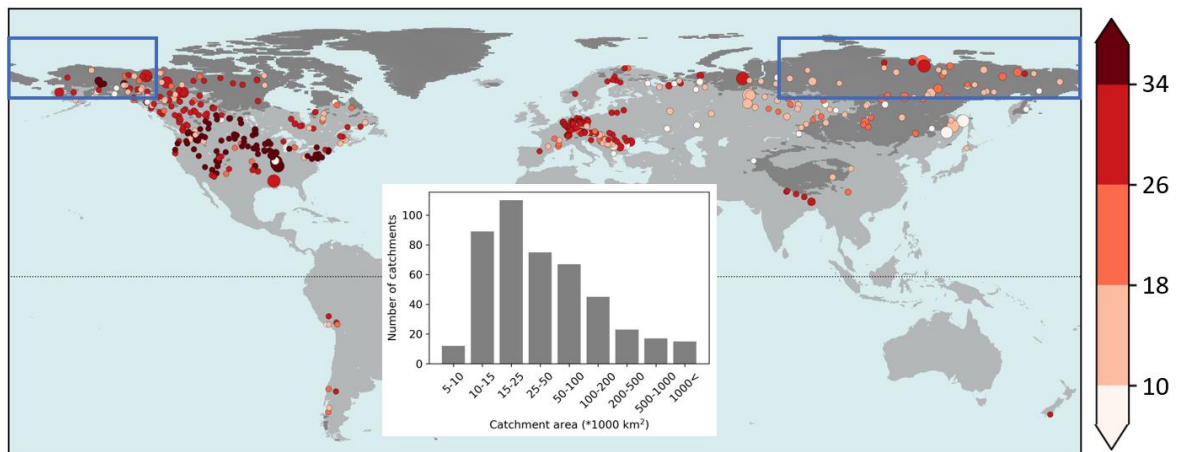


Figure 5-1. Stations used in this study with the number of river discharge observation years available (8 years is the minimum). In total, 453 catchments worldwide. The darker grey shading indicates the permafrost areas (defined as the area where the lowest soil layer's temperature is below 0 °C in the ERA5 climate mean on 1st of June, based on 1980-2018), while the blue rectangle shows the sensitivity area defined in the permafrost. The distribution of the catchment upstream area values is provided in the inset table (please note the area values are divided by 1000).

In total, 453 catchments were selected after the quality checks, almost entirely in the Northern Hemisphere (Figure 5-1). For the sensitivity analysis in the permafrost, a specific area was defined in the 60–80N belt containing parts of northern Siberia, Alaska and western Canada. It focuses on the coldest parts of the permafrost. The eastern area of Canada was omitted due to the large number of lakes in the area to avoid unduly influencing of the overall results. Figure 5-1 shows the bounding box, together with the permafrost, as defined by areas of below 0 °C climatological mean temperature in the lowest soil layer of ECLand in ERA5 for 1st of June. Within ECLand, the land use of the catchments in the permafrost sensitivity area are mainly a mixture of tundra and boreal forest vegetation, whereas the soil textures are medium-coarse and coarse soil types, which are

characterized by a relatively low field capacity and higher hydraulic conductivity (for details on land use and soil hydraulic properties see Balsamo et al., 2009 and Boussetta et al., 2021).

5.2.6 Verification statistics

Hydrological performance is assessed with the 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). It can be decomposed into three components, measuring the correlation, bias and variability errors, which makes it an easy to interpret metric, ideal for assessing hydrological dynamics:

$$kge = 1 - \sqrt{(pcorr - 1)^2 + \left(\frac{\mu_s}{\mu_o} - 1\right)^2 + \left(\frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} - 1\right)^2}. \quad (5-8)$$

In the *kge* decomposition, *pcorr* is the Pearson correlation coefficient between daily simulation (*s*) and observation (*o*) time series, measuring the temporal errors; μ is the mean and σ is the standard deviation of the time series. In Eq (5-8), $\frac{\mu_s}{\mu_o}$ is the bias ratio, while $\frac{\sigma_s/\mu_s}{\sigma_o/\mu_o}$ is the variability ratio, which highlight how close the mean and the variability (normalized by the means) are in the simulated and observed time series. The *kge* and its three components are all dimensionless. The correlation ranges from -1 to +1, with +1 showing perfectly strong linear relationship and 0 no linear relationship (-1 being perfect inverse relationship). In the *kge* definition, the bias and variability ratios both range from 0 to infinity, with 1 being the optimal value. The bias and variability errors, used in this study, were defined by the bias and variability ratio components of the *kge* as follows:

$$bias = \frac{\mu_s}{\mu_o} - 1 \text{ and } var = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} - 1, \quad (5-9)$$

This way, the optimal score value transforms to 0, highlighting the direction of the biases more intuitively, with negative values showing underprediction, while positive ones overprediction. To aid comparison between different experiments, the absolute version of the *bias* (*abias*) and *var* (*avar*) errors are also used. Change in these metrics can directly highlight improvement or deterioration, while they also share the optimal value of zero.

5.2.7 Daily climatology computation

The land-surface contribution to the water budget is diagnosed qualitatively at specific catchments by using daily climatological mean time series of simulated and observed river discharge, runoff (surface and subsurface components), snowpack water content and snowmelt, evaporation, soil temperature and water content (at different layers) and precipitation, computed from daily values

in the 1980-2018 period. All water related variables are converted into catchment totals as the sum of all the grid point values in the catchment, in order to compare them directly to river discharge. For temperature variables, the catchment averages are used.

These climatological means are computed for every day of the year (1 January - 31 December), by applying a 21-day window, centred over the day of the year. To aid direct comparison with observed river discharge, only those days of 1980-2018 are considered for the climatological mean computation, which have river discharge observations available. For the climatological mean, the minimum data length was lowered to 4 years, in order to maximize the likelihood of being able to compute the mean for every day of the year. With this choice, the climate sample size could range from 84 values (4 years, 21 values each) to over 800 (most years in 1980-2018) to compute the daily climate mean.

5.2.8 *Experimental setup*

In this study, the hydrological impact of the MLS is analysed on ECLand/CaMa-Flood coupled experiments. The runoff is produced by ECLand, while the CaMa-Flood model is used to produce river discharge by routing the runoff over the 15 arcmin (~25 km on the Equator) river network, which is an appropriate horizontal resolution for the related meteorological forcing of ERA5.

CaMa-Flood uses a 1-h time step and a 24-h output frequency to match the 24-h reporting frequency of the river discharge observations. All experiments are generated for the ERA5 period of 1979-2018, while for the hydrological analysis 1980-2018 is used with 1979 omitted to account for the spin-up in the simulations. Preliminary analysis showed that a 1-year spin up period is appropriate, as a longer spin up period did not have a large influence on results but considerably reduced the sample size.

In total, 13 experiments are produced and compared (Figure 5-2). Two experiments use SLS, while the other 11 are produced with variations of MLS and the soil freezing parametrization in ECLand, for analysing the hydrological sensitivity, focusing on permafrost areas. The very high computational cost of running these experiments meant that it was not possible to run all permutations of the schemes and parameters tested, instead possible modifications build on each other incrementally.

Single-layer, online, fully coupled with land data assimilation: SL-CDS

The first experiment involves a single CaMa-Flood run, using the original ERA5 runoff data (downloaded from the Copernicus Climate Data Store), which is produced online with land-atmosphere coupling, including atmospheric and land data assimilation and SLS (SL-CDS).

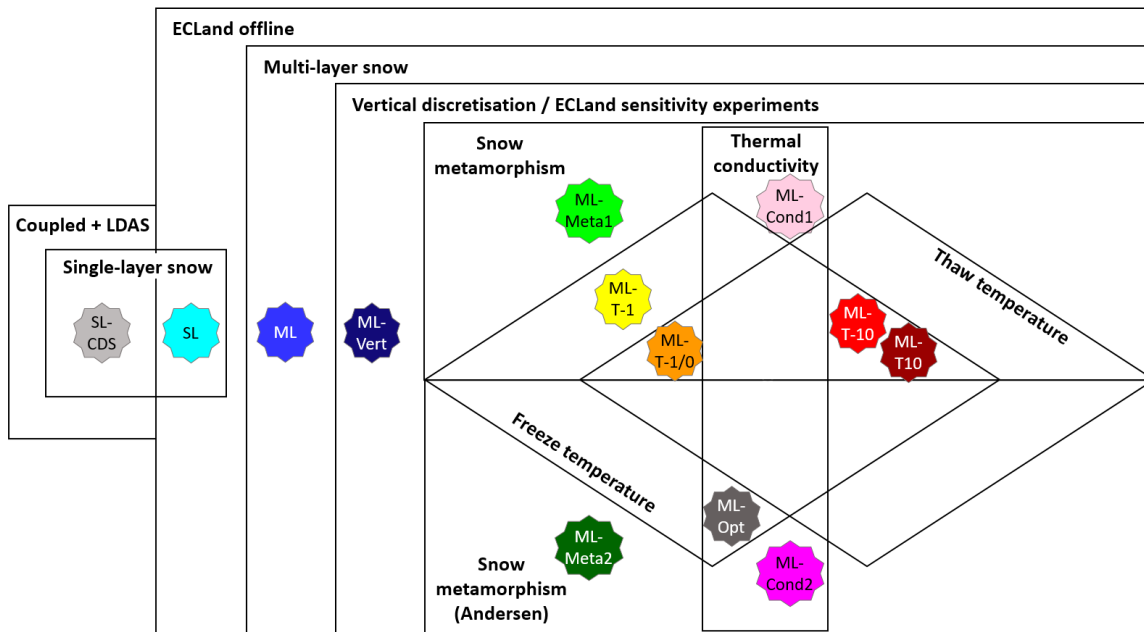


Figure 5-2. Experiments analysed in this study, describing the main simulation features including the snow and soil scheme modifications for sensitivity analysis in permafrost. The experiments are displayed with the star shapes, using a short name and a colour to help identifying them throughout the paper.

Offline experiments

All other experiments that follow include a surface only (offline) ECLand simulation, without land-atmosphere coupling and land data assimilation, to produce runoff, which is then routed with CaMa-Flood. This is because online experiments with coupling and data assimilation would be infeasible to run due to the very high computational cost. The offline experiments are initialized from the ERA5 state on 1 January 1979 and forced with ERA5 near-surface meteorological data on ~31 km horizontal resolution and hourly output frequency (see Section 5.2.1 for further details).

Single-layer snow scheme: SL

In order to compare the online and offline modelling approaches directly, one of the offline experiments is run with the SLS (SL), to be compared with SL-CDS (see Zsoter et al. (2019) for further details on online/offline comparison).

Multi-layer snow scheme: ML

The first MLS experiment uses the default snow parametrization, introduced by Arduini et al. (2019) and default ECLand soil freezing parametrizations (see Section 5.2.4) and can be considered the default multi-layer experiment (ML).

ECLand sensitivity experiments

The following 10 experiments are variations of the ECLand snow and soil freezing schemes, designed to evaluate the river discharge sensitivity in permafrost.

Vertical snow discretization: ML-Vert

The first change is for the vertical snow discretization in MLS (ML-Vert). It introduces thicker snow layers over complex terrains with deep snowpack and reduces issues related with excessive melting over mountainous regions (Boussetta et al., 2021). Complex terrain is defined as areas with standard deviation of the sub-grid scale orography over 50 m. For flat terrain and for $D_{sn} < 0.25$ m over complex terrain, the same vertical discretisation is used as in ML (described previously in Section 5.2.4). For $D_{sn} > 0.25$ m over complex terrain, $D_{max,i}$ and $D_{min,i}$ can vary with snow depth in Eq (5-1)-(5-3) as follows:

$$D_{min,i} = \min(0.25, 0.10 + \alpha_0(D_{sn} - 0.25)) \text{ for } i = 1 \quad (5-10)$$

$$D_{max,i} = \begin{cases} \min(0.25, 0.10 + \alpha_0(D_{sn} - 0.25)) & \text{for } i = 1 \\ \min(0.30, 0.15 + \alpha_0(D_{sn} - 0.25)) & \text{for } i = 2, \dots, N_{max} \end{cases}, \quad (5-11)$$

where $\alpha_0 = 0.1$ is a predefined parameter. For the example of $D_{sn} = 1.25$ m (as in Section 5.2.4), the thickness of the 5 layers changes to 0.20, 0.25, 0.25, 0.30 and 0.25 m, which means the snowpack is more evenly distributed for this large depth.

Destructive metamorphism of the snow: ML-Meta1 and ML-Meta2

The second group of changes is for the snow metamorphism. The parametrization for the destructive metamorphism of the snow (Eq (5-4)), as used in SL, SL-CDS and ML with the default value of $c_\xi = 460 \text{ m}^3 \text{ kg}^{-1}$, implies that the rate of snow density changes due to the 2nd density-dependent term in the exponential, and is active only for $\rho_\xi < 150 \text{ kg m}^{-3}$, that is for relatively fresh snowpack. Cao et al. (2020) pointed out, while evaluating soil temperature and snow characteristics of ERA5-Land in permafrost regions, that this can partly explain the underestimation of the snow density of ERA5-Land. They have argued that the underestimation of the snow density could lead to an overestimation of the thermal decoupling between the atmosphere and the soil underneath. This could contribute to the warm bias of soil temperature over permafrost regions as less heat is diffused from the soil towards the colder atmosphere above.

To address this, the impact of the representation of the snow density on river discharge is explored in two experiments, by changing the value of the parameter c_ξ . In the first experiment (ML-Meta1), the parameter c_ξ is varied for the five snow layers, using values closer to the $0.046 \text{ m}^3 \text{ kg}^{-1}$, reported in Anderson (1976), as follows:

$$c_\xi = (0.112, 0.152, 0.192, 0.288, 0.488) \text{ m}^3 \text{ kg}^{-1}. \quad (5-12)$$

These values are chosen as they were the best compromise in terms of land-surface and atmospheric impact (i.e., in particular 2-metre temperature) in coupled land-atmosphere forecast experiments, which were conducted for the foreseen implementation of the MLS in operational

weather forecasts at ECMWF. The decreasing values towards the first (top) snow layer imply that the destructive metamorphic process is more active for the top layer, whereas for the settled snow at the bottom of the snowpack it is less active, which makes the snow compaction process slower.

Another experiment (ML-Meta2) uses the original value of $c_{\xi} = 460 \text{ m}^3 \text{ kg}^{-1}$ for all snow layers, as reported by Anderson (1976), which is a commonly used approximation in land-surface models. This latter unified and low value option of c_{ξ} gives the opportunity to explore the maximum sensitivity to the destructive metamorphic process of the snow in the snow density representation, without considering the implication for the snow-atmosphere coupling (e.g., atmospheric scores), given that the feedback to the atmosphere is not considered in the offline experiments presented in this work. In coupled experiments, the impact of this change would be expected to be substantial on the near surface variables, like 2 m temperature, possibly requiring an additional tuning of other land-atmosphere coupling parameters. Both experiments build on ML-Vert by adding the snow metamorphism corrections to the vertical snow depth discretization adjustment.

Snow-soil thermal conductivity: ML-Cond1 and ML-Cond2

The next two experiments use the parameter $l_b = 0.5$ in the snow-soil thermal conductivity computation (Eq (5-6)), the same value as in SLS and reduced by half compared with MLS. Organic material distribution is highly variable both in horizontal and vertical (within the soil) scales and its handling requires more sophisticated parametrizations (Decharme et al., 2016).

For this reason, in these two experiments we relax the hypothesis of $l_b = 1$ in MLS, removing the additional thermal insulation effects caused by organic material, effectively considering half of the topmost soil layer in the computation of the conductance between the two media (see Eq (5-6)), consistently with what is done for the other snow-free land-surface tiles of ECLand. Effectively, this increases the thermal coupling between the snow and soil and thus also increases the heat flux between the two media in these experiments. Both experiments include the vertical discretization change (ML-Vert), while ML-Cond1 is run together with the first snow metamorphism change in ML-Meta1 and ML-Cond2 with the change in ML-Meta2.

Soil freeze and thaw temperatures: ML-T-1, ML-T-1/0, ML-T10 and ML-T-10

The next step in the experiment design is to test the sensitivity of river discharge simulations to the fraction of frozen water in the soil. Soil freezing and thawing is particularly important during the snowmelt season, when the thermal coupling of the topmost soil layers to the atmosphere increases quickly as the snow melts and phase changes of the soil water can occur. Out of the two temperature parameters, the freeze temperature has a more direct impact on runoff generation during the spring period, whereas the thaw temperature could be more important for energy fluxes

and the coupling to the atmosphere in the fall/winter season. Gouttevin et al. (2012) suggested that a shorter temperature interval for phase change works better for permafrost.

These soil temperature experiments are run including the changes in ML-Vert and ML-Meta1. First, while keeping the T_{Th} at the default value of +1 °C in Eq (5-7), the T_{Fr} parameter is increased from the default -3 to -1 °C (ML-T-1). Similarly, in another experiment, while keeping T_{Fr} at the higher level of -1 °C, T_{Th} is decreased from the default +1 to 0 °C (ML-T-1/0).

In addition, the extreme boundaries of the soil freezing contribution to river discharge are also tested, by setting the $[T_{Fr}, T_{Th}]$ phase change interval unreasonably high at [+10, +10.5] (ML-T10) and unreasonably low at [-10.5, -10] (ML-T-10). These extreme temperature thresholds allow the soil to remain almost always or never frozen, which consequently should increase or decrease the amount of infiltration to the soil and thus the amount of surface runoff to an extreme level.

Optimal combination: ML-Opt

Finally, a prospective experiment is defined with combinations of some evaluated changes that are expected to work best for permafrost. ML-Opt combines incremental changes in ML-Vert, ML-Meta2, ML-Cond2 and ML-T-1 into one experiment. This experiment aims at exploring the interactions among the proposed changes, as feedbacks between different processes can be highly non-linear. For instance, testing the proposed changes in combination can indicate if singular modifications that improve the river discharge simulation, actually (over-) compensate for other sources of errors.

5.3 Results

5.3.1 *Default multi-layer vs. single-layer snow schemes*

The default MLS generally improves on the SLS, mainly through better bias and variability, with the exception of some parts of the permafrost, where the multi-layer simulation is suboptimal.

The impact of the MLS is analysed first in this section by comparing the default parametrization option ML with the single-layer SL (Figure 5-3). The ML improves the river flow predictions for the majority of stations over the midlatitudes (about two thirds of them), with a larger cluster of improved catchments (with higher *kge*) present in western/central North America (Figure 5-3a). The *kge* mean, computed across 453 catchments, is 0.43 for ML while 0.40 for SL, highlighting a small overall improvement. However, in some higher latitude areas, especially in the northern half of Siberia in Asia and also near Alaska in North America (coinciding with the blue box permafrost sensitivity area in Figure 5-3), the river discharge performance is deteriorated in ML, with many catchments showing a drop of at least 0.05–0.1 (some even above 0.3) in *kge*.

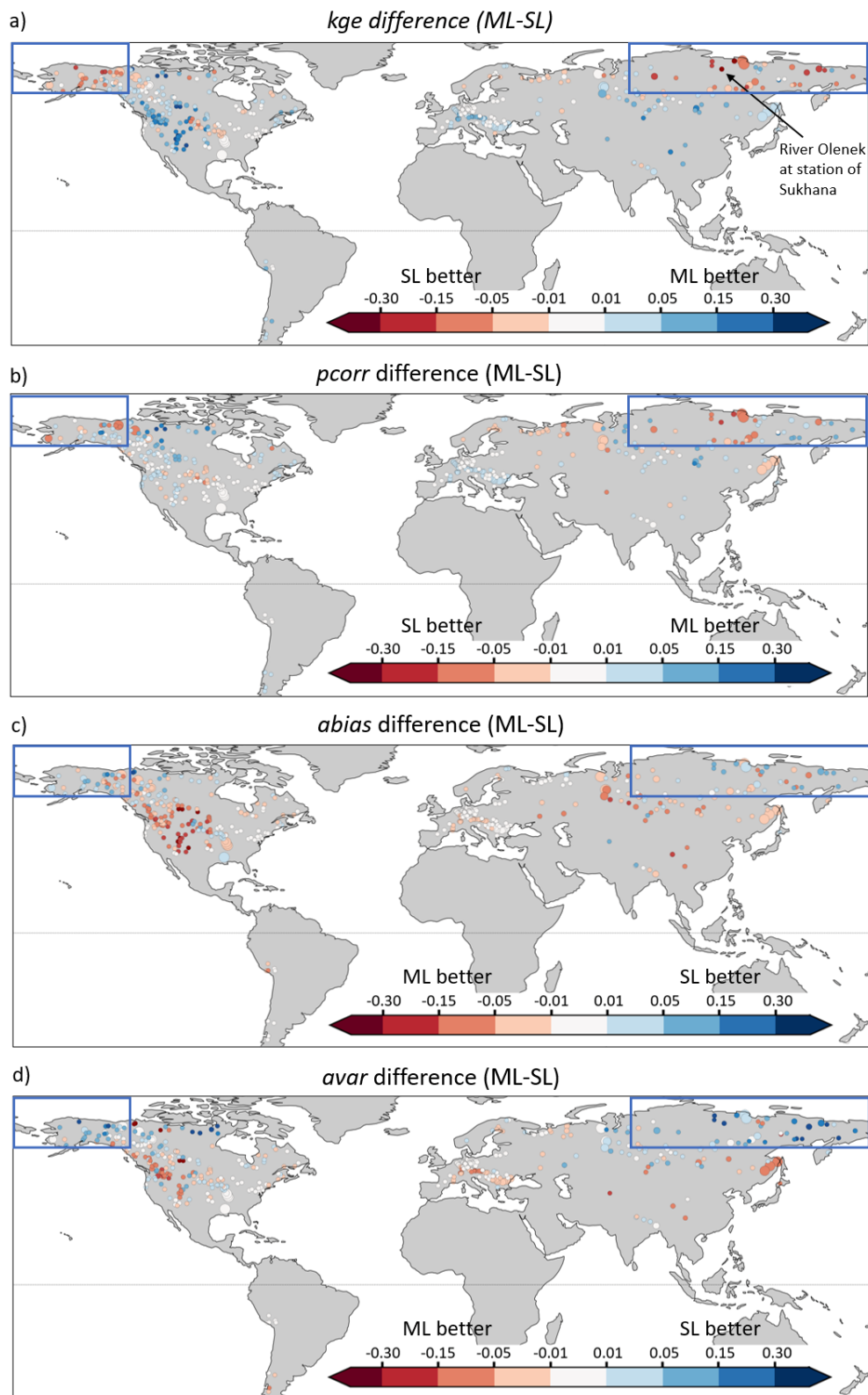


Figure 5-3. Difference of performance metrics between the default multi-layer (ML) and the single layer (SL) snow scheme experiments, across all 453 stations, calculated on daily river discharge over 1980–2018. (a) Modified Kling–Gupta efficiency (*kge*) and (b) Pearson correlation (*pcorr*), (c) absolute bias ratio (*abias*) and (d) absolute variability ratio (*avar*). Improvements in ML are indicated by blue dots in (a,b), while by red dots in (c,d). Size of the dots represent the catchment area. The sensitivity area in the permafrost is shown by blue rectangles, while the test catchment on the river Olenek at station Sukhana is indicated by black arrow in (a).

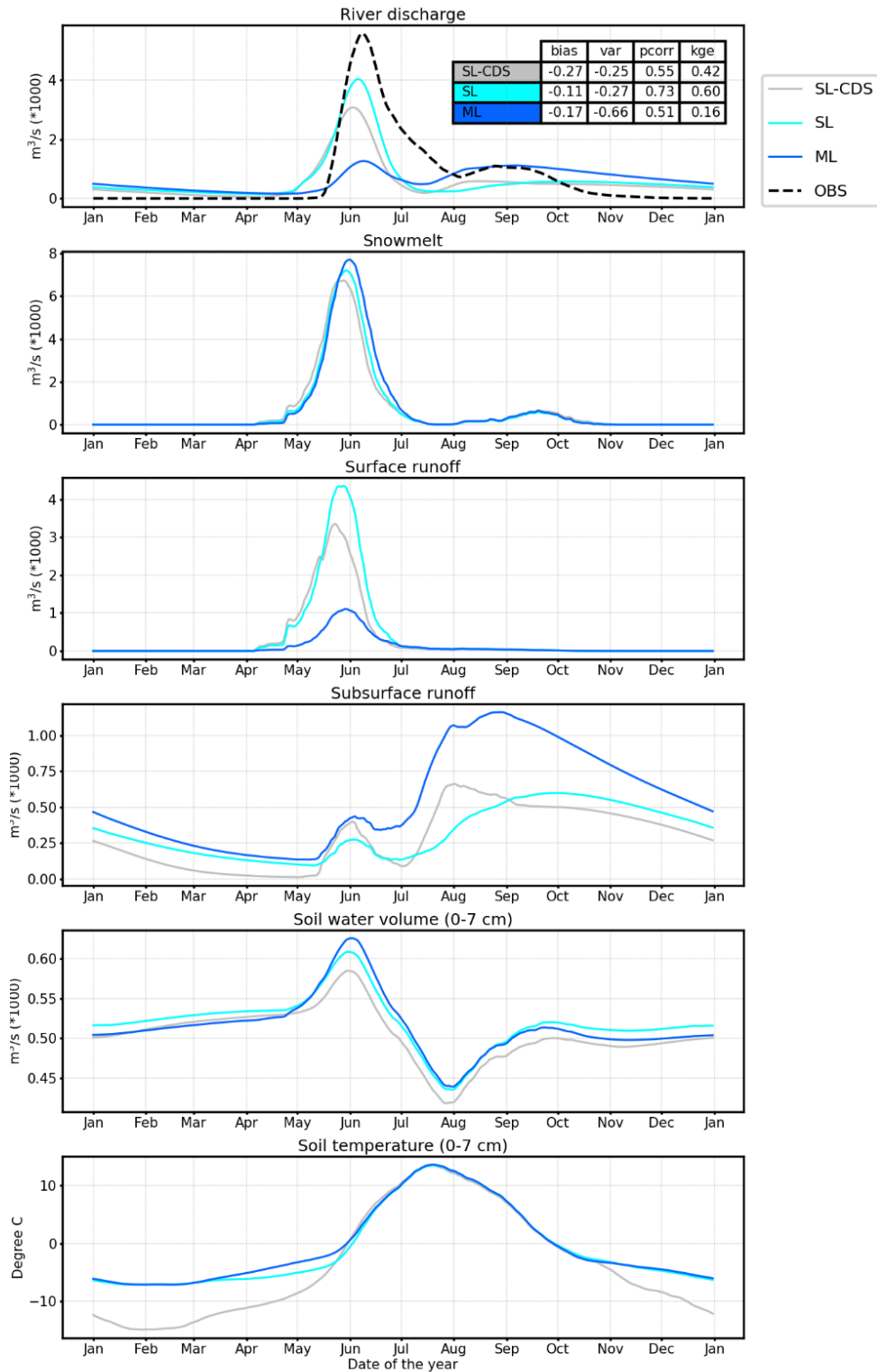


Figure 5-4. Daily climatological mean time series of (a) river discharge, (b) snowmelt, (c) surface runoff, (d) subsurface runoff, (e) water content and (f) temperature in the top 7 cm of the soil from SL-CDS, SL and ML experiments for the Olenek river at the station of Sukhana in eastern Siberia (with area of 127.000 km²). All water related variables are displayed as catchment totals in order to compare them directly to river discharge, while for soil temperature catchment averages are shown (please note the values are divided by 1000 for river discharge, snowmelt and subsurface runoff and by 1.000.000 for soil water volume). The kge and its bias, var and pcorr component scores are provided in an inset table for river discharge.

The improvements in midlatitudes seem to come mainly from the smaller bias errors (*abias* decreasing), especially pronounced in central North America (Figure 5-3c) and to a lesser extent from the smaller variability errors (*avar* decreasing; Figure 5-3d) and higher correlation (Figure 5-3b).

On the other hand, the deterioration in the higher latitudes over the permafrost seems to relate to much higher variability errors (many catchments with an increase of at least 0.3) and to somewhat higher bias errors and lower correlation in ML.

5.3.2 ML struggles in permafrost

Analysis in Siberia demonstrates that the deterioration of the daily river discharge representation and the largely missed snowmelt-driven flood wave is caused by too low surface runoff, which primarily comes from warmer soil in ML, allowing more water infiltrating into the soil and thus reducing surface runoff.

A test catchment on the Olenek river in Siberia (Sukhana station from within the permafrost sensitivity area; indicated in Figure 5-3) is used to demonstrate this large negative impact through analysing the daily climate time series of some key water budget related land-surface variables (snowmelt, surface and subsurface runoff, soil water content and soil temperature) for SL-CDS, SL and ML (Figure 5-4).

The ML-produced river discharge is clearly inferior compared with both SL and SL-CDS, as it shows a much larger underestimation of the observed flood peak in May-June (Figure 5-4a). Even though the secondary flood period is better represented by ML in August-September, due to the generally higher river discharge with the exception of May-June. The deterioration of the flow is reflected in the ML scores being the lowest across all three experiments (inset table in Figure 5-4a). The only exception is the bias ratio which is lowest for SL-CDS. This very low negative bias is a consequence of the snow data assimilation that removes water from the rivers in high latitudes. It is related to the current single layer snow model's tendency to melt the snow too slowly, which is compensated by the assimilation system during snowmelt periods, as documented in Zsoter et al. (2019).

The snowmelt peak in May-June is slightly higher and delayed in ML, compared with SL and even more so with SL-CDS (Figure 5-4b). At the same time, the soil is better insulated by the multi-layer snow in ML during March-May, as the temperature is higher by up to 2 °C than in SL (Figure 5-4f). As the snow is better insulated in ML, the melting will start later during spring, which then delays the faster soil temperature increase in ML, occurring in end of May. The surface runoff is highest in SL and lowest in ML during April-June, also highlighting a similar delay seen on the snowmelt peaks (Figure 5-4c).

The higher soil temperature in ML implies that a smaller fraction of the water is frozen in the soil, compared to SL, during the crucial snowmelt period and more water can infiltrate into the soil, which then reduces the amount of water that can runoff directly. The higher infiltration will increase the water content in the soil during May-July (Figure 5-4e), which will result in increased subsurface runoff (Figure 5-4d). This happens with a delay, as the water needs to reach the bottom of the soil to leave as subsurface runoff, producing a delayed peak to around end of July-September (Figure 5-4d).

The missing river discharge in ML is clearly related to the too low surface runoff during the April-June snowmelt, which period is clearly better represented in both SL and SL-CDS. The small differences in snowmelt cannot explain the lower surface runoff amounts. If anything, it should likely contribute to an increase in ML. In addition, the deficit in surface runoff is offset by the higher subsurface runoff in ML. However, the excess water from subsurface runoff spreads out over most parts of the year and thus cannot compensate for the missing surface runoff during the snowmelt flood period. Instead, the crucial aspect of the changes in ML is the higher soil temperature. This soil temperature difference is most critical in the middle of May, when the snowmelt is rapidly increasing in the catchment.

5.3.3 Improving the multi-layer snow scheme performance in permafrost

In this section, the ECLand sensitivity experiments (Figure 5-2 and Section 5.2.8) are evaluated for their ability to improve the river discharge simulation, focusing on permafrost.

5.3.3.1 Impact of the ECLand experiments on a test catchment in Siberia

The ECLand permafrost sensitivity experiments are demonstrated to increase the surface runoff in a test catchment in permafrost, by primarily making the soil colder through a series of incremental changes in the snow and soil freezing parametrizations, including modifications of the snow vertical discretization, snow density metamorphism, snow-soil thermal conductivity and soil freeze temperature.

In the Sukhana test catchment, the modifications are clearly effective in altering the amount of water distributed amongst different parts of the water budget (Figure 5-5). Generally, the impact on snowmelt (Figure 5-5b) is the smallest (maximum of a few days delay and a slight change in magnitude), while all other variables show much larger variability between the experiments (Figure 5-5a,c-f).

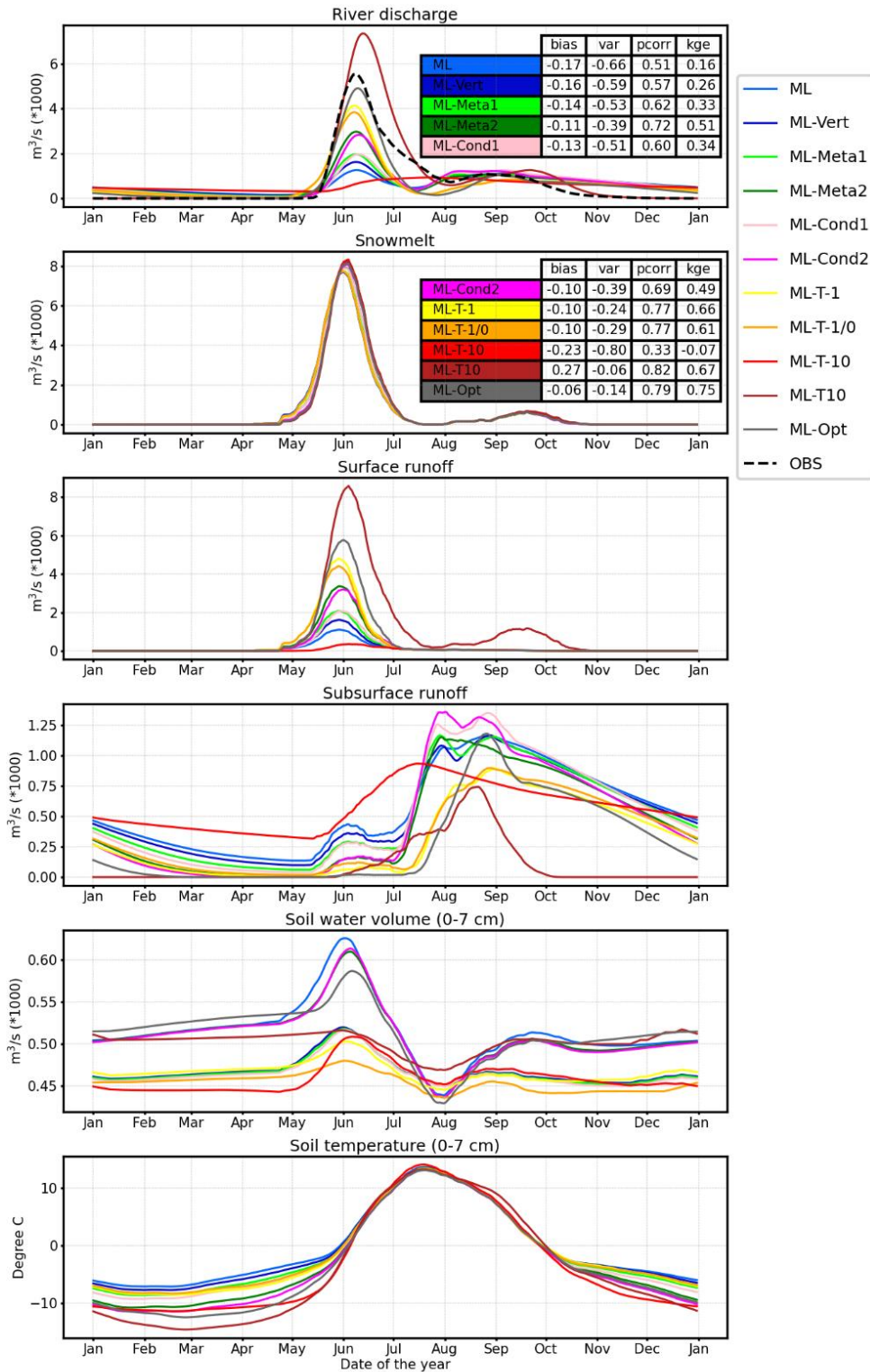


Figure 5-5. Daily climatological mean time series of (a) river discharge, (b) snowmelt, (c) surface runoff, (d) subsurface runoff, (e) water content temperature in the top 7 cm of the soil from ML, ML-Vert, ML-Meta1, ML-Meta2, ML-Cond1, ML-Cond2, ML-T-1, ML-T-1/0, ML-T-10, ML-T10 and ML-Opt experiments for the Olenek river at the station of Sukhana in eastern Siberia (with area of 127,000 km²). All water related variables are displayed as catchment totals in order to compare them directly to river discharge, while for soil temperature catchment averages are shown (please note the values are divided by 1000 for river discharge, snowmelt and subsurface runoff and by 1,000,000 for soil water volume). The kge and its bias, var and pcorr component scores are provided in an inset table for river discharge.

The surface runoff shows very large sensitivity to the ECLand changes in May-June, coinciding with the main snowmelt period (Figure 5-5c). The river discharge behaviour (Figure 5-5a) is directly determined by the variability in surface runoff in this period. The vertical discretization adjustment (ML-Vert), which is active for about 25% of the test catchment area, usually during October-May, and the combination with the first snow metamorphism change (ML-Meta1) both add small increase. However, ML-Vert combined with the second snow metamorphism setting (ML-Meta2) produces a much larger impact on surface runoff. On the other hand, the change in surface conductivity does not appear to be effective as both ML-Meta1 with ML-Cond1 and ML-Meta2 with ML-Cond2 (both pairs sharing the same snow metamorphism change respectively) show similar levels of surface runoff.

The experiments with the largest impact on surface runoff (and thus river discharge) are those including changes of the soil freezing. The experiments of ML-T-1 and ML-T-1/0 help reaching the surface runoff level of SL by resetting the freeze temperature to $-1\text{ }^{\circ}\text{C}$ and the thaw temperature to $0\text{ }^{\circ}\text{C}$. ML-Opt shows even further improvement with a surface runoff peak that closely matches the shape and magnitude of the observed river discharge until middle of June. This is achieved by the combined impact of the vertical discretization, second snow metamorphism, surface conductivity and freeze temperature changes. On the extreme end of the spectrum, the ML-T10 experiment shows too high surface runoff, by setting the freeze and thaw temperatures to an unrealistically high level (at around $+10\text{ }^{\circ}\text{C}$), consequently partitioning much of the runoff into surface runoff.

The soil temperature changes are in agreement with the surface runoff behaviour described above. The soil is warmest in ML, while it gradually gets colder by the ECLand modifications (Figure 5-5f). The soil is coldest for ML-Opt and the two extreme temperature experiments, ML-T-10 and ML-T10, during the main snowmelt season in March-May. The cooling soil means, the temperature will be more and more likely closer or below the freeze temperature threshold (T_{Fr}), which progressively decreases the infiltration into the soil. This leads to generally reduced soil water content (Figure 5-5e) and decreased subsurface runoff (Figure 5-5d), although with large variability depending on the actual parametrization changes and the time of year.

The *kge* and the three component scores (inset tables in Figure 5-5a,b) confirm the gradually improving behaviour of these experiments from ML to ML-Opt, through the introduction of the incremental ECLand parametrization changes. The *bias* improves from -0.17 to -0.06 , *var* from -0.66 to -0.17 , *pcorr* from 0.51 to 0.71 and finally the *kge* from 0.16 to 0.75 , representing a very large jump in skill.

5.3.3.2 Impact of the ECLand experiments in permafrost

The series of ECLand experiments are shown to provide widespread improvement for river discharge in the permafrost sensitivity area, based on all available catchments. They result mainly in reduced bias and variability errors, with optimal performance achieved by the ML-Opt experiment, including the combined changes of vertical discretization, second snow metamorphism, surface conductivity and freeze temperature.

The 69 catchments, found in the sensitivity area defined in the 60-80N latitude belt over the permafrost (Figure 5-1), were collectively analysed to see if the conclusions obtained from the test catchment in Siberia can be generalized to the permafrost. Figure 5-6 provides a visual summary of the overall performance in simulating river discharge, averaged over the sensitivity area, where the default configuration ML performs unfavourably. Each experiment is represented by a dot in a 3-dimensional graph showing *bias* (x-axis), *var* (y-axis) and *pcorr* (symbol size).

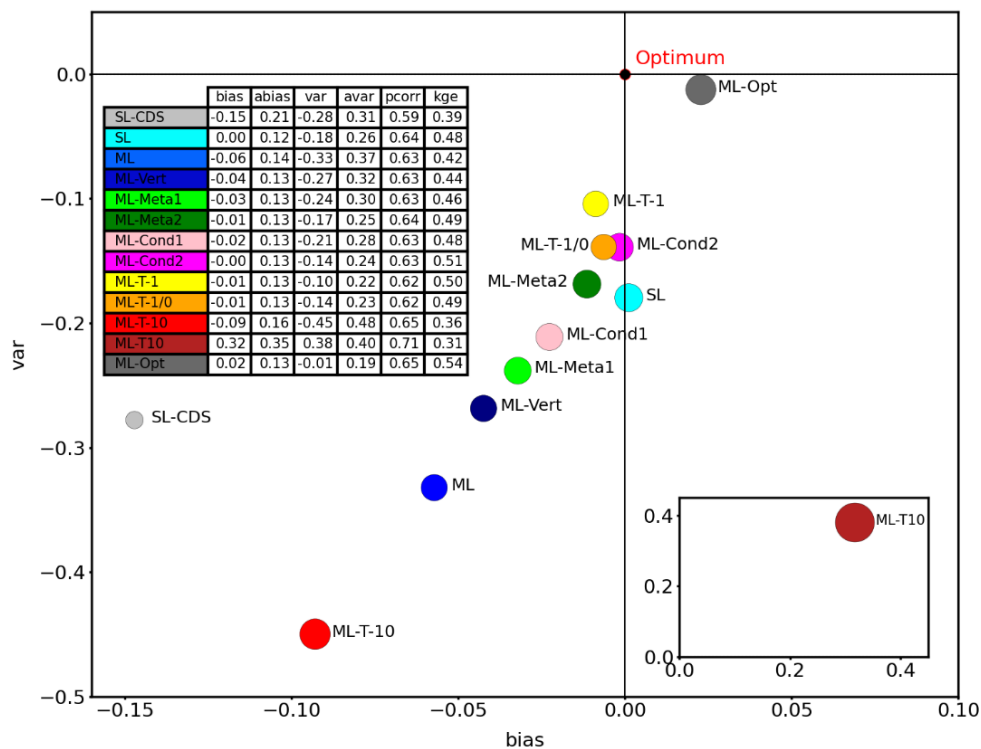


Figure 5-6. Scatter plot of area average scores for the 13 analysed experiments. The scores are computed from 69 catchments in the 60–80 N belt of the permafrost (see Figure 5-1 for the area). The size of the dots represents *pcorr*. The outlier ML-T10 is displayed in an inset graph for better readability. The area average score values are provided for all analysed experiments in the inset table.

The ECLand parametrization changes are effective to improve the simulation of river flow in the permafrost area, in particular the bias and variability errors, which improve from ML-T-10 to ML-Opt (see also the average scores of 69 catchments in an inset table in Figure 5-6). The correlation is less impacted by the modifications, the difference between all experiments is mostly within a few percent (the two extremes have average correlation of 0.59 (SL-CDS) and 0.71 (ML-T10)).

The furthest from the optimum [0; 0] *bias/var* point are the ML-T-10 and ML-T10 extreme temperature experiments. Nevertheless, the highest average correlation is achieved with the positive extreme, when the soil remains mostly frozen. The fact that an unrealistically extreme-setup simulation, with excess surface runoff, produces the highest correlation shows that the current land-surface process representation in ECLand is still suboptimal in the permafrost, even after the parametrization changes. Another outlier is the online produced SL-CDS, which has lower scores, especially the very low negative bias, due to the snow data assimilation removing water from the rivers in the northern latitudes, as a compensation for the slow snowmelt in SLS (Zsoter et al., 2019).

Overall, there is a notable improvement achieved by the ECLand experiments, shown by the increase in *kge* from 0.42 (ML) to 0.54 (ML-Opt), while the single-layer experiments highlight lower *kge* of 0.48 (SL) and 0.39 (SL-CDS).

5.3.4 Global impact of the ECLand experiments

The changes in the snow and soil parametrization schemes in ECLand could improve the river discharge in the permafrost sensitivity area in eastern Siberia and Alaska. However, in other areas these changes are dominantly suboptimal, as the difference between ML-Opt, the highest performing ECLand experiment, and ML highlights (Figure 5-7).

The worst ML-Opt performance is found in the western part of USA, Canada and catchments in central Asia, directly in those areas where ML could improve the most on SL (see Figure 5-2). Interestingly, the Canadian Arctic regions (Nunavut) do not show similarly improving results to other permafrost regions. One source of error, which can contribute to this, is the underlying physiographic data used to drive the simulations, e.g., lake cover. Boussetta et al. (2021) showed that recent lake cover datasets are substantially different from those used in this work for the Nunavut region. The use of more realistic physiographic data will be explored in future work. Furthermore, unrepresented physical processes in the land-surface model can affect the hydrological cycle of these regions, e.g., wind-driven sublimation of the snowpack which can change the amount of snow mass available for melting at the end of the season. Similarly, the handling of frozen lakes with, e.g., snow not accumulating on top of lake ice in ECLand, may introduce compensating errors with the changes tested in ML-Opt.

Nevertheless, even though only about one third of the stations show improvement by ML-Opt, the globally averaged scores show that ML is only slightly better (inset table in Figure 5-7). Moreover, the score variability amongst all the experiments, in general, is quite small as well. The only real exception is ML-T10 that has very low scores throughout. The geographical maps of the *kge*

differences between the experiments (from ML-Vert to ML-T10) and ML is provided in the Supplementary Materials (Figures 5-S1 - 5-S9).

The deterioration of ML-Opt and the other ECLand experiments in milder climate areas is related to the surface runoff increase during the snowmelt period, demonstrated earlier. Especially in catchments where the bias in ML (and similarly in SL) was originally positive, e.g., around the Rockies (as suggested by Zsoter et al. (2019)), any further increase in snowmelt-related surface runoff will dominantly be detrimental and contribute to further increased bias and thus lower *kge*.

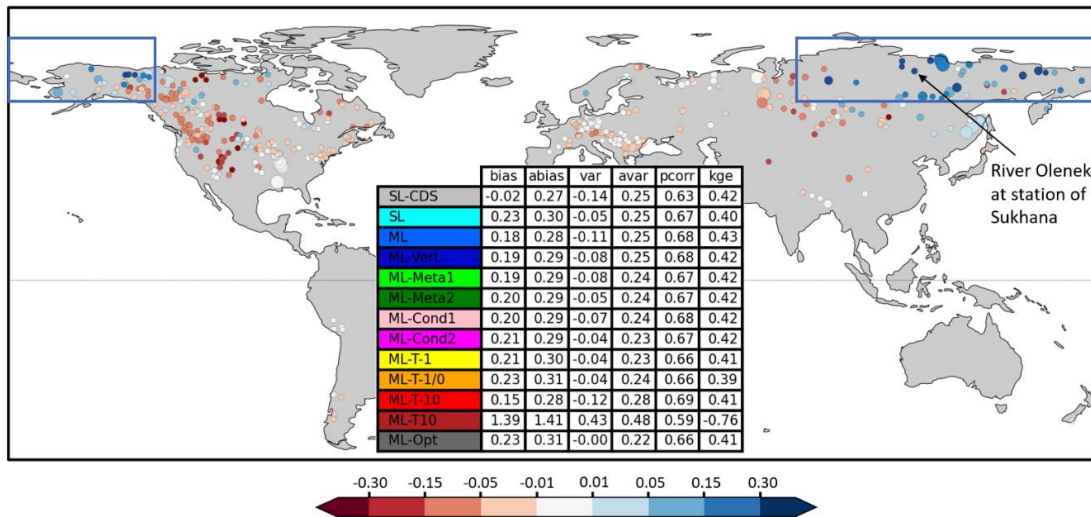


Figure 5-7. Difference of the modified Kling–Gupta efficiency (*kge*) between ML-Opt and ML. Blue areas show better performance in ML-Opt, while red indicates better performance in ML. The sensitivity area in the permafrost is shown by blue rectangles, while the test catchment on river Olenek at station Sukhana is indicated by black arrow. The area average scores are provided for all analysed experiments in the inset table.

5.4 Discussion

It was shown in this study that the MLS has an improved representation of the hydrology, with the notable exception of the coldest areas in permafrost, where the SLS is superior. To improve the MLS hydrological representation in permafrost, modifications of the ECLand snow and soil freezing schemes were tested. It was shown that a series of incremental changes could noticeably improve the quality of the river discharge simulation over a large area in permafrost, primarily through decreasing the soil temperature and thus increasing the amount of surface runoff in the critical spring snowmelt period.

Improving the hydrological process representation

The results have demonstrated that the use of uniform parameters in ECLand in the snow and soil freezing schemes, currently applied in ECLand, are too simplistic and will not work for both the permafrost and non-permafrost areas in the snow-impacted climate. Spatially variable parametrization for variables, such as c_{ξ} in the destructive metamorphism of the snow, l_b in the

snow-soil thermal conductivity and T_{Fr} freeze temperature in the soil freezing scheme, explored in this work, promise to bring a more balanced approach for delivering improved hydrological process representation.

However, this could technically be complex to implement in ECLand and needs a substantial amount of further research. As a first step, the parametrization part of ECLand has recently been refactored (by removing hard-coded parameter values) and the Multiscale Parameter Regionalization (MPR; Schweppe et al., 2021) has been implemented for estimation of spatially varying parameters. This change will make it easier in the future to work on aspects of ECLand, such as the one explored in this study for the snow and soil parametrization impact on hydrology.

This study has also contributed to the understanding of the hydrological importance of each tested change in the ECLand parametrization, which could be based on the average verification metrics in the permafrost (distance between dots in Figure 5-6), supported by the land-surface processes representation at the example catchment (Figure 5-5).

The thaw temperature adjustment from +1 to 0 brought deterioration, while all other changes could improve river discharge. The smallest improvement seems to come from the snow-soil thermal conductivity adjustment (updated l_b), closely followed by the first snow metamorphism change (variable c_s). The snow vertical discretization change looks to bring larger improvements (adjustment over complex terrain), while the second snow metamorphism change (uniform c_s) is even more beneficial in the permafrost. Finally, the biggest contribution appears to come from the adjustment of the lower temperature threshold in the soil freezing scheme (freeze temperature) from -3 to -1 °C.

Land-surface modelling challenges

Due to the presence of a unique soil column within a grid cell in EC-Land, the soil temperature may be largely affected by the repartitioning of the sub-grid tiled surfaces, as these can be characterized by different surface energy fluxes. In high latitude during transition seasons (e.g., snow accumulation and ablation), the tile fraction subdivision depends on the snow cover fraction, exerting a large control on the heat flux conducted through the soil (Slater et al., 2001; Yang et al., 1997; Bélair et al., 2003b). When the snow cover fraction is less than one, part of the soil directly interacts with the atmosphere above, without the thermal insulation effect of the snowpack. This in turn, may affect the soil temperature and therefore the fraction of frozen water within the soil. Future modelling work should evaluate the hydrological sensitivity to snow cover fraction parametrization in global land-surface simulations (see for example Niu and Yang, 2007), considering the combined effect that this can have on the soil below and the atmosphere above.

In addition, future modelling work should also evaluate the effect of unrepresented or poorly represented hydrological processes in ECLand. These include, for example wind-driven snow sublimation, snow interception by forests, as well as of a more physically complex representation of frozen/unfrozen water phases in the soil column, which all promise to bring benefits to the hydrological cycle, especially in permafrost regions (see for instance Koren et al., 2014; Krogh et al., 2017; Niu and Yang, 2006). Moreover, the impact of a deeper soil column, with additional vertical layers, has also been shown to work better for permafrost simulation in land-surface models (Melton et al., 2019). More complex models for hydrological applications in high-latitude regions have been proposed in the literature (e.g., Niu and Yang, 2007; Melton et al., 2019; Clark et al., 2021). Their implementation in ECLand could be considered in the future.

Relevance for ECMWF

The next operational IFS cycle (48r1) of ECMWF will include multi-layer snow representation. Based on this study, the slightly modified version of ML-Meta1 was selected for operationalization, which includes the snow vertical discretization and first snow metamorphism changes on top of the default ML configuration. The results of this study highlighted that this new snow model did show moderate hydrological improvements in permafrost and could still retain most of the good performance of the default MLS in other areas.

Earth system modelling implications

Modelling improvements in the Earth system process representation, such as the use of any prospective MLS version explored in this study, require thorough testing in coupled forecast mode with data assimilation, similarly as in Arduini et al. (2019). This is necessary to check the transferability of the offline-demonstrated hydrological improvements, and make sure other variables, such as 2-metre temperature and surface fluxes of sensible and latent heat, will not deteriorate.

This study has proven that further development of the snow and soil parametrizations in ECLand is crucial to achieve better hydrological performance everywhere globally; however, it could not be part of this study and will only be explored in the future as an important research area.

The success of helping the operational development of the ECLand model's snow component has shown that hydrological studies, such as the work presented in this paper, have great potential to help improve the land-surface realism in Earth system models and can contribute to improvements in not just the hydrological variables, such as river discharge, but potentially other components of the Earth system as well.

Limitations of the study

The authors acknowledge that even with the best care taken in the experimental setup of this study, some limitations remain. The network of analysed catchments is still under-representative in some areas, also with the relatively short minimum observation length of eight years as a compromise. Similarly, this study was not designed to be a full sensitivity experiment for the land-surface processes in the permafrost, which would have required a much higher computational cost. In addition, the choice of testing ECLand changes in MLS was because of practical reasons, but most of the modifications would be expected to show improvements in a similar manner even if tested within the SLS.

5.5 Conclusions

The representation of snow is a crucial aspect of land-surface modelling, as it has a strong influence on the energy and water balances. Snow schemes with multiple layers can better represent the snowpack evolution and bring improvements on single-layer schemes in simulating the snow processes and contribute to better soil freezing and hydrological cycle.

In this paper, the hydrological impact of the MLS, implemented in ECLand, was analyzed globally with ERA5-forced experiments over the period of 1979–2018. The CaMa-Flood model was used to generate river discharge from the ECLand runoff output, allowing direct comparison with gauged observations over more than 400 snow-impacted catchments. Different sensitivity experiments were conducted to evaluate the impact of changes in the ECLand snow and soil freezing schemes on the terrestrial hydrological processes, with particular focus on permafrost.

It was found that while the default MLS generally improves the river discharge simulation, mainly through better bias and variability errors, the performance is suboptimal in large parts of the high latitude permafrost regions. The analysis of the climatological mean time series, in a test catchment in Siberia, demonstrated that the largely underestimated snowmelt-driven floods in late spring to early summer are caused by too low surface runoff. The MLS better insulates the soil, which allows more water to infiltrate into the soil and thus surface runoff is reduced.

It was also found that the ECLand experiments provide widespread improvement for river discharge in the sensitivity area, defined in permafrost. The incremental changes of the snow vertical discretization, destructive metamorphism, snow-soil thermal conductivity and soil freeze temperature lead to gradually colder soil, which resulted in increased surface runoff and thus better river discharge simulation during the critical snowmelt-driven flood period. The ML-Opt experiment, as the combination of the best performing ECLand changes, has shown higher overall *kge*, mainly through reduced bias and variability errors.

The results, presented here, have directly influenced the MLS version that will be introduced in the next Integrated Forecasting System cycle of ECMWF (48r1). This has demonstrated that hydrological analyses, such as the work presented in this paper, can provide a useful platform to diagnose areas where improvements are needed in the land surface representation of the Earth system models.

Funding: Ervin Zsoter's PhD is supported by the Wilkie Calvert Co-Supported PhD Studentships at the University of Reading. Christel Prudhomme and Ervin Zsoter were supported by the Copernicus Emergency Management Service-Early Warning Systems (CEMS-EWS). Hannah Cloke is supported by the EVOFLOOD project: The Evolution of Global Flood Risk UK NERC, NE/S015590/1. Elisabeth Stephens and Hannah Cloke are supported by the FATHUM project: Forecasts for Anticipatory HUMANitarian Action funded by UK NERC as part of their Science for Humanitarian Emergencies & Resilience (SHEAR) program, NE/P000525/1.

Acknowledgements. We are grateful to The Global Runoff Data Centre, 56068 Koblenz, Germany for providing observations for our river discharge analysis.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Contribution of this chapter to the thesis. This chapter addressed the objective: "Evaluate the hydrological impact of the complexity of the snow scheme in the land-surface models in reanalysis simulations, with special focus on cold climate areas in permafrost." This work demonstrated that by replacing the single-layer snow scheme with a physically more complex multi-layer scheme can improve the river discharge reanalysis simulations. The chapter also highlighted that in the coldest areas of permafrost, river discharge could only be improved by further adjustments of the snow and soil freezing parametrisations.

5.6 Supplementary figures

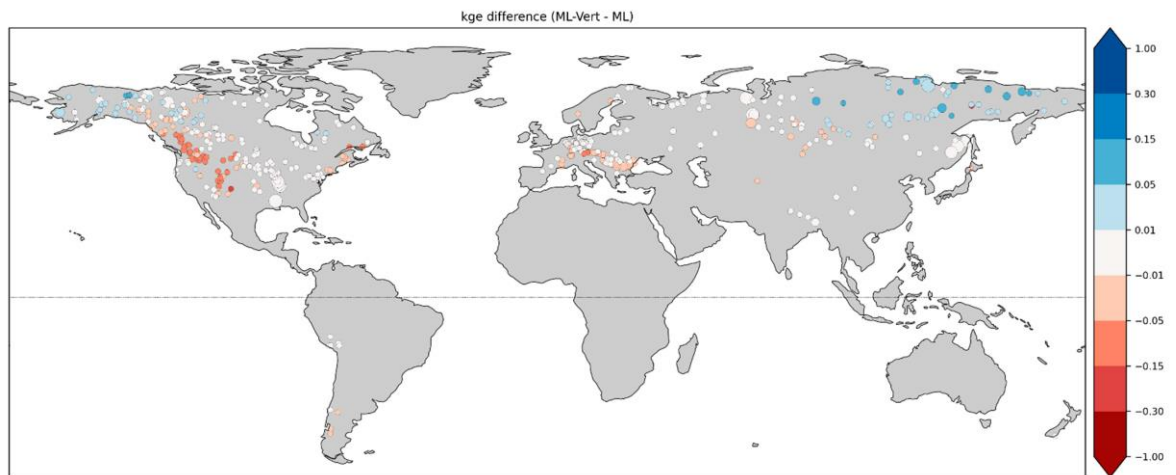


Figure 5-S1. Difference of the Kling-Gupta efficiency (kge) between ML-Vert, the experiment with modified snow vertical discretisation over complex terrain with at least 25 cm snow depth, and ML, the default multi-layer experiment, across all 453 stations, calculated on daily river discharge over 1980-2018. Improvements in ML-Vert are indicated by blue dots. Size of the dots represent the catchment area.

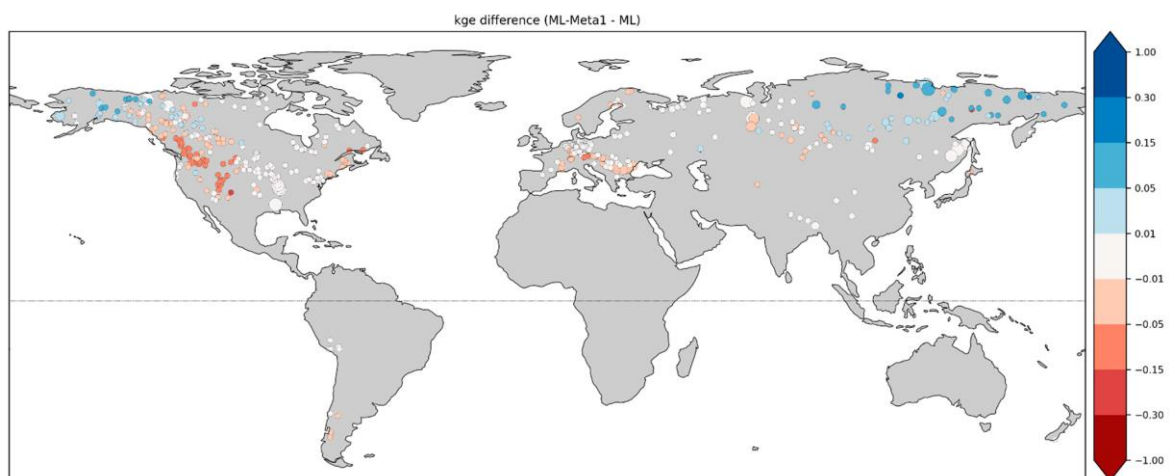


Figure 5-S2. As Figure 5-S1, but kge difference between ML-Meta1 (the 1st modification of the destructive metamorphism of the snow with variable c_{ξ} parameter values across the snow layers, together with ML-Vert) and ML.

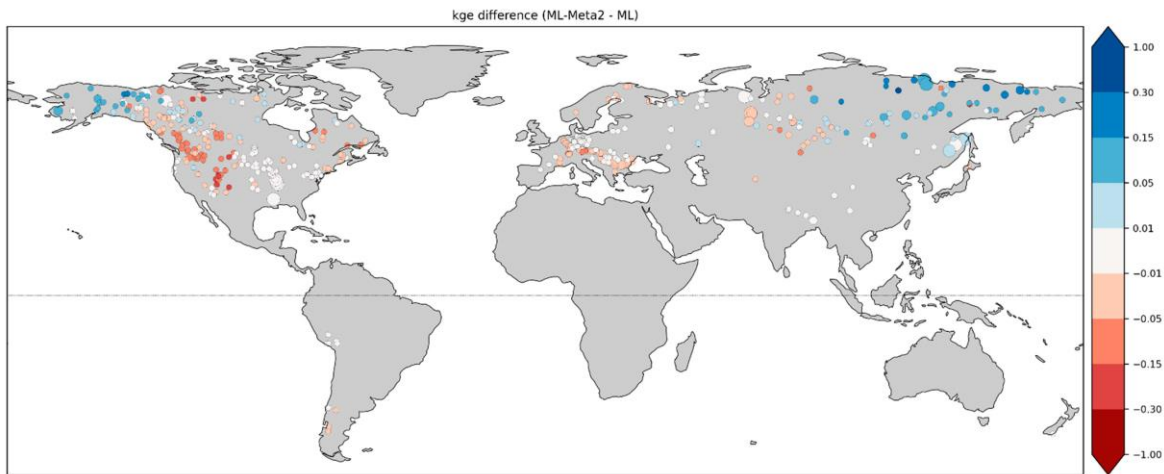


Figure 5-S3. As Figure 5-S1, but kge difference between ML-Meta2 (the 2nd modification of the destructive metamorphism of the snow with one c_{ξ} parameter value as in Anderson et al. (1976), together with ML-Vert) and ML

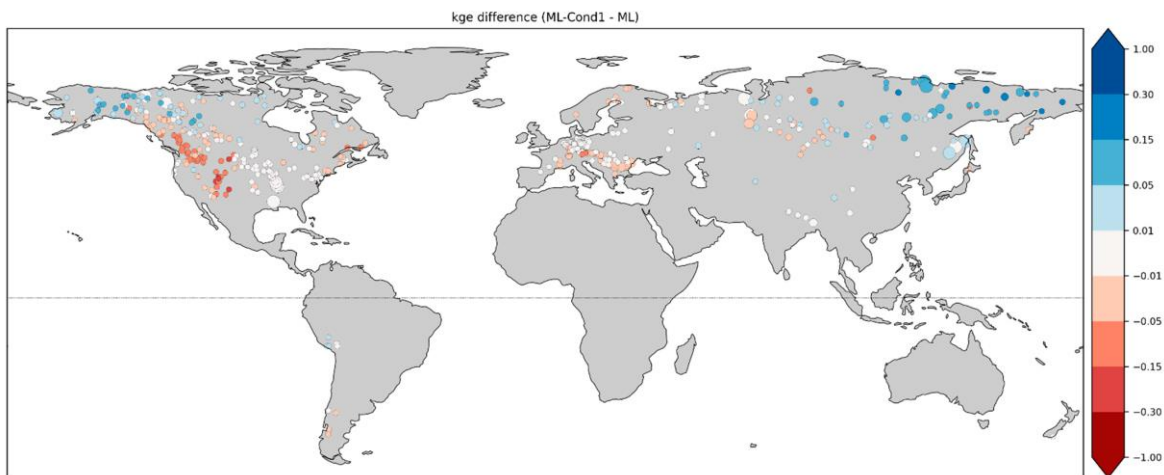


Figure 5-S4. As Figure 5-S1, but kge difference between ML-Cond1 (ML-Meta1 together with the snowsoil thermal conductivity computation with the revised parameter $l_b=0.5$) and ML.

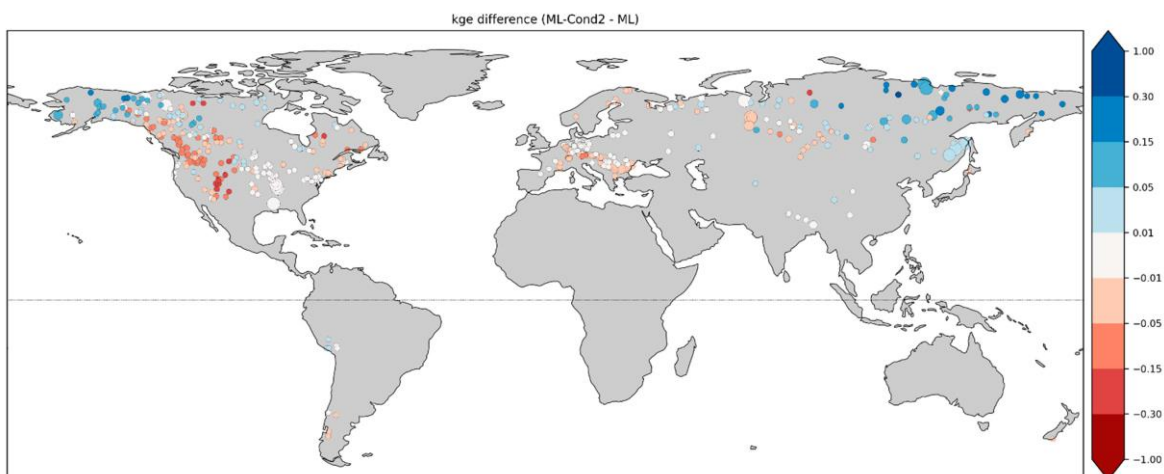


Figure 5-S5. As Figure 5-S1, but kge difference between ML-Cond2 (ML-Meta2 together with the snowsoil thermal conductivity computation with the revised parameter $l_b=0.5$) and ML.

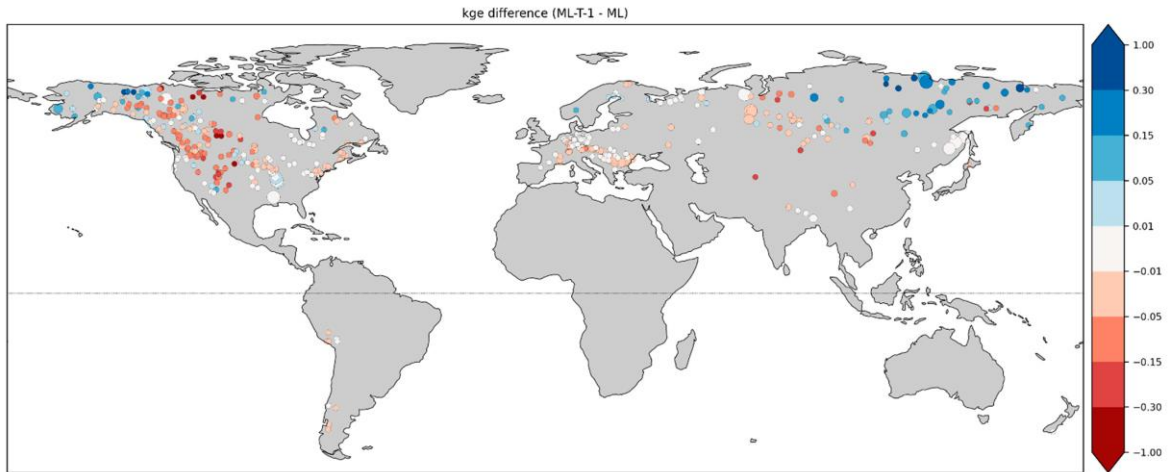


Figure 5-S6. As Figure 5-S1, but kge difference between ML-T-1 (change of the soil freeze parameter T_{Fr} from the default -3 to -1 °C added onto ML-Meta1) and ML.

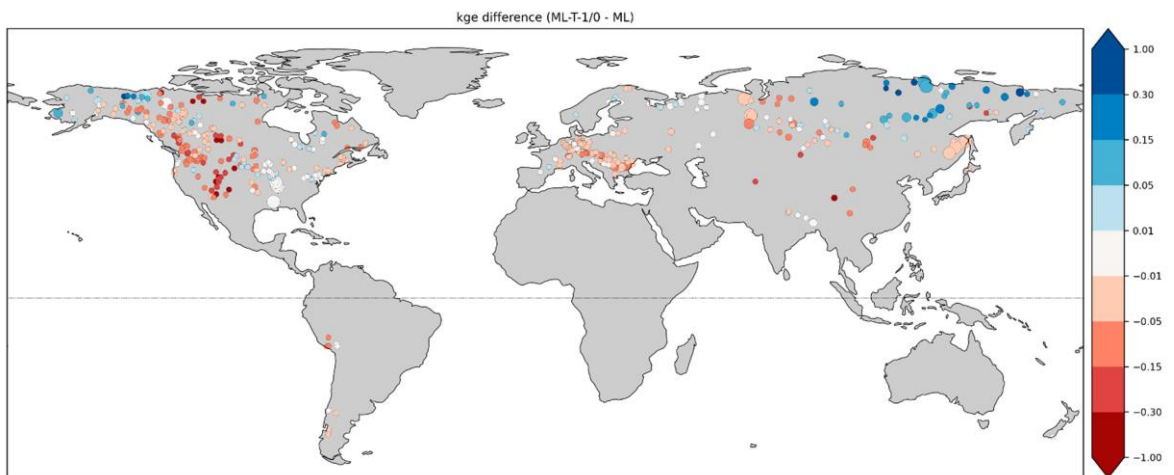


Figure 5-S7. As Figure 5-S1, but kge difference between ML-T-1/0 (change of the freeze temperature T_{Fr} to -1 °C and thaw temperature T_{Th} from the default +1 to 0 °C, added onto ML-Meta1) and ML.

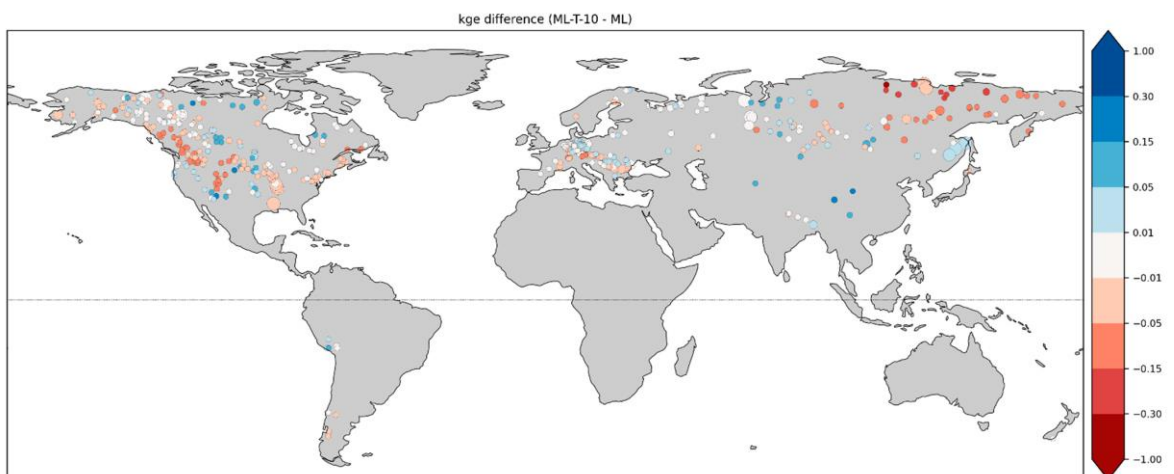


Figure 5-S8. As Figure 5-S1, but kge difference between ML-T-10 (T_{Fr} , T_{Th}) changed to [-10.5, -10], added onto ML-Meta1) and ML.

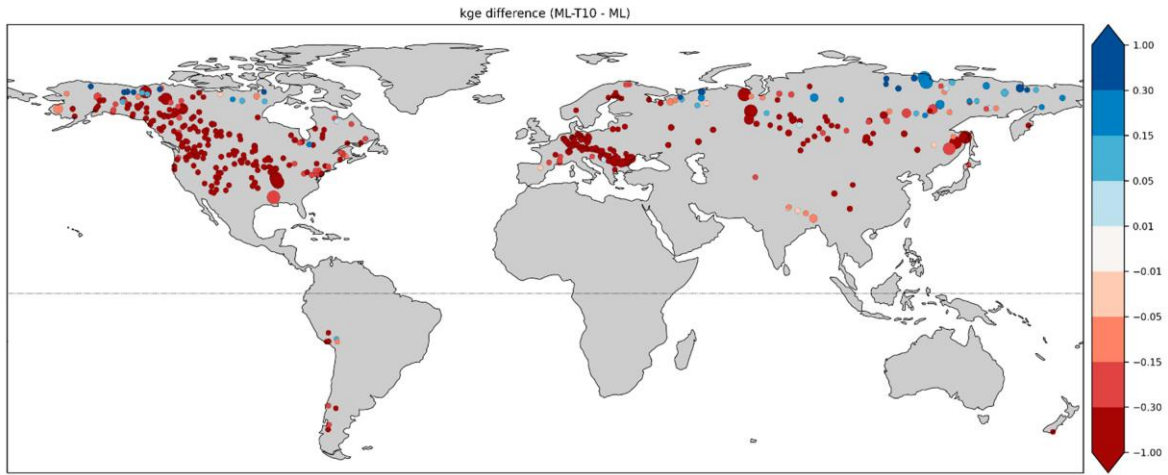


Figure 5-S9. As Figure 5-S1, but kge difference between ML-T10 ([TFr, TTh] changed to [+10, +10.5], added onto ML-Meta1) and ML.

Chapter 6 Trends in the GloFAS-ERA5 river discharge reanalysis

After analysing the hydrological contributions of land data assimilation and snow model complexity in Earth system models in Chapter 4 and Chapter 5, this chapter analyses the presence of non-stationarity in the reanalysis time series. Non-stationarity is an important aspect, as it is one of the potential limiting factors in the applicability of these Earth-system-based hydrological reanalysis data sets for the generation of flood hazard climatologies and the related flood warning thresholds. This chapter has been published as an ECMWF technical memoranda with the following reference (<https://centaur.reading.ac.uk/93047/>):

Zsótér, E., H. L. Cloke, C. Prudhomme, S. Harrigan, P. de Rosnay, J. Munoz-Sabater and E. Stephens, 2020b: Trends in the GloFAS-ERA5 river discharge reanalysis. *ECMWF Technical Memoranda*. 871. Technical Report, doi:10.21957/p9jrh0xp

The contributions of the authors of this paper are as follows: E.Z. designed and carried out the analysis and led the writing of the manuscript. H.C., E.S., S.H, P.deR. and J.M-S. all helped with the research methodology design. All authors assisted with writing the manuscript. Overall, 90% of the writing was undertaken by E.Zs.

The published article can be found in the Appendix A3.

Abstract. The main objective of this study is to analyse the GloFAS-ERA5 river discharge reanalysis for any noticeable change (including gradual trends or discontinuities) in the annual mean time series across the 1979-2018 (40-year) period, and to evaluate how realistic these are compared with available observed river discharge time series. These variabilities are quantified by linear regression in order to highlight any concerning features in the GloFAS-ERA5 time series. This work is particularly important for GloFAS, as large trends, discontinuities or other similar features could have a major consequence on the GloFAS flood thresholds in around 50% of catchments, which are based on GloFAS-ERA5, and thus subsequently on the issuing of flood warnings. In addition, this study also contributes to the understanding of the water cycle variable behaviour in ERA5 (driver of GloFAS-ERA5) and ERA5-Land (higher resolution land reanalysis forced by ERA5, produced offline) by exploring the linear trends in river discharge and related hydrological variables. In exploring the stability of the time series in ERA5, we seek to trigger potential further discussions and research studies, which subsequently should help with the planning and development for the next generation ECMWF reanalysis, ERA6.

6.1 Introduction

In November 2019, the GloFAS (Global Flood Awareness System) was upgraded to version 2.1 (<https://confluence.ecmwf.int/display/COPSRV/GloFAS+v2.1>). This included two important changes: the release of the official GloFAS-ERA5 river discharge reanalysis, and the revision of the flood threshold computation. The flood thresholds are predefined and specified as a magnitude of river flow for specific return periods; these are used operationally to highlight where upcoming flows may be severe and trigger alerts accordingly. They are computed from river discharge reanalysis by fitting an extreme value distribution onto the annual maxima time series. In the v2.1 upgrade, flood thresholds were recomputed using the 40-year (1979-2018) GloFAS-ERA5 river discharge reanalysis, which is an ERA5-forced hydrological simulation (Harrigan et al., 2020b). In addition, the analysis of the v2.1 vs. v2.0 thresholds revealed that over large parts of the world the GloFAS-ERA5 river discharge time series has very noticeable linear trends across the 40-year period. Linear trends can highlight noticeable change across the 40-year period, be that a gradual shift (i.e. a trend) or a discontinuity (i.e. a step change at one point in the time series). Any noticeable shift in the time series is particularly important as it can hinder the representativity of the GloFAS thresholds, through the characteristically different extreme flood event behaviour in different parts of the 40-year period.

In this study, the changes/shifts in the GloFAS-ERA5 annual mean time series are analysed and quantified by linear regression. The linear trend magnitudes, along the regression lines, are computed for river discharge, as well as for all ERA5 and ERA5-Land variables that directly affect the water budget: precipitation, snowfall, evaporation, 2m temperature, soil water content, runoff and snowmelt. The linear trends in the available river discharge observations are used as verifying truth and compared with the GloFAS-ERA5 river discharge (and also ERA5 runoff) trends.

6.2 Data and methods

6.2.1 *Global Flood Awareness System*

The Global Flood Awareness System (GloFAS; www.globalfloods.eu) is part of the Copernicus Emergency Management Service (CEMS) and has been developed in collaboration between ECMWF, the Joint Research Centre of the European Commission with help from research institutions such as the University of Reading (UoR). It monitors and forecasts floods across the world. GloFAS has two complementary systems:

- GloFAS 30-day, that includes daily ensemble flood forecast predictions, out to 30 days ahead, updated daily, based on the ECMWF medium- and extended-range ensemble runoff as input forcing

- GloFAS Seasonal, that provides ensemble hydrological forecasts of unusually low or high flow for calendar weeks up to 16 weeks ahead, updated monthly, based on the ECMWF SEAS5 ensemble runoff as input forcing
- GloFAS forecasts possible flood episodes and unusually high/low river flow for all major rivers of the world. It has been an operational service since April 2018 (following a pre-operational phase which started in 2011) with information shown on a dedicated web platform (www.globalfloods.eu; Figure 6-1).

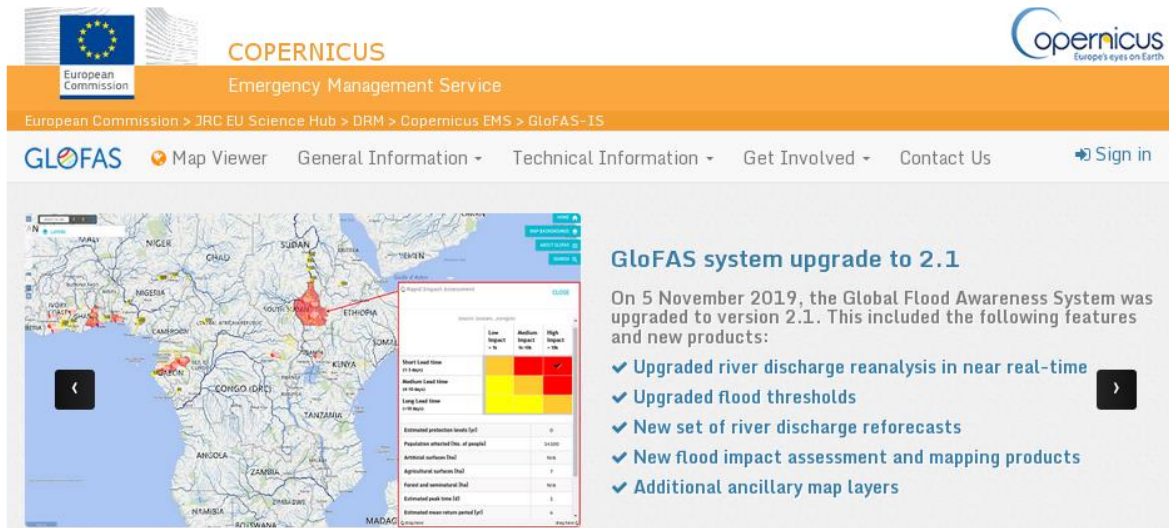


Figure 6-1. Snapshot of the GloFAS website's landing page.

GloFAS is designed to complement existing national and regional services, and to support national civil protection and international organisations in decision making before major flood events, particularly in large transnational river basins. Forecast information is used by a variety of decision makers, including national and regional water authorities, water resources managers, hydropower companies, civil protection and first line responders, and international humanitarian aid organisations. GloFAS only focuses on larger rivers (mainly over 10.000 km² catchment area) and does not provide real-time forecast information on flash flood risk, pluvial or coastal flooding.

GloFAS river discharge data is produced by coupling the Lisflood hydrological and channel routing model (van der Knijff et al., 2010) to the runoff output of the land-surface model of ECMWF NWP system. The river routing runs with surface and sub-surface runoff inputs on a 0.1×0.1° (~10 km resolution) global river network. The surface runoff component directly enters the river channel, while the sub-surface runoff first enters a groundwater module that outputs the water into the river channel after a time delay (Harrigan et al., 2020b).

The river state of the real time ensemble GloFAS forecasts is initialised from a hydrological simulation, forced by the fast release version of ERA5 (ERA5T) up until when it is available (a few days behind real time) and then subsequently by the ECMWF ensemble control forecast.

A long-term river discharge dataset is required in order to compute the return period flood thresholds in GloFAS 30-day: currently this is the GloFAS-ERA5 reanalysis. Further description on the GloFAS system is available on the GloFAS internal website at (<https://confluence.ecmwf.int/display/COPSRV/Global+Flood+Awareness+System>) and in Harrigan et al (2020b).

6.2.2 ERA5 and ERA5-Land

ERA5 is ECMWF's 5th generation global climate reanalysis (Hersbach et al., 2020). ERA5 covers the period January 1979 to near real time with a plan to extend to 1950. ERA5 includes one high-resolution component (~31 km horizontal resolution) and a lower resolution ensemble component with 10 members (~62 km horizontal resolution).

There are two flavours of ERA5 reanalysis available: the raw ERA5, based on consolidated, quality checked data, lagging ~2-3 months behind real-time; and ERA5T, which is available 5 days behind real time, but is not fully quality-checked (Hersbach et al., 2020). ERA5-Land (<https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land>) is also analysed for comparative purposes in this study. ERA5-Land is the offline land surface only improved version of ERA5.

ERA5 and ERA5-Land runoff are both produced by the HTESEL (Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land; Balsamo et al., 2009) land-surface model of the ECMWF Integrated Forecasting System (IFS).

The HTESEL scheme follows a mosaic (or tiling) approach where the grid boxes are divided into patches (or tiles), with up to six fractions over land (bare ground, low and high vegetation, intercepted water, shaded and exposed snow) and two extra tiles over water (open and frozen water) exchanging energy and water with the atmosphere. HTESEL is used within the ECMWF IFS in coupled atmosphere-surface mode on time ranges from medium-range to seasonal forecasts.

ERA5 is a coupled application which includes the operational land data assimilation system of ECMWF to assimilate conventional in-situ and satellite observations for land surface variables for the analysis of soil moisture, soil temperature and snow fields (de Rosnay et al., 2014).

ERA5-Land, on the other hand, is a product of an offline HTESEL simulation without atmosphere land coupling and land data assimilation, forced by the atmospheric variables (e.g. air temperature and radiation). ERA5-Land is produced at 9 km spatial resolution using downscaled ERA5 atmospheric forcing and a vertical lapse rate correction on temperature. There is no direct coupling or land data assimilation in ERA5-Land (there is only an indirect impact through the ERA5 forcing) and this can have a large impact on the hydrological cycle (Zsoter et al., 2019), and potentially also on the trends. Other major differences between ERA5 and ERA5-Land are the much higher

resolution and the lapse rate correction in ERA5-Land, which also can have a large impact on the water budget, especially in mountainous areas, changing the snow pack and snowmelt through the temperature differences.

ERA5-Land is analysed in a similar way to ERA5 for hydrological variables, apart from river discharge which is not yet produced for ERA5-Land. In addition, precipitation and snowfall are identical in ERA5 and ERA5-Land (apart from the downscaling) and therefore are not considered further within the ERA5-Land analysis.

6.2.3 GloFAS-ERA5 reanalysis

The GloFAS-ERA5 river discharge reanalysis is a product of the European Commission's Copernicus Emergency Management Service (CEMS) and is officially available in the Copernicus Climate Data Store free of charge after registration (<https://cds.climate.copernicus.eu/>).

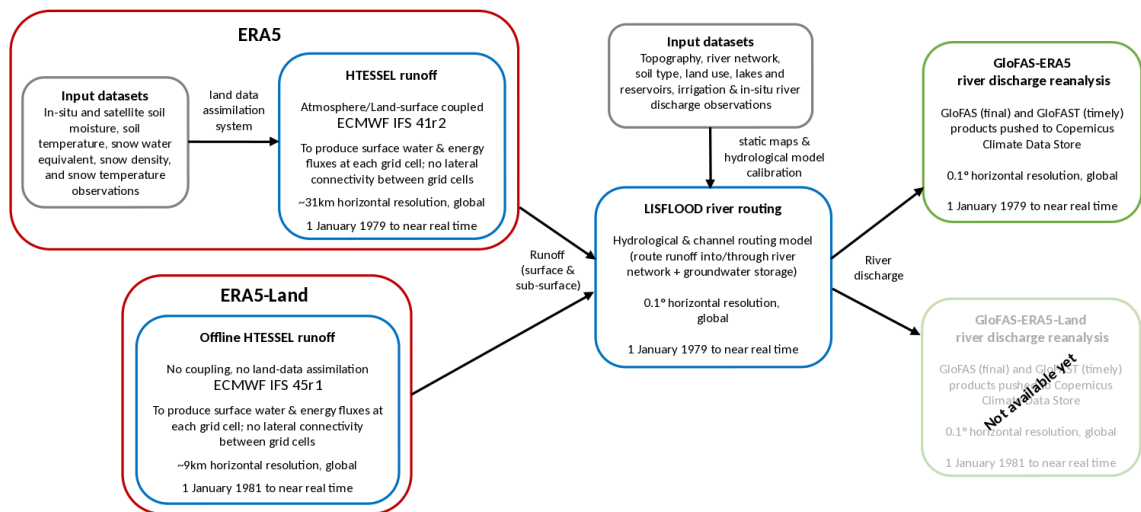


Figure 6-2. Schematic of the key components of the GloFAS-ERA5 river discharge reanalysis dataset production, including the ERA5 and ERA5-Land climate reanalyses. Adapted from Harrigan et al. (2020b).

GloFAS-ERA5 uses surface and sub-surface runoff input from the high-resolution component of the ERA5 reanalysis. It has a fast release version, GloFAS-ERA5T, forced by ERA5T, available 2-5-days behind real time, used in the initialisation of the GloFAS real time forecasts.

A schematic of the key components of the GloFAS-ERA5 and its potential extension GloFAS-ERA5-Land is given in Figure 6-2. GloFAS verification studies often use GloFAS-ERA5, serving as a proxy for river discharge observations. More detailed information on GloFAS-ERA5 is available in Harrigan et al. (2020b).

6.2.4 Trend analysis methodology

Although GloFAS-ERA5 starts in 1979, ERA5-Land data are only available from 1981, and thus the trend analysis is based on the common 1981-2018 period of 38 years. At the time of writing, no

ERA5-Land river discharge is available, and instead annual averages of runoff are used as a proxy, which is a very good estimate of the river discharge in annual means.

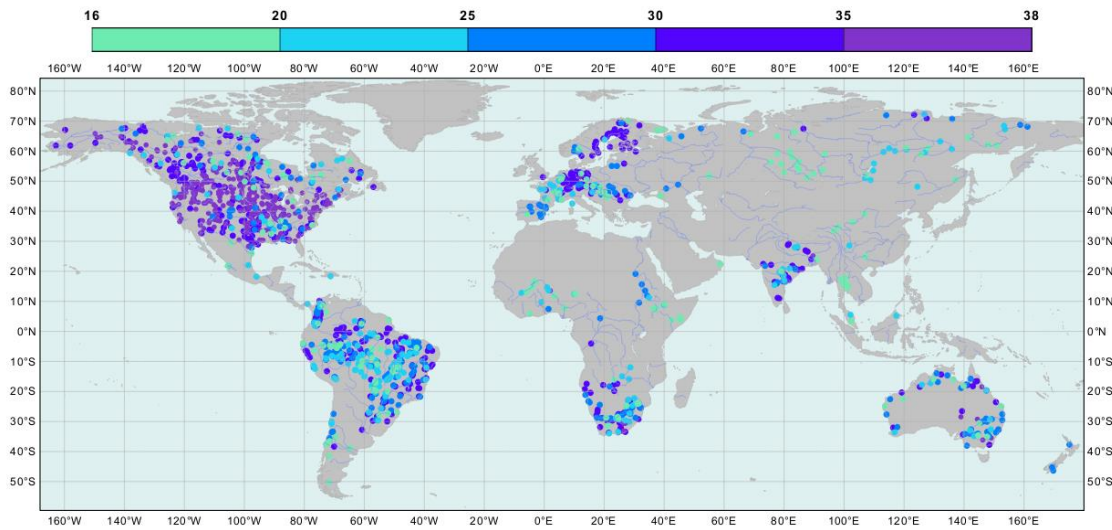


Figure 6-3. Length of observed river discharge records available for the trend analysis, represented at the catchment outlets. A total of 1324 stations with a minimum of 16 years was processed.

6.2.4.1 Choice of catchments

The GloFAS diagnostic catchments are used in this study to analyse the linear trends: a list of 6122 stations with catchment areas varying from about 1.000 km² to 5.4 million km². On a subset of these catchments, where river discharge observations were available (collected and managed by the JRC), the GloFAS-ERA5 river discharge trends are compared to the equivalent trends, determined from observed river discharge, available in the 1981-2018 period (Figure 6-3). A total of 1324 stations were selected with a minimum of 16 years of observed daily data. Unfortunately, large parts of the world are poorly observed with large gaps in space and time in the 38-year period (Lavers et al., 2019; Rodda et al., 1993; Pavelsky et al., 2014).

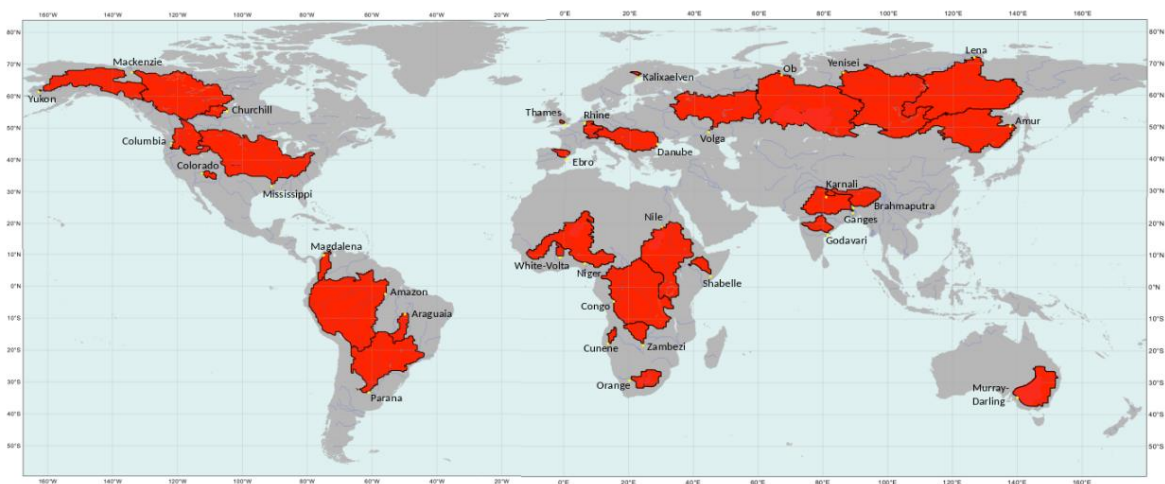


Figure 6-4. Global river catchments where the trends are provided in Table 6-2, Table 6-3 and Table B1 in Appendix B, and where the hydrographs are shown in the Appendix C with the annual mean time series.

Additionally, 33 major world river catchments with good observations are selected for providing detailed analysis of the annual mean time series and linear trend magnitudes (Figure 6-4). The catchments were selected to cover different parts of the world with the longest possible observation time series.

6.2.4.2 Choice of variables

In addition to the simulated and observed river discharge (Table 6-1), the surface variables that directly affect the water budget are analysed from ERA5 and ERA5-Land for linear trends. In the HTESSEL terrestrial water budget, precipitation (P) is the incoming water source. The water can stay in one of the water storage reservoirs or leave the land surface. Water reservoirs are the soil, the canopy interception and in solid form the snowpack (the water stored in the snowpack is the snow water equivalent, SWE). The interception accounts for only a very small fraction of the storage and thus it was left out of the analysis. From the soil, which has four layers in HTESSEL, two water reservoir versions are chosen to be analysed. The top layer (SWV7, 0-7cm), which provides an immediate impact to the atmosphere, and the combination of the top three layers (SWV100, 0-100cm) which represents the slower evolving part of the soil that is still more strongly connected to the atmosphere through the vegetation roots.

The snowfall, the solid part of the incoming precipitation (SF), contributes to building the snowpack. Some part of the rain (liquid part of precipitation) and the water from the melting snowpack (SMLT) leave the land surface system as surface runoff (SRO, the surface fraction of RO). Another fraction of the precipitation leaves as intercepted water evaporation. The remaining water (from the incoming rain and the snowmelt) enters the soil and contributes to the soil water reservoir.

Table 6-1. List of variables analysed for linear trends in this study with their short names, MARS archive codes (see <https://apps.ecmwf.int/codes/grib/param-db/>), the number of processed catchments and a short description.

Short name	Unit	MARS parameter	Number of catchments	Periods	Parameter description
DIS	m^3/s	240.024	5695	1981-2018, 1981-2003 and 2004-2018	GloFAS-ERA5 river discharge
OBS	m^3/s	-	1324	Only 1981-2018 (with variable length)	River discharge observations
DIS-match-OBS	m^3/s	240.024			River discharge observation dates matched GloFAS-ERA5 river discharge
RO-match-OBS	m^3/s	205.128			River discharge observation dates matched runoff outputs of HTESSEL

P	m^3/s	228.128	5698	1981-2018, 1981-2003 and 2004-2018	Precipitation
SF	m^3/s	144.128	3774	1981-2018, 1981-2003 and 2004-2018	Snowfall part of precipitation (same in ERA5 and ERA5-Land apart from downscaling)
RO	m^3/s	205.128	5694	1981-2018, 1981-2003 and 2004-2018	ERA5 sum of surface and sub-surface runoff outputs of HTESEL
RO-Land	m^3/s	205.128		1981-2018, 1981-2003 and 2004-2018	ERA5-Land sum of surface and sub-surface runoff outputs of HTESEL
SMLT	m^3/s	145.128	3789	1981-2018, 1981-2003 and 2004-2018	ERA5 snowmelt output of HTESEL
SMLT-Land	m^3/s	145.128		1981-2018, 1981-2003 and 2004-2018	ERA5-Land snowmelt output of HTESEL
E	m^3/s	182.128	5698	1981-2018, 1981-2003 and 2004-2018	ERA5 evaporation output of HTESEL
E-Land	m^3/s	182.128		1981-2018, 1981-2003 and 2004-2018	ERA5-Land evaporation output of HTESEL
P-E	m^3/s	-	5686	1981-2018, 1981-2003 and 2004-2018	ERA5 precipitation minus evaporation as net water flux to the land-surface in HTESEL
P-E-Land	m^3/s	-		1981-2018, 1981-2003 and 2004-2018	ERA5-Land precipitation minus evaporation as net water flux to the land-surface in HTESEL
SWV7	m^3/s	39.128	5698	1981-2018, 1981-2003 and 2004-2018	ERA5 water content in the top 7cm of the soil (layer 1) in HTESEL
SWV7-Land	m^3/s	39.128		1981-2018, 1981-2003 and 2004-2018	ERA5-Land water content in the top 7cm of the soil (layer 1) in HTESEL
SWV100	m^3/s	39-41.128	5698	1981-2018, 1981-2003 and 2004-2018	ERA5 water content in the top 1 meter of the soil (layers 1, 2 and 3 together) in HTESEL

SWV100-Land	m^3/s	39-41.128		1981-2018, 1981-2003 and 2004-2018	ERA5-Land water content in the top 1 meter of the soil (layers 1, 2 and 3 together) in HTESEL
T2	C°	167.128	5698	1981-2018, 1981-2003 and 2004-2018	ERA5 temperature output at 2 metres in HTESEL
T2-Land	C°	167.128		1981-2018, 1981-2003 and 2004-2018	ERA5-Land temperature output at 2 metres in HTESEL

Some of this water evaporates, either directly from the soil or through the vegetation as transpiration. In total, evaporation in HTESEL (E, where negative E means the land-surface losing water) is the sum of evaporation of the soil and the interception and plant transpiration. Finally, some of the water drains from the soil at the bottom at layer 4 and leave the system as subsurface runoff (SSRO, the subsurface fraction of RO). In order to compare directly with river discharge, the trend analysis is carried out on whole catchment values for each ERA5 and ERA5-Land water budget variable introduced above, after the values on the GloFAS grid are summed together in each of the catchments. This way, essentially the catchment average value is multiplied by the area for the water related variables. The only exception is 2m temperature which was analysed as area average.

6.2.4.3 Trend analysis presentation design

This analysis documents the linear trends in the GloFAS-ERA5 river discharge and in the main ERA5 and ERA5-Land water budget variables (Table 6-1). The trends are shown as maps for all these variables, where the catchment outlets are represented by circles, colour coded by the trend magnitudes, with the size representing the catchment area. In addition, the annual mean time series of all variables are also shown for some major world rivers in the Appendix C, supplemented by some trend related statistics.

As a preliminary check, before the linear trends are computed on the actual river catchments, the global land average annual mean time series of the water cycle variables are analysed for ERA5 (1979-2018) and ERA5-Land (1981-2018). All simulated variables are included in Table 6-1, extended by snow water equivalent (SWE), surface runoff (SRO) and subsurface runoff (SSRO). This step is carried out to identify if there is any major shift or discontinuity in the global time series. All land grid points are used in the averaging, with exception of the snow water equivalent (SWE), where a mask is applied to remove the glaciers, which are represented by a fixed value of 10 metre (of water equivalent) and thus would severely distort the average values.

The linear trends in the GloFAS-ERA5 river discharge simulations are also compared to the linear trends in the observed river discharge. The river discharge trend errors are assumed to be the

difference between the trends in the simulation and the observation annual mean time series. To allow fair comparison, this error computation is done on a special subset of the GloFAS-ERA5 time series that matches the dates when the observations are available (DIS-match-OBS). As trends in runoff and river discharge are expected to be very similar, the linear trend error, i.e. the difference between the runoff and observed river discharge trends, is calculated to assess how ERA5 and ERA5-Land estimate the changes in river discharge. For runoff trend errors, the observation matching time series are used for both ERA5 (RO-match-OBS) and ERA5-Land (RO-Land-match-OBS), similar to DIS-match-OBS. The impact by ERA5-Land is expressed as the difference between the absolute value of the ERA5-Land runoff linear trend error minus the ERA5 runoff linear trend error.

Even though reanalysis systems are designed with the intention of being independent of the changes in the observing system, it is inevitable that the 38-year reanalysis period has some inhomogeneities in the use of the different meteorological observations, including a known major change in the IMS snow observation use which was analysed further. The operational snow analysis was changed in the ECMWF IFS in 2004, when the 24-km Interactive Multi-Sensor Snow and Ice Mapping System (IMS) snow cover information was introduced, in addition to the SYNOP snow depth measurements.

This led to a more realistic representation of the extent of snow cover in the operational analysis (Drusch et al., 2004). Details on ECMWF's snow assimilation can be found in an ECMWF Newsletter article (de Rosnay et al., 2015). In ERA5, the higher resolution (4 km) IMS snow products could be used from 2004, as the high-resolution data was reprocessed and made available to this date.

Moreover, the snow scheme in ECMWF's HTESEL land-surface model is documented to melt the snow too slowly (Dutra et al., 2012). This, in combination with the use of the IMS snow cover data, could lead to a negative shift in ERA5 snowmelt from 2004, as the excess snow that is not melted by the model could then be removed by the assimilation in areas where in situ observations are not available.

This change is expected to produce a clear discontinuity in the snow related time series, and contribute indirectly to creating potential shifts in other variables. This would make it possible that the 38-year-based linear trend would reflect mostly the discontinuity. Therefore, in order to see the trends not dependent on this discontinuity, the 38-year period is split in two parts, 1981-2003 (period1, with 23 years) and 2004-2018 (period2, with 15 years). The linear trends are then computed for both periods in the same way as for the whole 38-year period for all variables other than the observation related ones (see Table 6-1).

6.2.4.4 Criteria for catchment inclusion

The availability of the river discharge observations varies from catchment to catchment and can be as low as 4-5 years (which is the minimum criteria to include them in the GloFAS observation database). For the trend computation a longer minimum length is needed. The criteria of 16 minimum available years, with at least 330 days available in each year, is set as a compromise considering both the observation availability and the minimum record length for reliable regression analysis (Dai et al., 2009).

For all analysed variables (Table 6-1), only catchments which have at least 1 m³/s whole catchment value (or river discharge) as sample mean over the 38-year period (or shorter for the observations) and also the 2-year return period flood thresholds are above 20 m³/s are considered for trend analysis. This filters out the very dry catchments for which trends would not be necessarily meaningful or representative, and similarly filters out snowfall and snowmelt for catchments in the warm climate, where the whole catchment values for these variables will be very small. However, for some of the catchments in the tropical belt the trend could still be computed, for those which have some areas over higher orography and thus some snow contribution. These snow related trends should only be interpreted with caution, as especially in the large catchments, such as the Amazon, the snow has extremely small contribution to the total river discharge.

With these criteria, about 1300 catchments could be used for observed river discharge, roughly 3800 catchments for snowfall and snowmelt and about 5700 catchments for all other water cycle variables (see Table 6-1 for the exact numbers). Regarding the gaps in the available river discharge observation time series, these were simply left out of the analysis, but the catchments that had at least 16 years of data in total (even if gaps in between) were used regardless of the gaps.

6.2.4.5 Definition of trend magnitudes

The trend magnitudes are defined after applying a linear regression to the annual mean time series sample. With this the trends are assumed to be linear. This assumption will not be true for all catchments, however, for the sake of this study, this is considered to be sufficient. Linear trends are expected to show if the time series is impacted by larger changes, discontinuities, etc., which could make the flood threshold computation problematic in GloFAS. To help the relative comparison between catchments and variables, the trend T is defined by (following Stahl et al., 2012):

$$T = (10 \cdot S) / M, \quad (6-1)$$

as the change over a fixed 10-year period, relative to the mean (M) of the n -year period. Here T stands for the trend magnitude at a catchment, S is the slope (annual change as a result of the linear regression), and $M = \text{Mean}(\text{Var}_1, \dots, \text{Var}_n)$, with Var_1 to Var_n denoting the annual mean values from year

1 to year n . The T value gives a measure of the change in a decade. For example, a value of ± 0.1 effectively means the variable increased (decreased) by 10% of the sample mean value over the course of the 10 years. The 10-year fixed period is chosen as a common ground to allow trend comparability for catchments with different observation length.

For 2m temperature, as the only non-water-related variable, the intercomparability with other variables is less important, and the division by the mean (M) would cause problems (being near 0 in some areas of the world), therefore the trend magnitude is defined as the temperature change in 10 years along the linear regression line ($T=10*S$, where M is replaced by 1).

An alternative trend magnitude was also calculated for all variables, including 2m temperature, by calculating the linear regression on the standardised variable (each annual mean value divided by the standard deviation of the annual mean time series) and T defined by the $10*S$. The standardised trend improves the comparability across variables with very different value ranges (e.g. river discharge and evaporation) and suffer less from the potential issue of division by near 0 values (as it can happen in some isolated cases for P-E). However, the M -based trend (Eq. 6-1) is more intuitive when interpreting the size of the trends and therefore was selected as the focus of our analysis. In the rest of the paper the 'raw' trend, i.e. the original definition of the linear trend with the raw variable (in Eq. 6.1), is analysed and displayed, with only few exceptions where the standardised trend is mentioned.

In Eq. 6-1, n is either 38 (1981-2018), 23 (1981-2003) or 15 (2004-2018) for all variables other than the river discharge observation related variables (OBS, DIS-match-OBS, RO-match-OBS and RO-Land-match-OBS; see in Table 6-1) where it varies between 16 and 38 for the catchments, depending on observation availability.

6.3 Results

6.3.1 Global land average annual mean time series for ERA5 and ERA5-Land

The discontinuity in the use of the IMS product from 2004 is a known issue in ERA5. This is clearly present in the global land average annual mean time series of the snow, both for SWE and SMLT (Figure 6-5). For snowmelt, the change happens from 2004, while for SWE from mainly 2005. This discontinuity, however, seems to be mostly embedded in a general decreasing trend for both variables. ERA5-Land, which does not have land data assimilation as in ERA5, does not show any sign of this change in 2004.

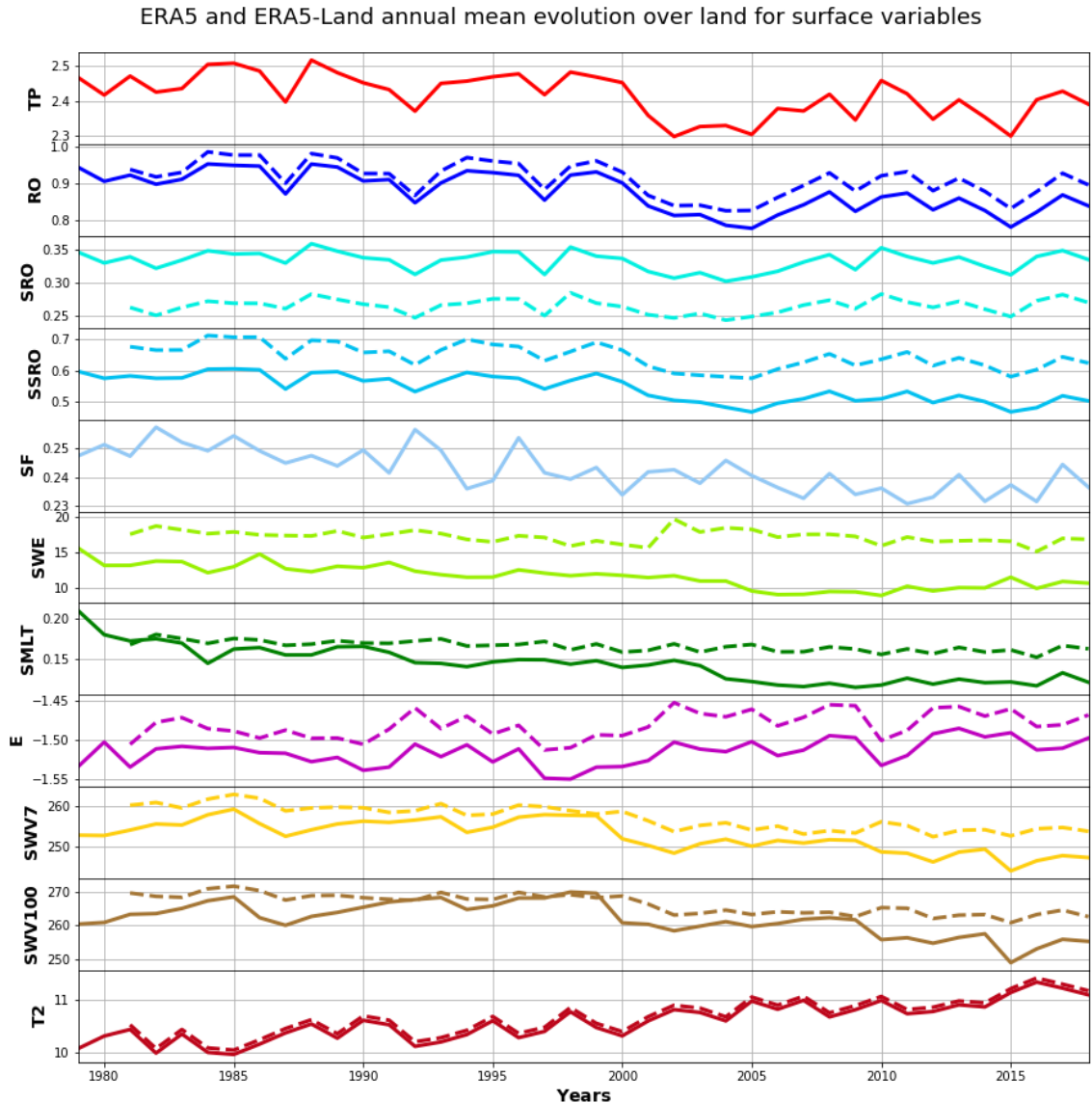


Figure 6-5. Annual mean time series of global land averages for precipitation (TP), runoff (RO), surface runoff (SRO), subsurface runoff (SSRO), snowfall (SF), snow water equivalent (SWE), snowmelt (SMLT), evaporation (E), soil water content in the top 7 cm (SWV7) and 100 cm (SWV100) and finally 2m temperature (T2) from ERA5 (solid lines) and ERA5-Land (dashed lines) for the 1979-2018 period (ERA5-Land is only available from 1981). TP and SF are the same in ERA5 and ERA5-Land (apart from downscaling to higher resolution in ERA5-Land), so only displayed for ERA5. All variables' unit is mm/day other than 2m temperature which has °C.

Another very clear shift is seen for precipitation and all the runoff variables in the 1999-2004 period, when these variables suffer a very large drop (Figure 6-5). There is no such tendency in snowfall, which suggests that the source of this drop must come mainly from warm climate areas where snow is not dominant.

Finally, as expected, the strong upward tendency is present for 2m temperature, with over 1 degree difference between the start and end years of the 40-year period (Figure 6-5).

ERA5-Land is slightly warmer and generally has more water in the water cycle. It produces more runoff, even though as a balance of much less surface runoff and lot more subsurface runoff. There is more snow in the snowpack with consequently more snowmelt. The soil also has more water but finally less evaporation than in ERA5 (Figure 6-5).

6.3.2 Trends in GloFAS-ERA5 river discharge

Over the 1981-2018 period, GloFAS-ERA5 river discharge shows a dominantly negative raw linear trend with almost 80% of catchments exhibiting negative trend magnitudes and about 40% showing a negative value stronger than -0.1 (fraction/decade), i.e. decreasing at least by 10% of the 1981-2018 mean value across the 10-year reference period (Figure 6-6).

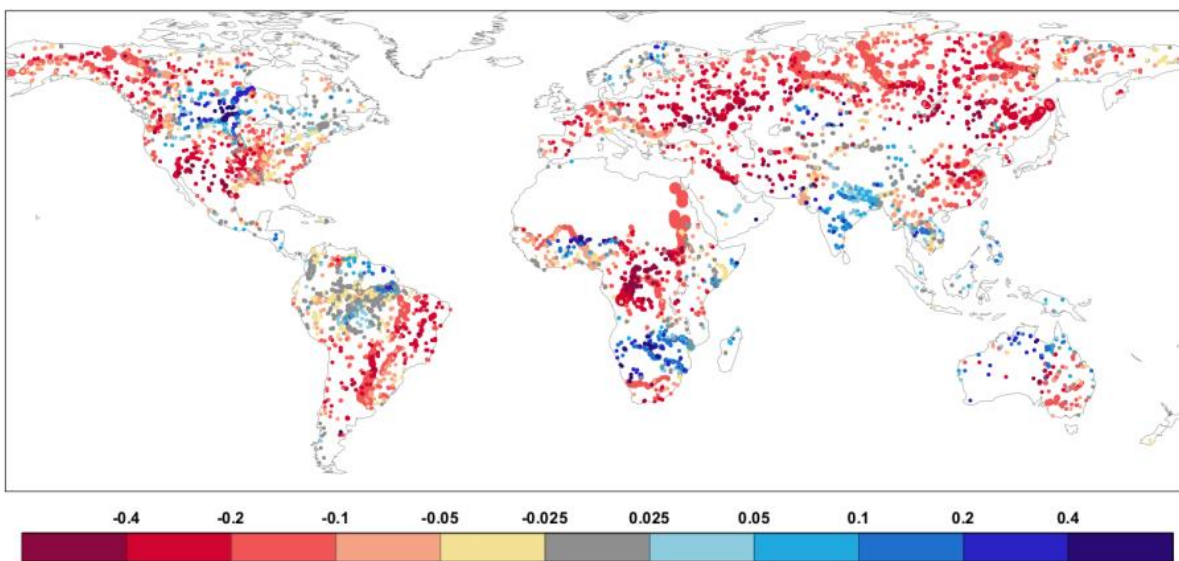


Figure 6-6. Raw linear trends (fraction/decade) at global river catchments for GloFAS-ERA5 river discharge, based on the 1981-2018 period. The circles represent the catchment outlets, while their size the catchment area.

The most negative river discharge trends are found in many of the larger world rivers such as: the Congo and the Nile in Africa; the Ob, Lena, Yenisei and Amur in Russia; the Dnieper and Volga in Europe; the Colorado, Mackenzie and the Yukon in North America; the Yellow and the Yangtze in China; the Tiger and Euphrates in the Middle East and the Sao Francisco, Tocantins and Paraguay in eastern South America.

Positive linear discharge trends can be seen in a few smaller areas, most notably around larger rivers like the Zambezi in South Africa, the Ganges in India and the Nelson in central Canada.

According to the DIS column in Table 6-2 (and also Table B1 with the standardised trends in Appendix B), only 8 out of 33 catchments show a positive trend (~25%), while 18 catchments (~55%) reveal larger trends at least with 10% decrease in the 10 years (or more than 1/3 of the standard deviation of the annual mean time series for the standardised trend in Table B1 in Appendix B), making it a very substantial drop in GloFAS-ERA5 river discharge for the whole 1981-2018 period.

Table 6-2. Raw linear trends (fraction/decade) for selected catchments (see Figure 6-4) for 1981-2018, for GloFASERA5 river discharge (DIS), ERA5 precipitation (P) and snowfall (SF) and both ERA5 and ERA5-Land snowmelt (SMLT), runoff (RO), evaporation (E), precipitation minus evaporation (P-E), soil moisture in the top 7 cm (SWV7) and 100 cm (SWV100) and 2m temperature (T2). Raw linear trends are also provided for observed river discharge (OBS) and the observation matched GloFAS-ERA5 river discharge (DISm) and ERA5 and ERA5-Land runoff (ROm). Differences in the absolute raw linear trend errors between ERA5-Land and ERA5 are also indicated (Imp). Empty cells correspond to cases for which trend computation was not possible. Coloured cells indicate negative (orange) and positive (blue) trends and decreasing (green) and increasing (purple) trend errors in ERA5-Land (Imp column). Where there is no raw trend, defined for absolute values less than 0.025, cells are not coloured. Darkening shades show increasing trend magnitudes.

Rivers		DIS	P	SF	SMLT		RO		E		P-E	
		ERA5	ERA5	ERA5	ERA5	ERA5-Land	ERA5	ERA5-Land	ERA5	ERA5-Land	ERA5	ERA5-Land
North Asia	Ob	-0.15	-0.01	-0.03	-0.13	-0.04	-0.11	-0.06	-0.023	-0.018	-0.07	-0.06
	Yenisei	-0.16	-0.02	-0.03	-0.18	-0.03	-0.15	-0.07	-0.016	-0.017	-0.07	-0.08
	Lena	-0.14	0.00	-0.03	-0.24	-0.04	-0.11	-0.04	-0.019	-0.023	-0.04	-0.05
	Amur	-0.26	-0.04	-0.04	-0.39	-0.07	-0.20	-0.16	-0.007	-0.004	-0.16	-0.17
North America	Yukon	-0.14	-0.01	-0.03	-0.13	-0.03	-0.13	-0.03	-0.001	-0.014	-0.01	-0.03
	Mackenzie	-0.12	0.00	-0.01	-0.05	0.00	-0.11	-0.02	-0.008	-0.009	-0.02	-0.03
	Churchill	0.12	0.03	0.03	0.07	0.02	0.07	0.07	-0.009	-0.009	0.15	0.19
	Columbia	-0.11	-0.02	0.01	-0.06	0.01	-0.10	-0.01	0.038	0.021	0.01	-0.01
	Colorado	-0.42	-0.13	-0.16	-0.15	-0.16	-0.36	-0.43	0.143	0.120	-0.05	-1.05
	Mississippi	-0.09	-0.02	-0.06	-0.08	-0.06	-0.08	-0.07	-0.005	0.006	-0.12	-0.07
Europe	Kalixaelven	0.02	0.03	0.02	-0.06	0.02	0.01	0.05	-0.022	-0.024	0.04	0.04
	Thames	-0.11	0.01	-0.12	0.10	-0.10	-0.11	0.00	0.003	-0.013	0.05	-0.01
	Rhine	-0.12	-0.03	-0.08	-0.15	-0.07	-0.10	-0.08	-0.019	-0.023	-0.08	-0.09
	Danube	-0.09	0.01	-0.04	-0.13	-0.05	-0.08	-0.03	-0.019	-0.026	0.00	-0.02
	Volga	-0.27	-0.03	-0.03	-0.12	-0.03	-0.22	-0.13	-0.024	-0.013	-0.16	-0.14
	Ebro	-0.20	0.00	0.04	-0.17	0.04	-0.15	-0.02	0.059	-0.002	0.21	0.00
South America	Magdalena	-0.03	-0.02				-0.04	-0.03	0.001	-0.008	-0.03	-0.04
	Amazon	-0.02	0.00	-0.12	-0.12	-0.12	-0.02	-0.01	-0.001	-0.004	-0.01	-0.01
	Araguaia	-0.15	-0.06				-0.13	-0.13	0.004	0.015	-0.18	-0.13
	Parana	-0.17	-0.05	-0.15	-0.16	-0.16	-0.14	-0.16	0.005	0.019	-0.21	-0.17
Africa	Nile	-0.11	-0.07				-0.10	-0.12	0.029	0.051	-0.22	-0.15
	Niger	-0.06	-0.02				-0.05	-0.07	-0.028	0.007	-0.35	-0.06
	Shabelle	-0.04	-0.03				-0.04	-0.03	0.021	0.027	-0.08	-0.03
	White-Volta	0.14	-0.03				0.12	0.07	-0.044	0.038	-0.67	0.25
	Congo	-0.30	-0.08				-0.25	-0.28	0.005	0.008	-0.33	-0.30
	Zambesi	0.08	0.01				0.06	0.08	0.022	0.010	0.12	0.06
	Cunene	0.12	0.05				0.09	0.15	0.045	-0.048	1.25	0.09
	Orange	-0.20	-0.07	-0.02	-0.04	-0.03	-0.16	-0.17	0.090	0.048	7.83	-0.47
South Asia	Brahmaputra	-0.02	-0.02	-0.05	-0.01	-0.05	-0.02	-0.03	-0.012	-0.007	-0.03	-0.03
	Ganges	0.07	0.03	-0.05	0.02	-0.04	0.06	0.05	-0.011	-0.006	0.06	0.05
	Godavari	0.10	0.05				0.10	0.11	-0.018	-0.015	0.13	0.12
	Karnali	0.03	0.01	-0.08	-0.05	-0.07	0.03	0.01	-0.012	-0.008	0.01	0.01
Australia	Murray	-0.11	-0.07	-0.13	-0.19	-0.13	-0.10	-0.14	0.035	0.056	-0.11	-0.32

(continues on next page)

Rivers		SWV 0-7		SWV 0-100		T2		OBS	DISm	ROm		
		ERA5	ERA5-Land	ERA5	ERA5-Land	ERA5	ERA5-Land	-	ERA5	ERA5	ERA5-Land	Imp
North Asia	Ob	-0.006	-0.006	-0.007	-0.009	0.22	0.22	0.02	-0.16	-0.12	-0.04	-0.081
	Yenisei	-0.010	-0.010	-0.006	-0.014	0.36	0.35	0.04	-0.15	-0.14	-0.05	-0.097
	Lena	-0.004	-0.001	-0.002	-0.001	0.37	0.36	0.04	-0.17	-0.12	-0.06	-0.065
	Amur	-0.024	-0.019	-0.012	-0.024	0.32	0.31	-0.33	-0.65	-0.52	-0.40	-0.128
North America	Yukon	-0.009	-0.002	-0.014	-0.001	0.40	0.39	0.02	-0.20	-0.18	-0.04	-0.140
	Mackenzie	-0.006	-0.003	-0.007	-0.003	0.34	0.33	0.04	-0.12	-0.10	-0.01	-0.095
	Churchill	0.003	0.006	0.007	0.008	0.13	0.12	0.20	0.22	0.11	0.07	0.045
	Columbia	-0.027	-0.015	-0.027	-0.011	0.34	0.32	-0.02	-0.13	-0.11	-0.03	-0.078
	Colorado	-0.110	-0.094	-0.104	-0.057	0.43	0.44	-0.27	-0.43	-0.37	-0.45	0.073
	Mississippi	-0.014	-0.016	-0.007	-0.019	0.24	0.25	0.03	-0.11	-0.10	-0.09	-0.006
Europe	Kalixaelven	0.004	-0.001	0.004	0.000	0.48	0.51	0.03	-0.13	-0.10	0.03	-0.125
	Thames	-0.009	-0.001	-0.022	-0.004	0.26	0.25	0.00	-0.20	-0.18	-0.06	-0.118
	Rhine	-0.011	-0.010	-0.015	-0.012	0.40	0.39	-0.06	-0.14	-0.12	-0.07	-0.043
	Danube	-0.010	-0.006	-0.017	-0.006	0.45	0.45	0.08	-0.09	-0.07	0.00	-0.063
	Volga	-0.015	-0.016	-0.015	-0.019	0.35	0.35	-0.02	-0.32	-0.27	-0.14	-0.133
Ebro	-0.032	-0.011	-0.075	-0.013	0.29	0.28	-0.07	-0.36	-0.29	-0.13	-0.167	
South America	Magdalena	-0.012	-0.008	-0.016	-0.008	0.21	0.21	0.01	0.00	-0.01	0.00	-0.003
	Amazon	-0.010	-0.009	-0.011	-0.009	0.19	0.19	0.02	-0.04	-0.04	-0.04	-0.005
	Araguaia	-0.021	-0.022	-0.013	-0.016	0.30	0.32	-0.03	-0.11	-0.11	-0.11	-0.002
	Parana	-0.023	-0.027	-0.021	-0.031	0.27	0.29	-0.05	-0.19	-0.16	-0.18	0.015
Africa	Nile	-0.029	-0.033	-0.023	-0.035	0.38	0.39	0.12	-0.13	-0.12	-0.14	0.026
	Niger	0.000	-0.008	0.015	-0.006	0.27	0.28	0.28	-0.09	-0.06	-0.10	0.032
	Shabelle	-0.017	-0.018	-0.006	-0.006	0.23	0.24	0.27	-0.12	-0.12	-0.11	-0.013
	White-Volta	-0.005	-0.019	0.044	0.000	0.17	0.21	-0.08	-0.12	-0.10	-0.14	0.033
	Congo	-0.023	-0.024	-0.028	-0.034	0.28	0.28	0.05	-0.29	-0.24	-0.27	0.032
	Zambesi	-0.013	-0.007	-0.009	0.001	0.20	0.19	0.22	0.10	0.08	0.10	-0.020
	Cunene	-0.007	0.018	-0.028	0.036	0.22	0.18	0.21	0.16	0.14	0.19	-0.057
Orange	-0.036	-0.030	-0.064	-0.025	0.27	0.26	0.13	-0.54	-0.40	-0.36	-0.034	
South Asia	Brahmaputra	0.007	0.003	0.007	0.003	0.21	0.21	-0.05	-0.02	-0.02	-0.03	-0.010
	Ganges	0.009	0.004	0.011	0.007	0.14	0.14	-0.01	0.08	0.07	0.07	-0.002
	Godavari	0.009	0.006	0.014	0.014	0.05	0.05	-0.02	0.16	0.15	0.16	0.015
	Karnali	0.000	-0.007	0.003	-0.004	0.30	0.28	-0.01	0.04	0.04	0.02	-0.018
Australia	Murray	-0.020	-0.032	-0.032	-0.029	0.30	0.31	-0.28	-0.18	-0.14	-0.16	-0.028



6.3.3 Trends in precipitation, snowfall and 2m temperature

Precipitation shows the same trend signs to river discharge over much of the world (Figure 6-7a). The raw linear trend magnitudes are generally smaller than in river discharge, but a lot of the differences are in fact related to the larger M-term in Eq. 6-1, as precipitation is generally much larger than river discharge. The standardised linear trends highlight (see Table B1 for the selected global catchments in Appendix B) that in tropical and subtropical areas of the world, including Africa, Central and South America, South Asia and Australia, precipitation and river discharge trends are very similar, while over the Northern Extratropics river discharge trends tend to be more pronounced than precipitation.

The largest linear trends (both raw and standardised) in precipitation are for the Colorado and Rio Grande basins in the United States and also the Congo and upper Nile basins in Central Africa, where the negative raw trend is between -0.1 and -0.2 (10-20% decrease in just 10 years, see also the P

column in Table 6-2; or over 0.5 standard deviation of the annual mean time series sample, see Table B1 with the standardised trends in Appendix B).

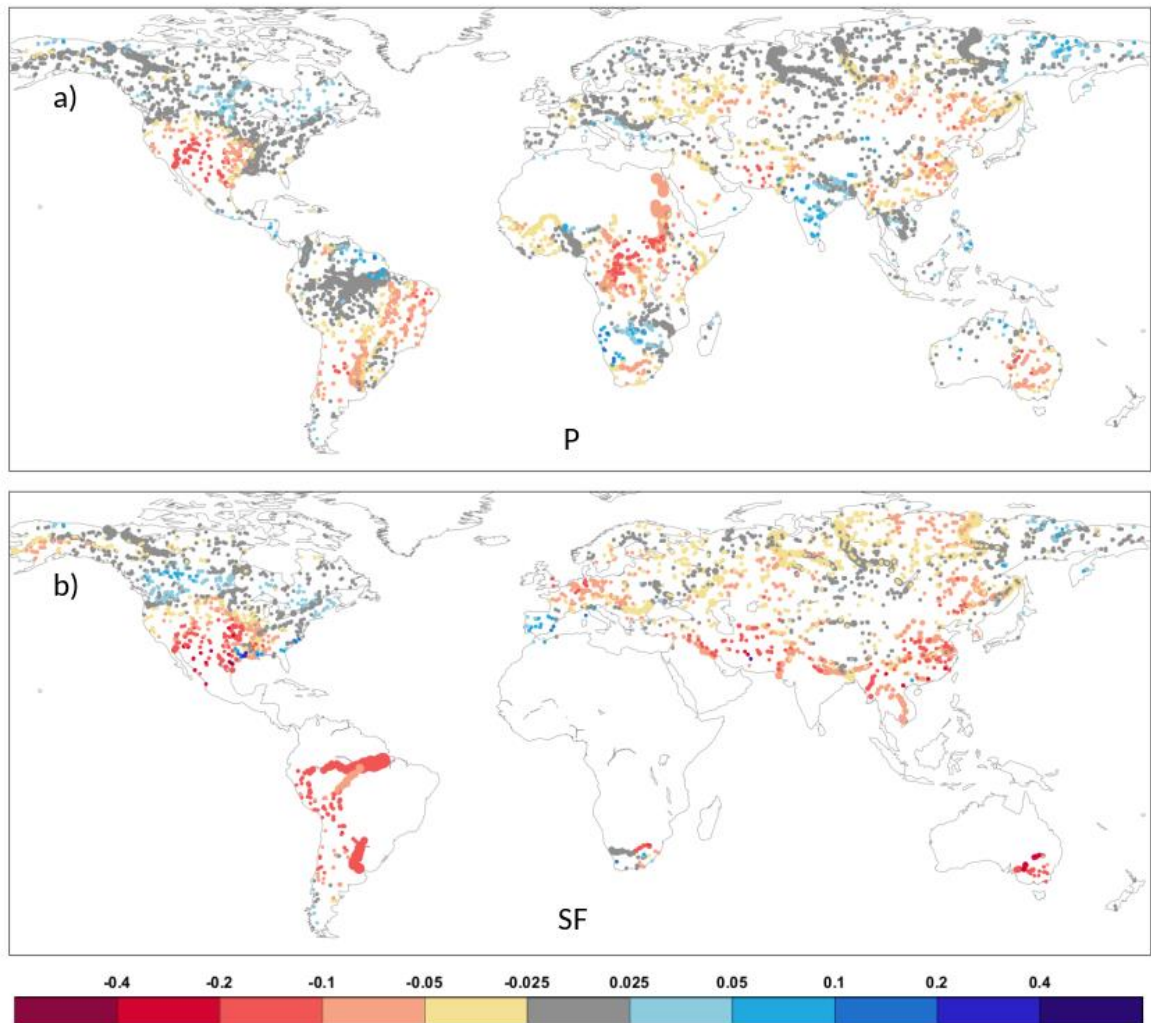


Figure 6-7. Raw linear trends (fraction/decade) at global river catchments for ERA5 a) Precipitation (P) and b) Snowfall (SF), based on the 1981-2018 period. The circles represent the catchment outlets, while their size the catchment area.

Although some of the tropical and subtropical basins, mainly the rivers downstream of the Andes in South America and areas in South Asia, show clearly more negative trends, the snowfall contribution to precipitation is usually very little in those areas (Figure 6-7b). On the other hand, in the more snow dominant higher latitude areas of the Northern Hemisphere, some areas such as Alaska, central parts of Europe or northern parts of Russia, there is a moderate tendency to more negative trends.

The main reason for the decreasing snowfall is very likely to be the generally increasing temperature. Figure A1 in the Appendix A highlights that the 2m temperature trend is positive almost everywhere. The largest positive trends, above 0.4 degree change in 10 years, are in the Nile basin in Africa, the southwest part of the USA, the eastern parts of Europe, the Middle-East and

also many of the smaller rivers in the northern-most latitudes. Trends in ERA5 and ERA5-Land are generally similar, if anything, ERA5-Land tends to be slightly warmer (see the annual mean time series for the selected global river basins in Appendix C), but the trends are almost the same everywhere. These differences can originate from the snow cover and evaporation processes which can differ in coupled (ERA5) vs offline (ERA5-Land) systems.

These large global temperature increases in ERA5 are consistent with the scientific literature (e.g. Hansen et al., 2006; Parmesan and Yohe, 2003, IPCC, 2007 and 2014).

It is difficult to determine how realistic the identified ERA5 precipitation and snowfall trends are, due to the sparse precipitation observing network and the highly variable quality of the satellite derived precipitation data (Sun et al., 2018).

The AR4 (IPCC, 2007) and AR5 (IPCC, 2014) IPCC reports disagree on precipitation trends in many parts of the world, the worst being West Africa where significant positive (AR4) trends turn to significant negative (AR5). Part of the reason is the different periods (1979-2005 in AR4 - Figure 3.13 vs. 1951-2010 in AR5 - Figure 1.1), however, the shortness of the first period (through capturing more of the natural climate variability as trends) and the differences between the used data sets must have also contributed. The estimation uncertainties in precipitation changes were acknowledged in the AR5 report when it concluded that 'Confidence in precipitation change averaged over global land areas since 1901 is low prior to 1951 and medium afterwards'. In fact, neither of the two IPCC reports are similar to the ERA5 precipitation linear trends presented here. In particular, there is no clear sign of either of the two most prominent trend areas in ERA5. While the southwestern USA is represented in AR4, although with smaller negative values in AR4 than in ERA5 (3-10% vs. 10-20%), the large negative ERA5 trends in central Africa are not there in any of these two sources.

Nguyen et al. (2018) analysed precipitation trends with satellite derived data in the 1983-2015 period, which is directly comparable with our period. They concluded that although only few percent of the land mass show pixel-by-pixel significant trends, this increases by regional- or catchment-based analysis, but even on the large catchment-scale (over 200 large rivers), only a smaller fraction has significantly large trends. The catchment-scale precipitation trends in Nguyen et al. (2018) show a lot of similarities, at least in sign, to the ERA5 precipitation linear trends in Figure 6-7a. For example, the trend patterns are similar in most parts of Australia, Central and South Africa, South America, but also in Europe and large parts of Asia. Major differences are present mainly in Central America, around Canada and Central Asia where the trends show mainly opposite signs between ERA5 (Figure 6-7a) and the satellite-derived precipitation in Nguyen et al. (2018).

The dominantly decreasing trend in ERA5 is not supported by either of GPCC or GPCP, two of the available global precipitation estimate data sets, as described in Hersbach et al. (2020); see Figure 23. ERA5 seems to produce more precipitation than either GPCC or GPCP, and the difference gets smaller from 2000. It seems, after the large decline around 2000-2002 (see also Figure 6-5), the ERA5 precipitation is more realistic in the 21st century, which could potentially come from some changes in the used satellite data in ERA5. Nevertheless, further analysis is going to be needed in the future to better understand the behaviour of the ERA5 precipitation changes.

6.3.4 Trends in runoff and snowmelt

The linear trends for runoff in ERA5 (Figure 6-8a) are almost identical to the GloFAS-ERA5 river discharge trends (Figure 6-6). This is expected, as the annual mean values of these two variables can usually differ only little, the river routing's time delay is averaged out by computing the mean over the whole year. However, the trends for runoff in ERA5-Land (Figure 6-8b) are different from ERA5 in the Northern Extratropics, namely the ERA5 trends are more negative in Alaska, western Canada and most of northern Eurasia. As the atmospheric forcing are the same in ERA5 and ERA5-Land, and the land-surface model is also mainly the same, differences in runoff and other surface variables will come from the missing coupling and land data assimilation and the much higher resolution and lapse-rate correction in ERA5-Land. The land data assimilation impact on the hydrological cycle can be substantial, considering both snow and soil moisture, as shown in Zsoter et al. (2019), and the resolution change, through the different temperature conditions with the lapse rate correction, is also expected to have a potentially large impact.

As the ERA5 and ERA5-Land runoff trends are very similar in the tropics and subtropics (see Figure 6-8 and also the bottom half of Table 6-2 and Table B1 in Appendix B), the likely culprit for the differences in the higher latitudes is the handling of the snow with the possible differences in snowmelt.

The runoff raw linear trend differences seem to come from snowmelt which are very different in ERA5 (Figure 6-9a) and ERA5-Land (Figure 6-9b). The snowmelt trends in ERA5 are much more negatively oriented in the Northern Extratropics such as the Lena, Amur (in Russia) or the Yukon (in Alaska) rivers. The exceptions are the central part of North America and some catchments in Northern Europe and near the Himalaya, where ERA5 has more positive trends (see the Churchill and Thames rivers in Table 6-2 and Table B1 in Appendix B; SMLT columns), where snowmelt appears to have increased in the last ~40 years in ERA5. These ERA5 snowmelt linear trends can be partially explained by the changes in snowfall, i.e. the amount of snow available to melt, as the sign of the snowfall and snowmelt trends is the same over large parts of the world. However, the

magnitude is clearly different in the Northern Hemisphere, with the snowmelt trends being lot more pronounced.

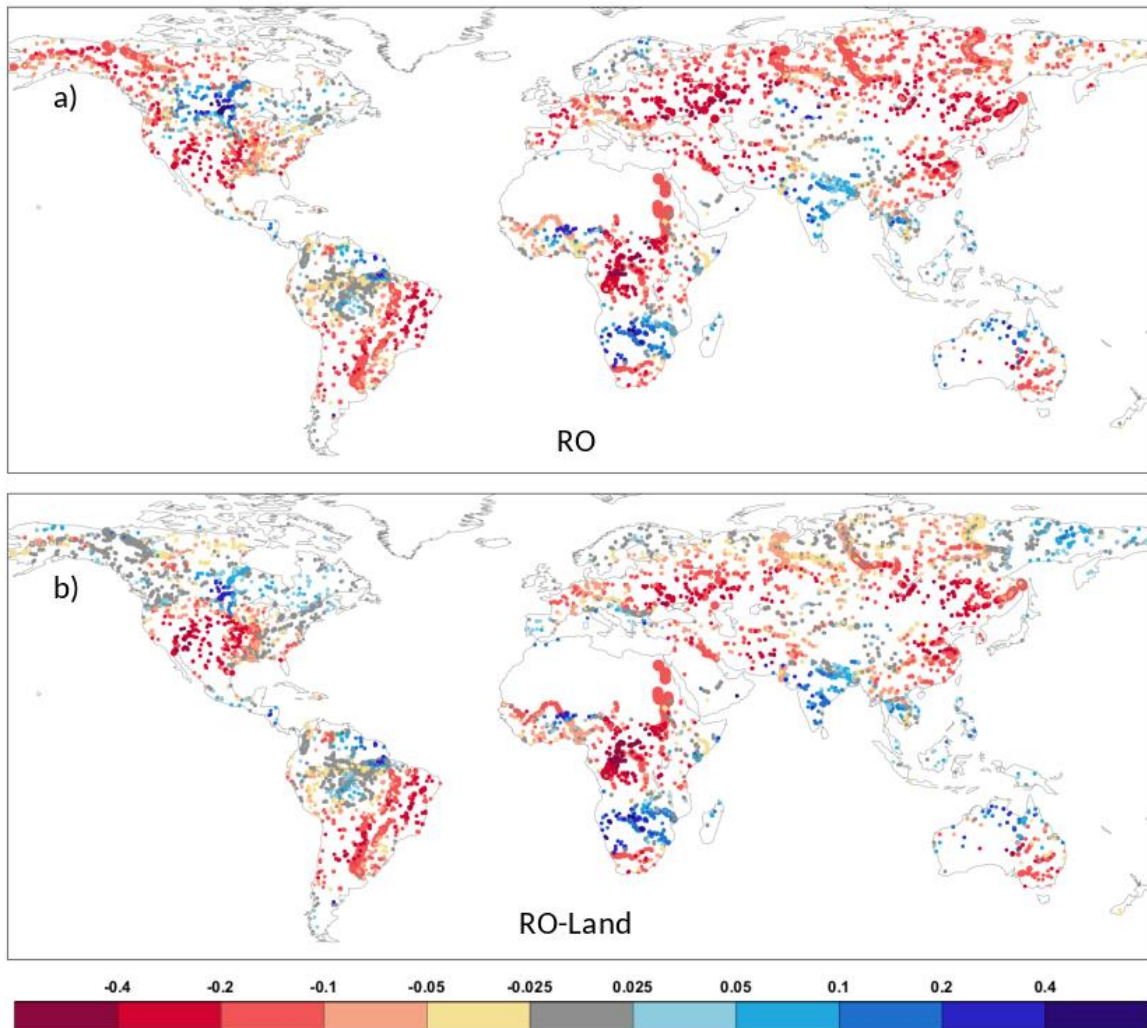


Figure 6-8. Raw linear trends (fraction/decade) at global river catchments for runoff in a) ERA5 (RO) and b) ERA5-Land (RO-Land), based on the 1981-2018 period. The circles represent the catchment outlets, while their size the catchment area.

The scientific literature has documented trends in the snowpack related variables (snowfall, snow depth, snow cover) extensively. Studies are either based on available in situ observations, or satellite derived measurements. However, only snow cover extent (e.g. the IMS snow cover product used at ECMWF) seems to be reliable enough to be used quantitatively from satellites and snow water equivalent (SWE) and snow depth (SD) have higher uncertainty and thus offer limited help (Hancock et al., 2013).

The 5th IPCC report states that there is very high confidence that the extent of Northern Hemisphere snow cover has decreased since the mid-20th century (IPCC, 2013, see Figure SPM.3). The declining snow cover is strongly related to the increasing global temperatures.

For example, Kunkel et al. (2016) provided a summary of several snow climatology studies and also found a strong decrease in maximum seasonal snow depth in the studied North America and Europe. Connolly et al. (2019) evaluated the snow cover extent trends in the Northern Hemisphere based on the Rutgers University snow cover dataset and found that the trend in the satellite-derived observations is poorly explained by the CMIP5 climate models, the models exhibiting a steadier decline during the 1967-2018 period.

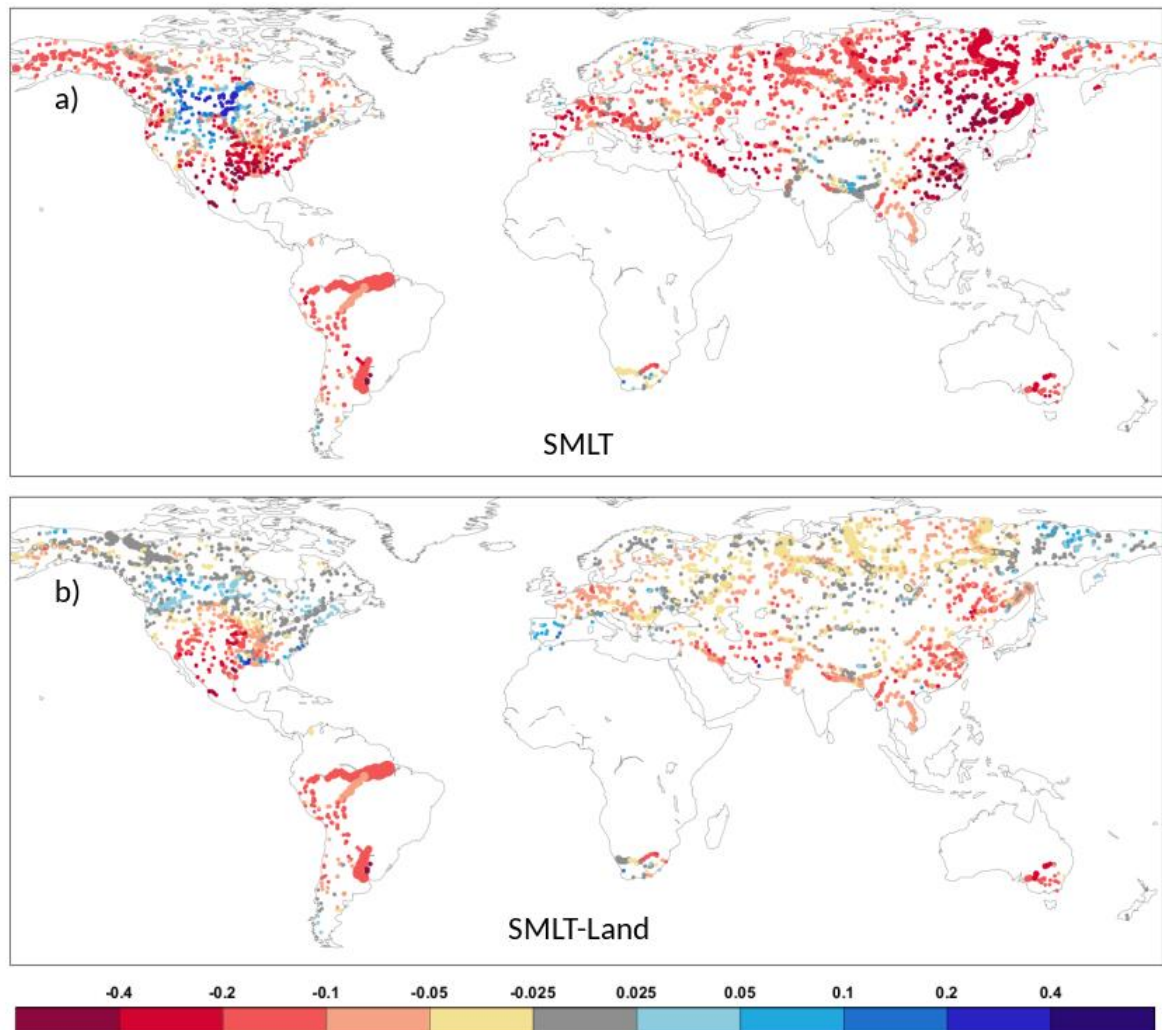


Figure 6-9. Raw linear trends (fraction/decade) at global river catchments for snowmelt in a) ERA5 (SMLT) and b) ERA5-Land (SMLT-Land), based on the 1981-2018 period. The circles represent the catchment outlets, while their size the catchment area.

In our study no snow cover extent is analysed, however, both the snow water equivalent and snowmelt similarly show a steady decline over the global land areas in Figure 6-5, at least until 2004. Similarly, Knowles (2015) evaluated the trends in the GHCND climate observations for snow related variables in the 1950-2010 period for the United States. It was found that both snowfall and snow depth showed more negative than positive trends, which is in agreement with our analysis,

even though the geographical distribution of their trends is different to the linear trends documented for snowfall and also for snowmelt in this study (Figure 6-7b and Figure 6-9).

It can be concluded that the ERA5-Land runoff linear trends (the signs at least) are generally similar to the precipitation trends everywhere in the world. In ERA5, however, the runoff (and thus river discharge) trends seem to be dominated by precipitation in the tropics and subtropics, while the trends in higher latitudes resemble the snowmelt trends, as a consequence of the additional land data assimilation (see the SMLT and RO columns in Table 6-2 and Table B1 in Appendix B). Note that the attribution of the ERA5/ERA5-Land trend differences to the observed and simulated trends in existing studies, and the interpretation of these global and regional differences, were beyond the scope of this study, but would be a very important further piece of analysis.

6.3.5 Trends in evaporation, precipitation minus evaporation and soil moisture

The evaporation (E) and the two soil moisture variables (top 0-7 cm, SWV7 and top 0-100 cm, SWV100 layers) all show smaller raw linear trends for both ERA5 and ERA5-Land (Figure A2 and A4-5 in Appendix A). The smaller trend magnitudes mainly come from the much larger volume in these variables, thus a larger M term in Eq. 6-1 and thus a smaller trend magnitude (T) in relative terms. This is supported by Table B1 in Appendix B which shows that the standardised trends, defined by using the standardised variables, are generally in the same magnitude range as the other variables. More noticeable trends (both raw and standardised) are present e.g. in the Nile basin and Middle East regions or the southwestern United States area for evaporation and the soil water content variables, both showing decrease. Please be aware that for evaporation positive trend (such as in these areas) actually means increase of negative values, which equals to decreasing evaporation (as the analysed evaporation variable is generally negative everywhere).

The scientific literature agrees that during the last few decades the global land evaporation has generally increased (e.g. Jung et al., 2010; Zhang et al., 2016; Anabalon and Sharma et al., 2017). However, there seems to be evidence that this stopped in the 1998-2008 period (Jung et al., 2010), and then evaporation is relatively stable since (Javadian et al., 2020).

In contrast, the ERA5 and ERA5-Land global land average evaporation trend (in Figure 6-5) is more positive than negative, somewhat contrary to the literature (showing slightly decreasing absolute values). However, the first period shows a small increase until 1998, followed by a marked decrease and then mainly no change in the last 10 years which is broadly similar to what is reported in the literature.

ERA5/ERA5-Land agrees with the large negative evaporation trends shown by Zhang et al. (2016) in the Middle East, western United States and the generally positive trends in the northern latitudes

and also in Southeast Asia. However other areas show marked differences, especially parts of Africa and much of Australia.

Soil moisture has been shown to be generally decreasing in the last few decades by several studies (e.g. Feng and Zhang et al., 2016; Albergel et al., 2013; Pan et al., 2019; Dorigo et al., 2012), agreeing with the dominantly negative trends documented in this study. About 30% of land is shown to have significant trend in these studies, a majority of being negative. However, there are large differences in the actual pattern depending on the data set used. For example, as shown in Albergel (2013), the ERAI-Land, MERRA-Land and SM-MW (a microwave-based multisatellite surface soil moisture dataset) all show marked differences in the trend patterns. Similarly, the ERA5 and ERA5-Land soil moisture trends show notable similarities to the SM-MW trends only in Central Asia, the Middle East, central South America, otherwise the match is poor.

The precipitation minus evaporation (P-E), i.e. the water source to the land-surface, and the two analysed soil moisture variables show roughly similar trend signs (see Figure A3-A5 in appendix A). Moreover, ERA5 and ERA5-Land are broadly similar for all these four variables and only exhibit a few regional variations, most notably in Africa where they change noticeably in ERA5-Land. This happens over areas like the Niger, White-Volta and Cunene rivers, where evaporation even changes trend sign from negative to positive (Niger, White-Volta), which actually means decreasing amount of water leaving the land-surface through evaporation in ERA5-Land, as evaporation is dominantly negative over the world, or the Cunene river where evaporation changes from positive to negative, or the Nile river where the positive trend gets more pronounced, all these changes coinciding with also large swings in soil water content (see Table 2 and Table B1 in Appendix B; the E, P-E, SVW7 and SWV100 columns).

6.3.6 Trends in observed river discharge

The obvious question about the large trends in GloFAS-ERA5 river discharge is whether they are also present in the observations. Figure 6-10 highlights the match between the simulated and observed raw linear river discharge trends.

The observed trends (Figure 6-10a) show a mixed picture with few more positive than negative changes. Almost all catchments show positive trends in Africa (Congo, Nile, Niger, Orange rivers), but also many in Russia, Canada, northern Australia and some in Amazonia (see for trend details in Table 6-2; OBS).

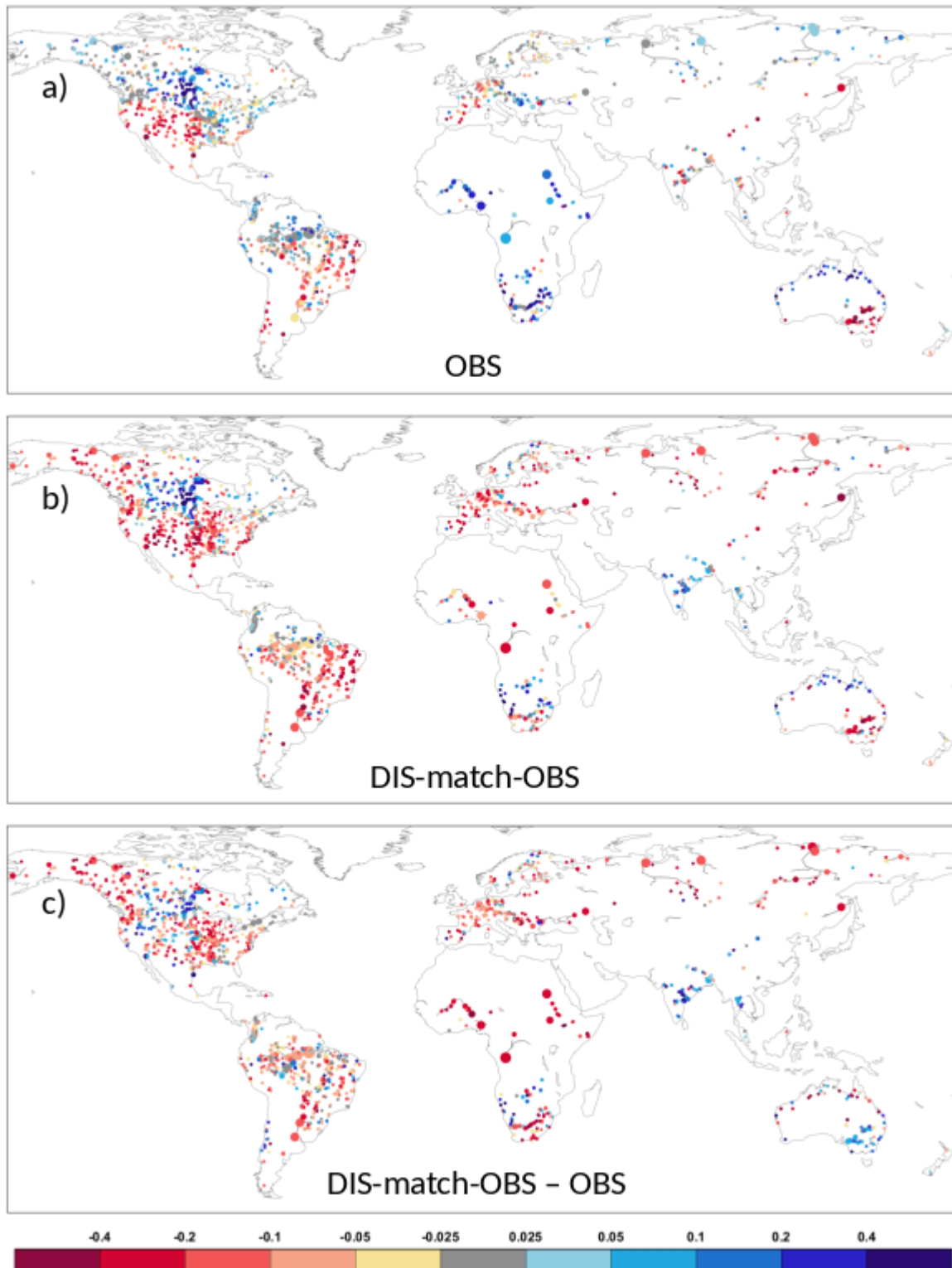


Figure 6-10. Raw linear trends (fraction/decade) at global river catchments for a) river discharge observations (OBS), b) GloFAS-ERA5 river discharge matching the available observations (DIS-match-OBS) and c) the difference (DIS-match-OBS minus OBS) catchments that have at least 16 years of available observations based on the 1981-2018 period. The circles represent the catchment outlets, while their size the catchment area.

Other scientific studies, for example Su et al. (2018) or Dai et al. (2009) found similar results after analysing hundreds of the world's largest ocean-reaching rivers with mixed positive and negative

trends. Although the main emphasis of their trends is shifted to generally more rivers showing negative than positive trends, this could likely be related to the different period (1948-2004) or the different geographical distribution of the analysed catchments.

In contrast, the GloFAS-ERA5 linear trends are dominantly negative, even though these trends are calculated over the exact same periods as for the observations. The observation-matched-period-based GloFAS-ERA5 raw linear trend (Figure 6-10b) can be quite different to the 38-year-based version in Figure 6-6, but the overall pattern is the same in both. The GloFAS-ERA5 trends are dominantly negative, while the observation trends are more positive than negative, therefore the difference between them is overwhelmingly negative (Figure 6-10c). This can be demonstrated by the rather different colours in Table 6-2 and Table B1 in Appendix B (columns of OBS and DIS-m as abbreviated from DIS-match-OBS), the simulation trend being dominantly orange (negative), while the observation trend is blue (positive). Clusters of positive differences (i.e. observations have a stronger tendency to increase) can mainly be seen in South-Asia, southern Australia and parts of central North America (Figure 6-10c).

The trends in the GloFAS-ERA5 river discharge are thus only a poor match for the trends of the available observations. Apart from the likely reason of the unrealistic trends in the ERA5 forcing, some of this can be explained by the inadequate handling of the human influence in GloFAS-ERA5, which in some areas can have very large impact on river discharge, even though this is not necessarily will impact the sign of the trends. For example, see the Nile river in Appendix C, which has observed river discharge that is only a fraction of the GloFAS-ERA5 value. A large part of this comes from the fact that the river is highly regulated with also irrigation being important in the area.

6.3.7 Trend error comparison ERA5 vs ERA5-Land

It was shown earlier that the ERA5 and ERA5-Land snowmelt trends are markedly different in the Northern Hemisphere higher latitudes, which then directly influences the runoff trends. The error of ERA5 and ERA5-Land runoff raw linear trends, computed against the observed river discharge raw linear trends, are compared in Figure 6-11 (the raw trend values are shown in Table 6-2, while the standardised trends in Table B1 for selected catchments in Appendix B, with green and purple colours). It shows the difference of the absolute trend errors, with blue (green in Table 6-2 and Table B1 in Appendix B) catchments showing where the ERA5 runoff trends are closer to the observed trends, while red (purple in Table 6-2 and Table B1 in appendix B) showing where the ERA5-Land trends are closer. It is clear that, due to the large difference in snowmelt, the ERA5-Land runoff linear trend is clearly closer to the observed trends in the higher latitudes (see the

dominantly green cells in Table 6-2 and Table B1 in appendix B, ROm/Imp column, over North Asia, North America and Europe). However, in the tropical and subtropical areas, and in the central part of the United States, ERA5 is closer to the observations or the two have similar trends (see in Table 6-2 and Table B1 in Appendix B the slightly more purple colours over South America, Africa, South Asia and Australia).

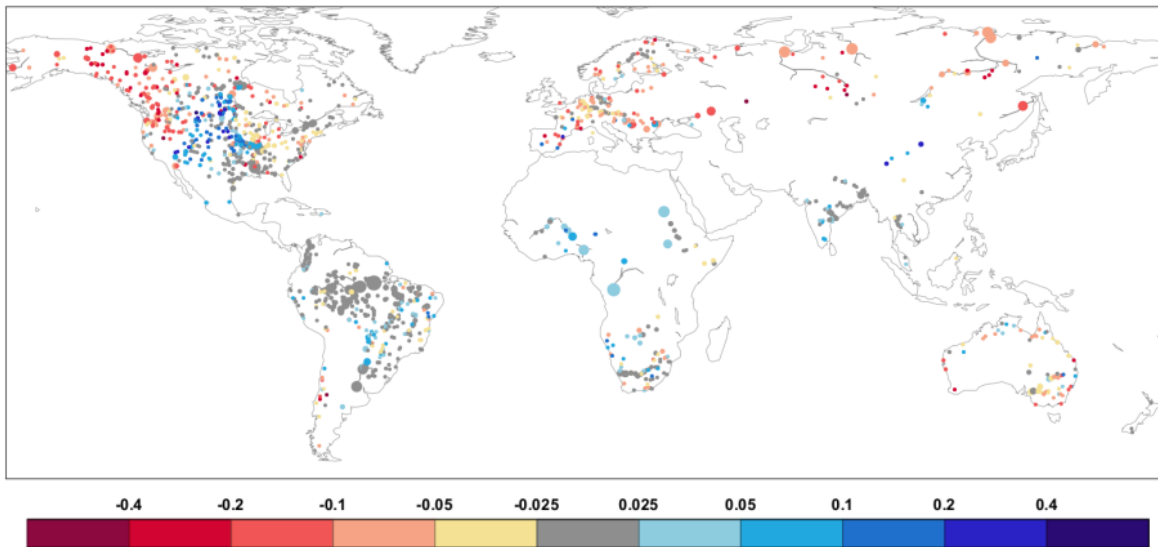


Figure 6-11. Difference between ERA5-Land and ERA5 absolute raw linear trend errors (simulated trend minus observed trend; fraction/decade) at catchments that have at least 16 years of available river discharge observations based on the 1981-2018 period. The circles represent the catchment outlets, while their size the catchment area.

6.3.8 ERA5 and ERA5-Land trends in 1981-2003 and 2004-2018

In this section the trend is compared in the two subperiods, 1981-2003 (period1, 23 years) and 2004-2018 (period2, 15 years), after removing the impact of the IMS satellite snow cover caused discontinuity in the time series.

The linear trends are generally different in the two periods, and usually larger for period2, although this shorter period would be expected to show larger random variability anyway. Figure 6-12 shows the raw linear trends for a few variables, while Table 6-3 highlights them for several variables for the selected global rivers, similarly to Table 6-2 and Table B1 in Appendix B (for location see Figure 6-4).

For precipitation, the rivers in southwestern United States, eastern South America, central Africa, the Middle East and eastern Australia stand out with their large negative trends in period1 (Figure 6-12a). In fact, these trends are very similar to the original 38-year trends in Figure 6-8a, for the majority of the world, especially for the stand-out negative areas. The second part of the 38-year period, on the other hand, does seem to show less stand-out geographical areas, the picture more

mixed, even though the trend values are quite large, likely related to the shortness of the period. For snowmelt (Figure 6-12c-d) the same is valid.

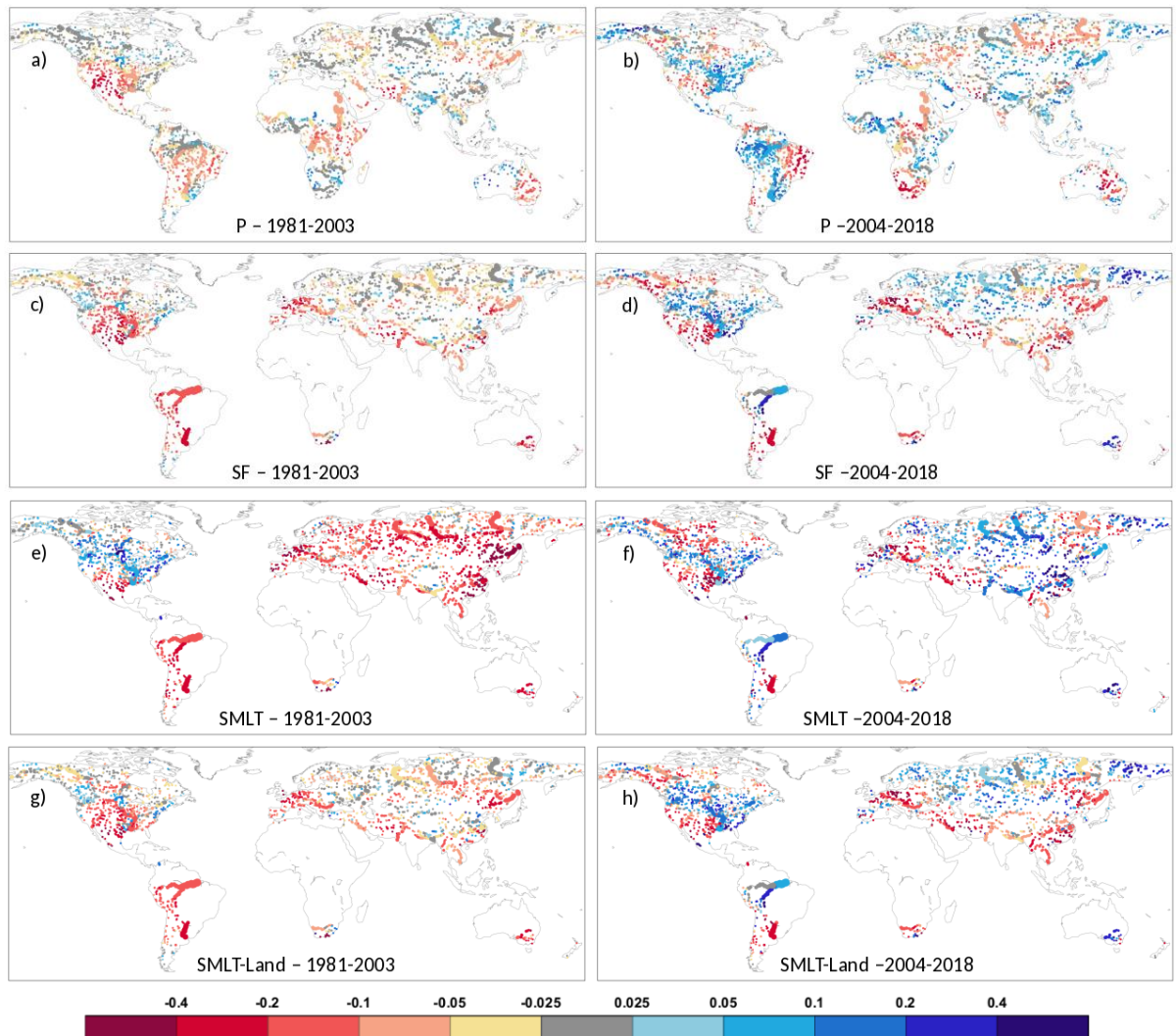


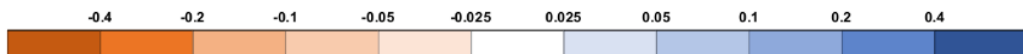
Figure 6-12. Raw linear trends (fraction/decade) at global river catchments for 1981-2003 (left column) and 2004-2018 (right column) for ERA5 precipitation (a-b), snowfall (c-d) and for ERA5 (e-f) and ERA5-Land (g-h) snowmelt. The circles represent the catchment outlets, while their size the catchment area.

Regarding snowmelt, the behaviour of ERA5 (Figure 6-12e-f) and ERA5-Land (Figure 6-12g-h) clearly differ, even after splitting the 38-year period in two. While period 2 behaves similarly for both (compare Figure 6-12f and Figure 6-12h), the magnitude of the linear trends is in the same range, period 1 shows larger differences between ERA5 and ERA5-Land (compare Figure 6-12e and Figure 6-12g), Eurasia is more negative, while North America is somewhat more positive. This is also visible amongst the selected catchments in Table 6-3 (SMLT and SMLT-Land columns), where the 1981-2003 column for ERA5 is more orange than for ERA5-Land over Europe and North Asia, and generally more blue for North America. The pronounced differences between the ERA5 and ERA5-

Land snowmelt for period1 suggests that the snow assimilation likely plays a role in producing the negative trends even before the introduction of the IMS snow product in 2004.

Table 6-3. Raw linear trends (fraction/decade) for selected catchments (see Figure 6-4) for two periods, 1981-2003 and 2004-2018, for GloFAS-ERA5 river discharge (DIS), runoff for ERA5 (RO) and ERA5-Land (RO-Land), precipitation (P), snowfall (SF) and snowmelt for ERA5 (SMLT) and ERA5-Land (SMLT-Land). Empty cells correspond to cases for which trend computation was not possible. Coloured cells indicate negative (orange) and positive (blue) trends. Where there is no raw linear trend, defined for absolute values less than 0.025, cells are not coloured. Darkening shades show increasing trend magnitudes. For location of the catchments see Figure 6-4.

Rivers		DIS		RO		RO-Land		P		SF		SMLT		SMLT-Land	
		1981-2003	2004-2018	1981-2003	2004-2018	1981-2003	2004-2018	1981-2003	2004-2018	1981-2003	2004-2018	1981-2003	2004-2018	1981-2003	2004-2018
North Asia	Ob	-0.13	0.01	-0.10	0.00	0.00	0.01	0.01	0.01	-0.03	0.04	-0.20	0.06	-0.03	0.03
	Yenisei	-0.13	-0.16	-0.12	-0.12	-0.06	-0.14	-0.02	-0.07	-0.04	0.00	-0.18	0.10	-0.05	0.02
	Lena	-0.18	-0.19	-0.13	-0.17	-0.07	-0.17	-0.01	-0.07	-0.01	-0.04	-0.14	-0.06	-0.02	-0.04
	Amur	-0.28	0.21	-0.23	0.15	-0.20	0.09	-0.06	0.06	-0.08	-0.16	-0.47	0.09	-0.14	-0.15
North America	Yukon	-0.03	0.07	-0.03	0.10	-0.04	0.12	-0.02	0.08	-0.03	-0.06	0.02	0.01	-0.03	-0.07
	Mackenzie	0.00	-0.27	0.00	-0.19	0.00	-0.08	0.00	-0.01	-0.04	-0.08	0.01	-0.15	-0.04	-0.05
	Churchill	0.65	-1.14	0.29	-0.53	0.04	-0.10	0.01	-0.02	-0.14	0.03	0.03	-0.17	-0.17	0.06
	Columbia	-0.02	0.14	-0.02	0.12	-0.07	0.17	-0.07	0.04	0.01	0.06	0.10	0.09	0.01	0.09
	Colorado	-0.62	-0.64	-0.56	-0.52	-0.68	-0.47	-0.27	-0.10	-0.22	-0.29	-0.20	-0.28	-0.22	-0.28
	Mississippi	0.01	0.13	-0.01	0.15	-0.12	0.22	-0.06	0.09	-0.12	0.07	0.08	0.03	-0.12	0.09
Europe	Kalixaelven	-0.27	0.61	-0.20	0.45	-0.03	0.12	-0.01	0.03	0.01	0.09	-0.13	0.55	0.01	0.19
	Thames	-0.07	0.39	-0.07	0.36	0.03	0.37	0.02	0.02	-0.67	0.21	-0.94	-0.09	-0.65	0.17
	Rhine	-0.11	0.16	-0.09	0.13	-0.07	0.04	-0.02	-0.03	-0.19	-0.18	-0.37	0.02	-0.18	-0.13
	Danube	-0.07	-0.04	-0.07	-0.08	-0.05	-0.11	0.01	-0.03	-0.09	-0.12	-0.15	-0.12	-0.10	-0.13
	Volga	-0.32	-0.11	-0.27	-0.10	-0.11	-0.11	-0.03	-0.06	0.01	0.05	-0.14	0.04	0.00	0.03
	Ebro	-0.21	0.52	-0.16	0.42	-0.14	0.33	-0.01	0.17	-0.06	0.12	-0.19	0.60	-0.08	0.14
South America	Magdalena	-0.03	-0.17	-0.04	-0.17	-0.05	-0.17	-0.03	-0.11						
	Amazon	-0.03	0.10	-0.03	0.10	-0.03	0.11	-0.02	0.06	-0.15	0.07	-0.17	0.10	-0.16	0.07
	Araguaia	-0.16	-0.33	-0.15	-0.29	-0.17	-0.28	-0.08	-0.09						
	Parana	-0.12	0.11	-0.11	0.07	-0.12	0.09	-0.05	0.05	-0.21	-0.26	-0.23	-0.26	-0.21	-0.26
Africa	Nile	-0.12	-0.01	-0.12	-0.04	-0.14	-0.05	-0.07	-0.07						
	Niger	-0.01	0.09	0.00	0.08	-0.02	0.08	0.00	0.08						
	Shabelle	-0.13	-0.04	-0.13	-0.04	-0.13	-0.05	-0.09	0.03						
	White-Volta	0.26	0.54	0.22	0.46	0.17	0.51	0.00	0.17						
	Congo	-0.24	-0.16	-0.20	-0.16	-0.23	-0.17	-0.08	-0.05						
	Zambesi	-0.05	0.30	-0.03	0.20	0.01	0.21	-0.02	0.01						
	Cunene	0.22	-0.31	0.20	-0.29	0.26	-0.16	0.10	-0.05						
	Orange	0.17	-0.53	0.08	-0.42	0.08	-0.43	0.00	-0.22	-0.07	-0.15	-0.09	-0.09	-0.07	-0.14
South Asia	Brahmaputra	0.01	0.03	0.01	0.03	0.01	0.01	0.00	0.01	-0.02	-0.06	-0.04	0.06	-0.02	-0.09
	Ganges	0.12	0.12	0.10	0.11	0.12	0.10	0.05	0.04	-0.03	-0.04	-0.04	0.18	-0.01	-0.03
	Godavari	0.07	0.10	0.07	0.09	0.10	0.11	0.06	0.01						
	Karnali	0.07	0.11	0.07	0.11	0.06	0.10	0.04	0.05	-0.04	-0.09	-0.04	0.03	-0.03	-0.06
Australia	Murray	-0.22	0.38	-0.17	0.19	-0.23	0.12	-0.10	-0.05	-0.19	0.21	-0.27	0.23	-0.19	0.22



6.4 Conclusions

This study has analysed the GloFAS-ERA5 river discharge reanalysis data set, the related ERA5 and ERA5-Land surface variables, and the available river discharge observations, for noticeable changes in the time series characterised by linear trends, in the 38-year period of 1981-2018, also including the 1981-2003 and 2004-2018 subperiods. It was found that the river discharge simulation shows a dominantly negative change across the world during 1981-2018, with some major world rivers having quite substantial decrease (Yenisei, Volga, Congo, Amur, Colorado, Yukon, Nile, Lena, Yellow, see Table 6-2 and Table B1 in Appendix B).

* For a geographical summary of the trends, Figure 6-A is added in Appendix A8 showing trends for the all the analysed surface variables in both ERA5 and ERA5-Land for all the 33 selected catchments. Please be aware that this figure does not appear in the published version of this chapter.

The river discharge observations do not support such dominantly negative linear trends, and although varied, observations show overall more positive than negative changes in the 38-year period. The scientific literature generally documents a similar behaviour to the observational trend analysis presented here, with mixed trends globally, but with slightly more major rivers showing negative than positive changes in observed river discharge during the period of 1948-2004 (Su et al., 2018; Dai et al., 2009).

The linear trends in GloFAS-ERA5 seem to be driven by changes in precipitation over the tropical and subtropical areas of the world, with the snowmelt changes showing a very strong influence in determining the river discharge trends in the northern latitudes of the Northern Hemisphere. The reason for this atypical behaviour in the northern latitudes is likely to be related to changes in the snowmelt producing processes, including the snow assimilation.

The snowmelt exhibits a pronounced negative linear trend in large parts of the world in ERA5, while this is not present in ERA5-Land. This suggests that the negative trends are, at least partially, related to the snow assimilation tendency to remove water from the water cycle as a consequence of the suboptimal snow scheme in HTESSEL (Zsoter et al., 2019). Some of the issues stem from the use of the IMS snow product from 2004 in the snow assimilation, which creates a discontinuity in the ERA5 time series (see Figure 6-5).

It has to be acknowledged that such discontinuity can make any trend analysis unreliable. However, as Figure 6-5 suggests, the snow evolution in ERA5 and ERA5-Land seems to be more complex than a single discontinuity in 2004, so linear trends can still deliver valuable information even on the whole 1981-2018 period.

After splitting the period into 1981-2003 and 2004-2018 and removing the impact of this discontinuity on the linear trends, it could still be shown, that even in the first subperiod of 1981-2003 there is a large area globally with significant negative snowmelt trends in ERA5, mainly in Asia and Europe, which is not present in ERA5-Land. This highlights that there should be other contributing factors in generating such negative trends in the ERA5 snowmelt, other than just the introduction of the IMS satellite product. Potentially the generally lower temperatures in mountains due to the higher orography in ERA5-Land could also contribute to this by decreasing the snowmelt amount compared with ERA5.

Two particularly interesting areas with the largest linear trends, that were highlighted by this study, are the central region of Africa (e.g. Congo and Nile river basins), and the southwestern part of the United States (e.g. Colorado river basin). Both these areas show very dominant and large negative trends both in precipitation and river discharge, but also in runoff by not just ERA5 but also equally by ERA5-Land. However, based on the limited analysis of the scientific literature, there was no indication of such strong precipitation trends in the explored studies, and similarly no such large trends were shown either in the river discharge observations available in this study.

It will require more work in the future to better identify the underlying reasons for these very dominant negative trends. Moreover, it would also be beneficial to repeat the analysis including an improved precipitation observation data set, preferably one that is high quality and merges several of the available gauge- and/or satellite-based data sets, such as the MSWEP (Beck et al., 2019), or the latest bias corrected ERA5 data set, WFDE5 (Cucchi et al., 2020). This would allow us to directly evaluate the quality of the ERA5 trends against the best available observation estimates.

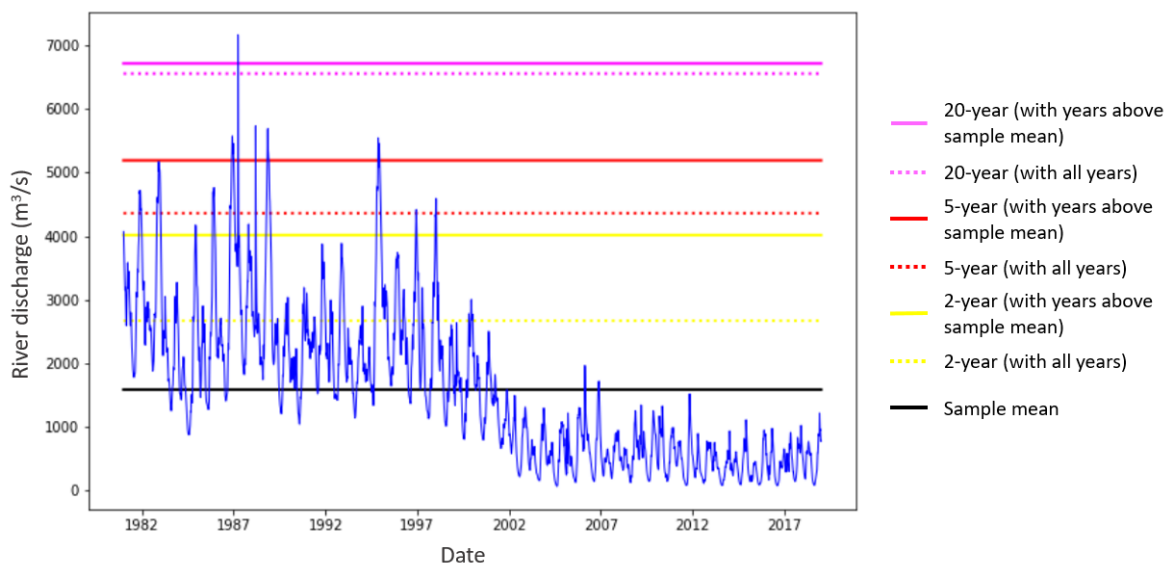


Figure 6-13. Example of the daily GloFAS-ERA5 river discharge (m^3/s) time series for an upstream catchment in the Congo river basin and the 2- (yellow lines), 5- (red lines) and 20-year (magenta lines) flood thresholds based on two different fitting methods. The dashed lines represent the flood thresholds fitted using all annual maxima, while the solid lines were produced using only annual maxima over the sample mean discharge shown by the black line.

A potential driver of the exposed ERA5 trends is the method of production: ERA5 is produced in several streams that were later merged into one consolidated data set. These streams have a year overlap to allow for long enough spin-up. According to Hersbach et al. (2020), in the deep soil, where spin-up can take several years, discontinuities could be observed. The deep soil can certainly impact on the runoff through the sub-surface runoff, which can impact on river discharge, however,

it is not expected to have any noticeable impact on variables such as precipitation or evaporation which vary on a significantly shorter timescale.

The GloFAS-ERA5 linear trends, presented here, have a direct impact on the quality of the GloFAS flood warnings through the use of the flood thresholds. The presence of significant trends, or a very substantial regime change, such as in the example provided in Figure 6-13, makes it difficult to produce flood thresholds that correctly represent the extreme event behaviour in the forecasts. In the provided example, the river discharge level collapses to less than a third after 2000. Thus, the thresholds will be much too high and represent only the first half of the period. In this case, the real time GloFAS forecast will likely be similar to the latter part of the reanalysis period and will hardly ever exceed these flood thresholds, making the flood warnings very unreliable.

In other catchments, where the river discharge is significantly increasing, the situation is the opposite. In such catchments the real time forecasts will be mostly similar to the latter part of the reanalysis and the flood thresholds will be biased towards the lower earlier years. In this case, the thresholds will be little too low and the real time GloFAS forecasts are expected to show too frequent flood events, making the forecasts unreliable again.

The example in Figure 6-13 is a very extreme one from the upstream part of the Congo basin. However, it is not an isolated case causing problems, as areas where the linear trend shows at least $\sim 20\%$ change in the ~ 40 -year period, extreme value fitting to compute the flood thresholds is likely to provide us with difficulties. There is no scientific basis for this 20% minimum value (equivalent to ~ 0.05 raw linear trend magnitude), further tests would be needed to identify the expected forecast reliability loss due to such trends in the reanalysis time series. Based on the current reanalysis period, as used in GloFAS, at least 50% of land areas show raw linear trends higher (lower) than 0.05 (-0.05) (see Figure 6-6 and also Table 6-2) causing a significant issue in the operational running of GloFAS.

The extension of ERA5 back to 1950 (Hersbach et al., 2020) will provide an increase in the period used in the flood thresholds computation. However, it is strongly recommended that hydrological trends, with any potential discontinuities or regime changes due to merging different streams in the ERA5 production, should be analysed in that extended data set before making any change in GloFAS.

In addition, it will be important to explore ways to better derive the flood thresholds for catchments that are impacted by large trends in the GloFAS-ERA5 river discharge. An option could be to limit the period for the flood threshold computation to most recent decades, on a case by case basis (for single catchments, or maybe whole regions) selecting the period length that provides historical time

series with low enough trend magnitudes. Based on this study, and the current ERA5 reanalysis, this could be as short as the last 15-17 years from ~2004, which although not ideal for estimating longer return periods, may provide improvements to make the GloFAS flood warnings more reliable in the future.

Zsoter et al. (2020a) recommends that river discharge ensemble reforecasts should be used to compute flood thresholds instead of reanalysis, which would help to create more reliable flood warnings especially at longer lead times. The use of the reforecasts would be beneficial, as although they are initialised from GloFAS-ERA5 and are thus bound to inherit any trend problems that are present in this reanalysis, they are only generated on the most recent 20-year period (currently 1999-2018), which would undoubtedly lessen the trend impact documented in this study.

Finally, the documented river discharge linear trends in the GloFAS-ERA5 dataset are large enough to warrant further investigation of the underlying causes to the general behaviour of the water cycle variables in ERA5. This is crucial in order to provide improvements in hydrological variables such as river discharge, especially in the context of any future version of ECWMF reanalysis data sets, such as ERA6.

Contribution of this chapter to the thesis. This chapter addressed the objective: "Evaluate the relevant trends in hydrological reanalyses for river discharge and other related land-surface variables and analyse how the interactions amongst these variables contribute to the trends." This work demonstrated that trends are widespread in a state-of-the art reanalysis time series for river discharge and the related land-surface variables. These trends have the potential to negatively impact the usability of the flood thresholds generated from these time series. The next chapter will take this further and explore the impact that the threshold generation method can have on flood forecast reliability and skill, by focussing on the used data source and sampling strategy.

Chapter 7 Using ensemble reforecasts to generate flood thresholds for improved global flood forecasting

Chapter 6 showed that non-stationarity, measured by linear trends, is widespread in the analysed GloFAS-ERA5 hydrological reanalysis. Another aspect of the global flood forecasting systems that has the potential to hinder the usability of the reanalysis-derived flood thresholds is non-stationary forecast biases. This chapter will analyse this, by comparing the use of the traditional reanalysis-based thresholds to a newly designed, innovative, ensemble-reforecast-based methodology, that has the potential to increase the reliability and skill of the flood forecasts. This chapter has been published in *Journal of Flood Risk Management (JoFRM)* with the following reference (<https://centaur.reading.ac.uk/91631/>):

Zsoter, E., Prudhomme, C., Stephens, E., Pappenberger, F. and Cloke, H, 2020a: Using ensemble reforecasts to generate flood thresholds for improved global flood forecasting, *J. Flood Risk Management*, doi:10.1111/jfr3.12658

This paper was chosen as one of the three highly commended articles of 2021 as part of the Journal of Flood Risk Management's outstanding paper award.

The contributions of the authors of this paper are as follows: E.Z. designed the experiment, carried out the flood threshold analysis, and led the writing of the manuscript. H.C. and E.S. assisted with posing the research question and designing the analysis. C.P. and F.P. helped designing the research methodology. All authors assisted with writing the manuscript. Overall, 90% of the writing was undertaken by E.Zs. The published article can be found in the Appendix A4.

Abstract. Global flood forecasting systems rely on predefining flood thresholds to highlight potential upcoming flood events. Existing methods for flood threshold definition are often based on reanalysis datasets using a single threshold across all forecast lead times, such as in the Global Flood Awareness System. This leads to inconsistencies between how the extreme flood events are represented in the flood thresholds and the ensemble forecasts. This paper explores the potential benefits of using river flow ensemble reforecasts to generate flood thresholds that can deliver improved reliability and skill, increasing the confidence in the forecasts for humanitarian and civil protection partners. The choice of dataset and methods used to sample annual maxima in the threshold computation, both for reanalysis and reforecast, is analysed in terms of threshold magnitude, forecast reliability, and skill for different flood severity levels and lead times. The variability of threshold magnitudes, when estimated from the different annual maxima samples, can be extremely large, as can the subsequent impact on forecast skill. Reanalysis-based thresholds should only be used for the first few days, after which ensemble-reforecast-based thresholds, that

vary with forecast lead time and can account for the forecast bias trends, provide more reliable and skilful flood forecasts.

7.1 Introduction

Flood forecasting systems use meteorological data and hydrological modelling to deliver forecasts of river discharge and other hydrological variables such as inundation or soil moisture. They provide early flood warnings on time scales up to several weeks ahead, essential for managing flood risk at local, regional, and recently also on the global scale (Emerton et al., 2016).

The state-of-the-art systems in use today provide an ensemble of equally likely solutions that can be used to define occurrence probabilities for certain flood events (Cloke and Pappenberger, 2009; Wu et al., 2020). These flood events are defined by comparing the forecast time series with flood thresholds, usually based on a return period magnitude or a quantile.

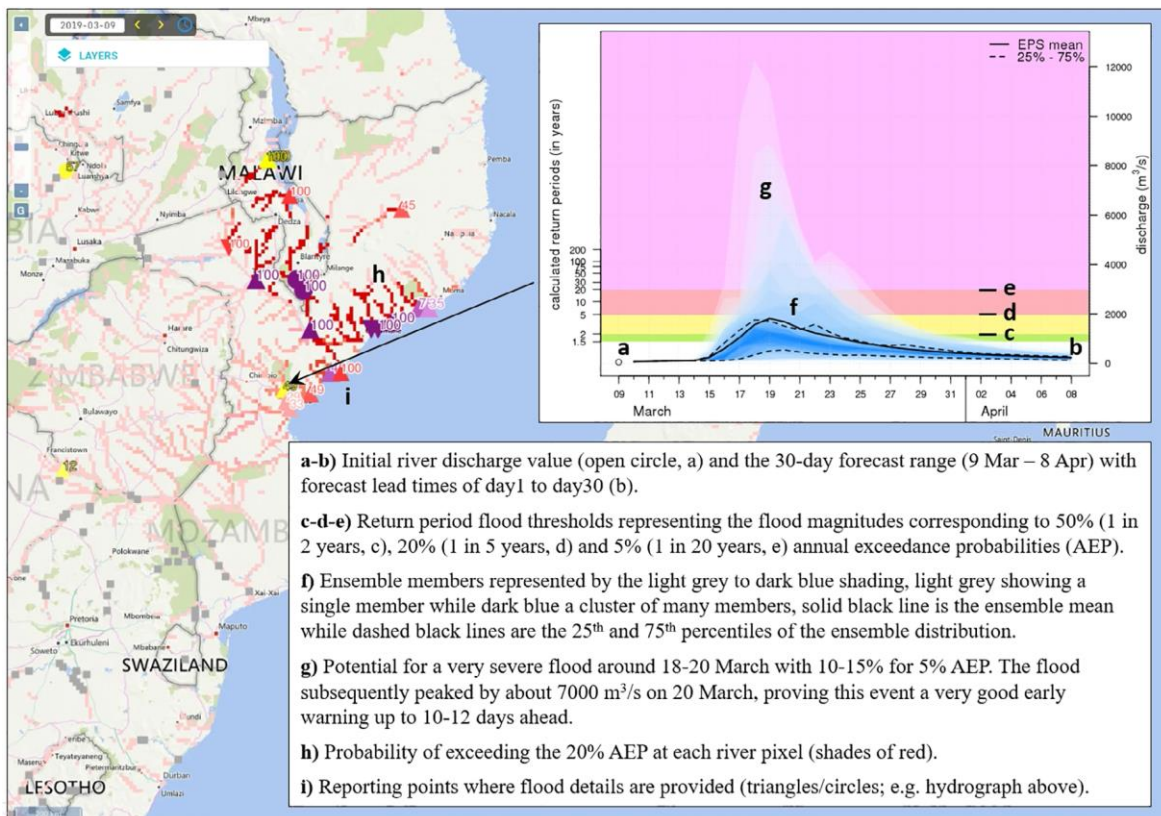


Figure 7-1. GloFAS forecast on 9 March 2019 for Mozambique showing flood predictions related to tropical cyclone Idai. As an example, the inset diagram shows the hydrograph for a river point near the coast in Mozambique for the 30-day period of 9 March to 8 April. GloFAS forecasts are openly accessible on www.globalfloods.eu.

In the Global Flood Awareness System of the Copernicus Emergency Management Service (GloFAS; Alfieri et al., 2013, Hirpa et al., 2018b), the severity of the predicted flood is defined according to a set of three thresholds, as shown in Figure 7-1 for the example of tropical cyclone Idai in Mozambique in March 2019. These thresholds are computed from a 40-year long river discharge

Chapter 7 Using ensemble reforecasts to generate flood thresholds for improved global flood forecasting

reanalysis (Harrigan et al., 2020b). The hydrograph in Figure 7-1 shows the predicted river discharge for the next 30 days, highlighting a severe flood event around 18–21 March with 10–15% chance of exceeding the 5% annual exceedance probability (AEP) threshold.

The flood thresholds, defined according to flood magnitude of selected return periods (or flood quantiles), and used in many of the existing flood prediction systems (GloFAS Alfieri et al., 2013; EFAS Thielen et al., 2009; WW-HYPE Arheimer et al., 2020), are determined by flood frequency analysis, usually by fitting an extreme value distribution on a set of annual maxima, sampled from a time series as long as possible. These quantities describe the likelihood of different flood magnitudes occurring locally based on a “climatological” data set over a long period of time (preferably 30 years or more; World Meteorological Organisation [WMO], 2017). Traditionally, flood thresholds are produced from observations or deterministic model reanalysis (Alfieri et al., 2015). River discharge observations can provide a solution only at certain locations, whereas hydrological model simulations, forced with meteorological observations, can cover a whole geographical domain, delivering flood thresholds at every model river point or catchment.

Because the flood thresholds determine the severity of the forecasted flood signal, these flood thresholds should ideally represent extreme events the same way as they occur in the forecasts. If this is not the case and the different biases make an event of the same magnitude occur with a different frequency in the climatological data set that was used to compute the thresholds and in the forecasts (e.g., the 5% AEP flood magnitude happens more often in the forecasts than the expected 5% probability in a given year), then the flood forecast probabilities could become unreliable (e.g., leading to flood signals that often overestimate the flood severity). In the case of the example in Figure 7-1, this could mean that the predicted severe flood event should in fact appear significantly less extreme as the high severity would only be a consequence of the unrealistically low thresholds.

The extreme event representation of flood thresholds can be heavily influenced by the data set and method used to derive them. The value of the flood quantiles can be impacted by the choice of the extreme value statistical distribution (Papalexiou and Koutsoyiannis, 2013), the data set that is used for the annual maxima extraction (observation, reanalysis or forecasts; see, for example, Hirpa et al., 2016) and its length (Kjeldsen et al., 2014). These can all lead to potential differences in the flood threshold magnitudes, subsequently resulting in differences in the forecast probabilities to exceed the thresholds, and ultimately causing an impact on the quality of the flood warnings.

By definition, conventional observation- or reanalysis-based data sets provide a single time series to compute flood thresholds, meaning that only one set of thresholds, with different severities, is

going to be applied to all lead times in the forecast range. This might cause further inconsistencies if the forecast biases have trends across lead times. For example, forecasts might show increasing river discharge overprediction with lead time, which would result in a growing number of forecasts exceeding the 5% AEP flood threshold (which stays unchanged as it is computed from the reanalysis time series), with an increasingly higher frequency than the expected 5% of the years occurrence on average (Alfieri et al., 2019). As trends and biases in a forecast are model specific, using different meteorological forcing models within the same forecasting system (such as in the European Flood Awareness System, EFAS, Thielen et al., 2009) might cause even more complex inconsistencies between the observation- or reanalysis-based flood thresholds and the forecasts.

Bias correction methods can help to achieve consistency between forecasts and thresholds (e.g., Verkade et al., 2013; Yuan and Wood, 2012). They have the potential to make the extreme event representation of the forecasts and the climatology, that is used to define the thresholds, similar. However, bias correction, even in its simplest form with only hydrological output postprocessing (without correction of the meteorological forcing data), would introduce further complexity into the river discharge production chain with its associated uncertainties.

Alternative approaches have been investigated (e.g., Alfieri et al., 2019). Generally, flood thresholds are not produced from forecasts. Part of the reason could be the limited sample of available historical forecasts and the convenience for the users to work with only a single threshold set that does not show evolution with lead time. The consequence is that, as said earlier, the same threshold is applied to all forecast lead times. However, Alfieri et al. (2019) showed that range-dependent, reforecast-based thresholds were substantially different from unique reanalysis-based thresholds in two thirds of the global rivers. Moreover, despite the recent advancement of ensemble-based forecast systems, ensemble forecasts are generally not considered in the flood threshold generation. However, this can be a problem as ensembles can have different biases to single deterministic forecasts (Leutbecher et al., 2017), which can further contribute to the extreme event representation inconsistencies between reanalysis-based thresholds and ensemble forecasts.

The use of ensemble reforecasts in generating the climatological sample can provide a range-dependent threshold set (e.g., as in Emerton et al., 2018 and Tsonevsky et al., 2018), which has the potential to overcome the issues associated with extreme event identification. In addition, multi-value ensembles can also contribute to increased effective sample size, from which to define flood thresholds, and therefore help to improve the representation of extreme events (Zsoter et al., 2014). This could be important for very extreme events which might not occur in the typical 30–50-year-long sample of traditional observation or reanalysis time series (e.g., the median length of the daily data in the GRDC archive is 39 years, as of January 22, 2020 at www.bafg.de/GRDC).

In this study, the potential benefits of using river discharge ensemble reforecasts to define flood thresholds are analysed globally. Two main research questions were explored in our study, targeting specifically the sampling strategy to extract the annual maxima sample on which the flood frequency analysis is conducted:

- How adequate it is to use a reanalysis dataset to define flood thresholds and apply them for all forecast lead times?
- How best to use reforecast ensemble information in the flood threshold generation to improve flood forecast performance?

The work is carried out in the context of GloFAS, for a 30-day forecast range, with a selection of over 5,000 catchments. The impacts of the choice of data source (reanalysis or reforecasts) and of the annual maximum sampling strategies (from the reforecasts) are analysed by comparing the flood threshold magnitudes and the resulting forecast reliability and skill benefits for four different flood severity levels.

7.2 System description, datasets and methods

This section describes the data sets, methods, and experimental set-up used to generate flood thresholds and compare their value and impact on the flood forecast skill.

7.2.1 *GloFAS*

GloFAS is part of the Copernicus Emergency Management Service (CEMS) and has been developed by the Joint Research Centre of the European Commission and the European Centre for Medium-Range Weather Forecasts (ECMWF) with help from research institutions such as the University of Reading (UoR; e.g., Stephens et al., 2015; Emerton et al., 2017 and Towner et al., 2019). It is a probabilistic hydrological prediction system, which has a 30-day (Alfieri et al., 2013) and a seasonal component (Emerton et al., 2018). This study is based on the 30-day component, which predicts daily flood occurrences on the global scale.

In GloFAS, ensemble runoff outputs from the HTESSSEL land surface model (the Hydrology-Tiled ECMWF Scheme for Surface Exchange over Land; Balsamo et al., 2009; Balsamo et al., 2011) are coupled to the Lisflood hydrological model (van der Knijff et al., 2010) to produce an ensemble of daily river discharge across a global river network at 0.1° resolution (Alfieri et al., 2013; Hirpa et al., 2018b). To detect the likelihood of high flow situations, to forecast flood events, the real time river discharge forecasts are compared with a set of flood thresholds derived from a 40-year long climatological simulation, a daily river discharge reanalysis time series.

7.2.2 *River discharge reanalysis*

The GloFAS-ERA5 river discharge reanalysis (Harrigan et al., 2020b) is produced with ERA5 forcing, ECMWF's fifth generation global climate reanalysis (Hersbach et al., 2018; Hersbach et al., 2020), which is part of the EU-funded Copernicus Climate Change Service (C3S). ERA5 covers the period 1979 to present and is updated with two to 3 months delay. ERA5 is open access (<https://climate.copernicus.eu/>) and includes one high resolution component and a lower resolution ensemble component with 10 members. The GloFAS-ERA5 uses the high-resolution ERA5 component at ~31 km horizontal resolution with the configuration of the GloFAS operational forecasting systems. GloFAS-ERA5 is a key component of GloFAS verification, serving as a proxy for river discharge observations and it is also openly available from the Copernicus Climate Change Service Climate Data Store (Harrigan et al., 2020b).

7.2.3 Ensemble river discharge reforecasts

The ensemble river discharge reforecasts are GloFAS reforecasts produced for the 20-year period of 1997–2016. These are 30-day river discharge forecasts generated for past dates by the same GloFAS system that is used for the real time forecasts. They are initialised from GloFAS-ERA5 and forced by runoff from the twice weekly (Monday and Thursdays in 2017), 11-member, 20-year ECMWF meteorological ensemble reforecasts (Vitart, 2014). This data set includes a batch of 20 reforecasts (one for each year in 1997–2016) for each Monday and Thursday in 2017. Altogether 2080, 11-member, 30-day reforecasts were produced for the 20-year period (104 in each year).

7.2.4 Flood thresholds

In the 30-day GloFAS, flood quantiles of three severity levels (2-, 5-, and 20-year return periods) are used as flood thresholds. Flood quantiles are commonly used in risk analysis, typically estimated using time series data of generally twice the length of the return period of interest. Because of the relatively short length of daily discharge data available, the 10-year return period severity was also considered in this study.

A return period T is an estimate of the likelihood of an event to occur (Gumbel, 1941), expressed as average number of years for an event of same or higher magnitude to occur. It can also be expressed as an Annual Exceedance Probability AEP (given by $AEP = 100/T$). To facilitate the interpretation, AEP is used in the rest of this study.

The flood thresholds were computed as currently done operationally in GloFAS: the Gumbel Extreme Value Distribution (EVD) is fitted to the annual maximum river discharge sample using the method of L-Moments (Hosking, 1990). This method is appropriate for relatively small sample sizes, such as used in GloFAS (Alfieri et al., 2019) and in this study (20 years of reforecast data from 1997 to 2016).

7.2.5 River catchments

The study is based on the GloFAS network (Figure 7-2), a set of 6,122 catchments of which about one-third are always highlighted on GloFAS website as reporting points (www.globalfloods.eu). This network provides a global coverage and includes all points where daily historical river discharge observations are made available to the GloFAS team. Catchments that have 50% AEP magnitude below 20 m³/s in GloFAS-ERA5, that is, too dry or too small, were excluded from the study, resulting in 5,665 catchments in total for the analysis.

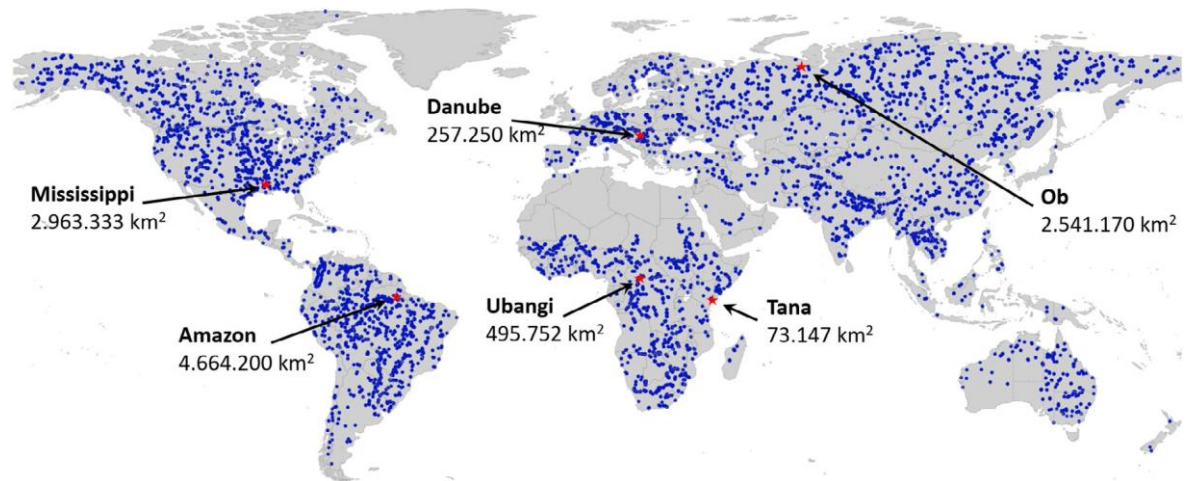


Figure 7-2. The 5,665 GloFAS stations used in this study. The six contrasting catchments of Figure 5 are indicated by red stars, along with the river names and GloFAS upstream areas.

7.2.6 Analysis methods

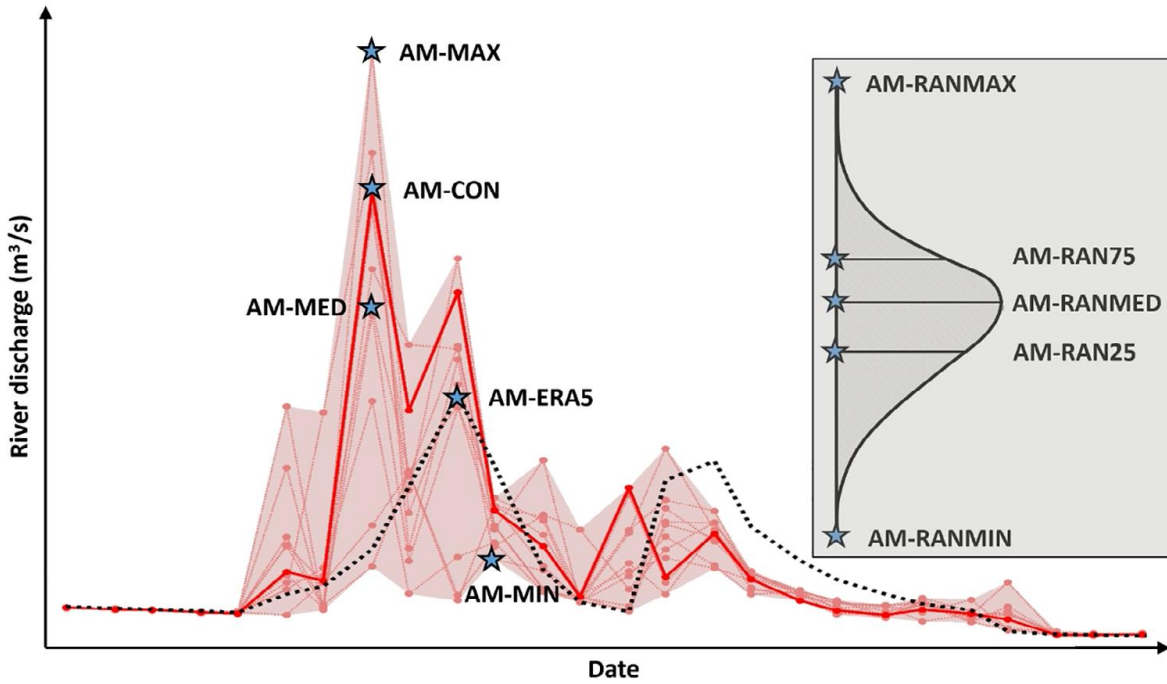
The impact of flood threshold estimation for all four severity levels was analysed for each catchment by direct comparison of the quantile magnitudes. Additionally, the flood forecast performance was evaluated for day 1 to day 30 lead times by:

- Comparing the number of events forecasted (i.e., when the discharge exceeds the flood threshold) with the number of events identified in the benchmark set (i.e., GloFAS-ERA5 river discharge reanalysis which is the nearest equivalent to the “observations”), expressed as percentage occurrence frequency (or event/forecast probability). This step analyses the simplified forecast reliability with only one probability category;
- Calculating the Brier score (Murphy, 1973) and the reliability diagram (Hsu and Murphy, 1986). This step assesses both the skill and reliability in the resulting probability forecasts of exceeding the thresholds.

7.2.7 Experimental set-up

For consistency and comparability, the annual maxima sampling for the flood threshold computation was done from daily time series containing only the calendar days corresponding to

the dates of the day 1 to day 30 reforecast values (for all Monday and Thursday reforecast runs of 1997–2016).



Version	Time series set	Version	Time series set
T-ERA5	Benchmark set	T-RANMIN	Minimum extended reforecast set
T-CON	Control member reforecast set	T-RAN25	25th percentile extended reforecast set
T-MIN	Minimum member reforecast set	T-RANMED	Median extended reforecast set
T-MED	Median member reforecast set	T-RAN75	75th percentile extended reforecast set
T-MAX	Maximum member reforecast set	T-RANMAX	Maximum extended reforecast set

Figure 7-3. Schematic of the annual maximum sampling for flood threshold estimates from daily river discharge time series. Dotted black line: GloFAS-ERA5, solid red line: GloFAS reforecast control member, light red lines: GloFAS reforecast perturbed ensemble members. Small red dots show individual daily river discharge values in the ensemble reforecasts. The x-axis shows the date of the forecasts, while the y-axis the river discharge values.

For each lead time, three sets of time series were used:

- Benchmark set: GloFAS-ERA5 river discharge reanalysis (independent from the lead time). This is as close as possible to the flood thresholds used operationally in GloFAS and can be considered as proxy observation-based thresholds;
- Reforecast set: the time series of the control member, plus three time series corresponding to the minimum, median and maximum values from each run of the 11-member GloFAS reforecasts;
- Extended reforecast set: 1,000 time series, each generated with randomly selecting one of the 11 ensemble members from each GloFAS reforecast.

After applying the flood threshold generation method, described earlier, this resulted in 1 + 4 + 1,000 threshold values summarised graphically in Figure 3 for the annual maxima selection differences. From the 1,000 random-member-based thresholds only the minimum, 25th percentile, median, 75th percentile, and maximum values were analysed further. The exercise was conducted on all study catchments, flood severity levels, and forecast lead times. The major methodological steps of this study are provided in Table 1.

Table 7-1. Major methodological steps of this study.

Steps	Description
Setup	Ensemble reforecasts for day 1 to day 30 lead times, over 20 years (1997-2016), with 104 forecasts in each year, flood thresholds computed by fitting an extreme value distribution on the 20 annual maxima, for 5665 global catchments and 4 return periods (50, 20, 10 and 5% AEP)
Benchmark (reanalysis) thresholds	Produce reanalysis-based reference thresholds (T-ERA5) for all lead times, always with the days of the reforecasts, to guarantee homogeneous samples
Reforecast thresholds	Produce ensemble-forecast-based alternative thresholds (T-CON, T-MIN, T-MED, T-MAX)
Extended reforecast thresholds	Produce random-ensemble-member-based thresholds 1000 times (T-RAN)
Extended reforecast threshold distribution	Define the key statistics of the extended reforecast threshold distribution (T-RANMIN, T-RAN25, T-RANMED, T-RAN75 and T-RANMAX)
Probabilities	Compute the exceedance probabilities over the 20-year period with all threshold versions (10 in total)
Scores	Compute the Brier scores and produce the reliability diagrams with all threshold versions (10 in total), for all catchments including a global average

Note: For all lead times, catchments and return periods if not otherwise stated.

7.3 Results

The impact of the data set and sampling strategy choice in the flood threshold generation was analysed globally on selected river catchments, with the flood threshold magnitude and forecast skill compared geographically.

7.3.1 How similar are the flood thresholds?

In this section, we analyse the impact of the annual maxima sampling strategy on the flood threshold magnitude for day 1 and day 30 lead times, focusing on the 10% AEP severity level.

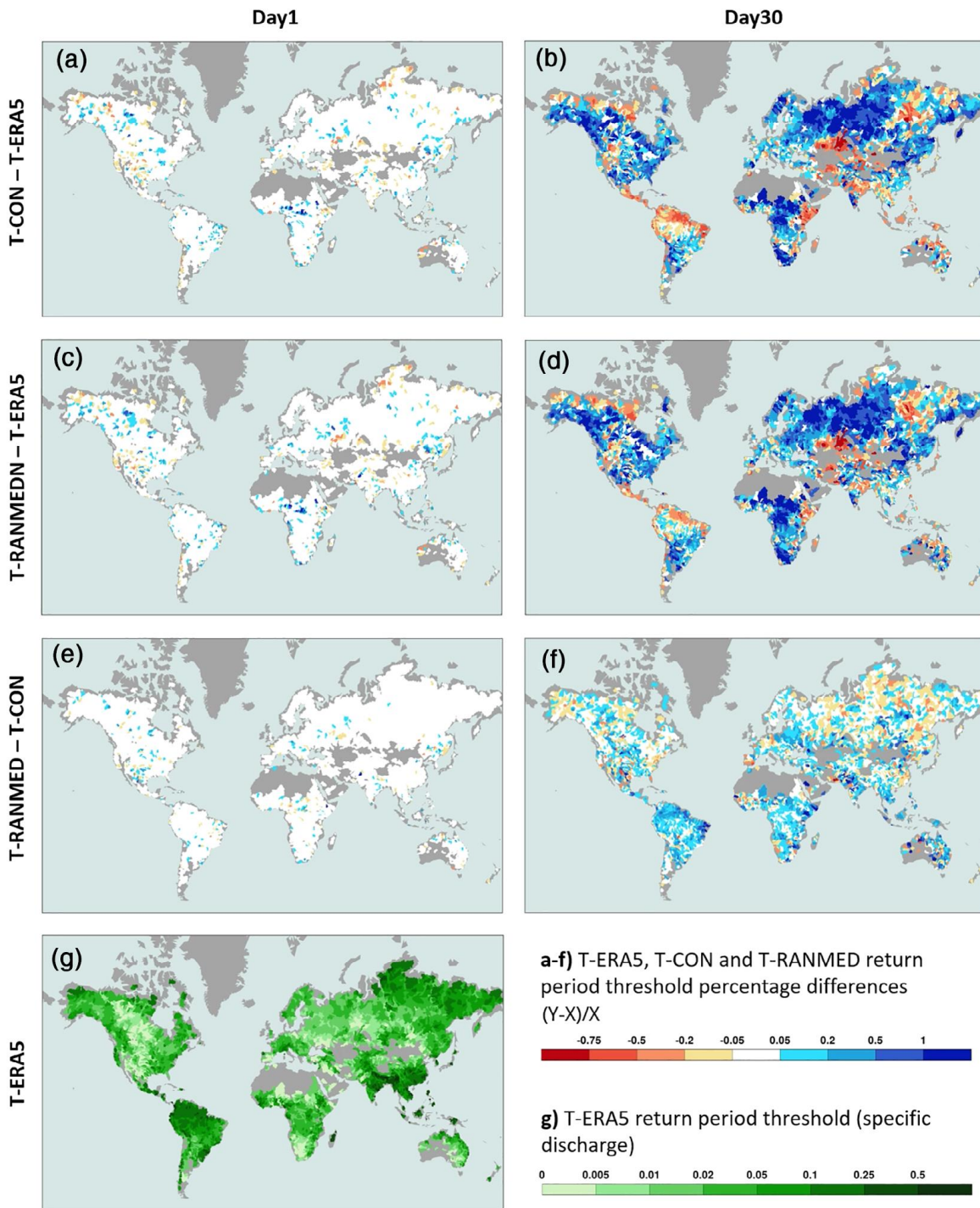


Figure 7-4. Percentage difference of 10% AEP flood thresholds between (a, b) T-CON and T-ERA5, (c, d) T-RANMED and T-ERA5 and (e, f) T-RANMED and T-CON based on the 1997–2016 period. The left column (a, c, e) is for day 1 while the right one (b, d, f) is for day 30 lead time. Percentage differences of orange (blue) colour palette mean lower (higher) flood thresholds respectively in T-CON (vs. T-ERA5) and in T-RANMED (vs. T-ERA5 and T-CON). Panel (g) shows the reference T-ERA5 threshold magnitudes as specific river discharge (river discharge divided by the upstream area in km² in order to scale better between different catchment sizes).

Flood threshold magnitudes, derived from reforecasts, depend on lead-time with values less than 5% different from those derived from the reanalysis (T-ERA5) for day 1 (Figure 7-4a,c,e), but

exceeding 50% difference over large parts of the world by day 30 (Figure 7-4b,d,f), regardless of the ensemble reforecast sampling strategy. However, there is a large spatial variability, with most of the world showing reforecast flood thresholds larger than T-ERA5, except in north Canada, Central and northern South America, Central Asia, the Horn of Arica, and some of east Russia. This confirms earlier finding from Alfieri et al. (2019) that for extended-range lead times, the flood frequency distribution of hydrological forecasts is not well represented by reanalysis simulations.

Figure 7-4 also shows that using the control member of the reforecasts to derive the flood threshold (T-CON) leads to different results than sampling the full reforecast ensemble (T-RAN family). Specifically, T-CON is systematically smaller than T-RANMED in most regions (Figure 7-4e,f), with differences growing with lead time, generally below 20% but reaching 100% in some places by day 30. This suggests that the control forecasts do not fully represent the flood frequency distribution of the ensemble reforecasts and their use could potentially lower the forecast skill.

Results for other severity levels (5, 20, and 50% AEPs) show very similar behaviour. Although there are some variabilities across the severity levels with differences in flood threshold magnitude increasing with the severity level, the percentage differences appear to be in the same order of magnitude (Figures 7-S1 - 7-S3).

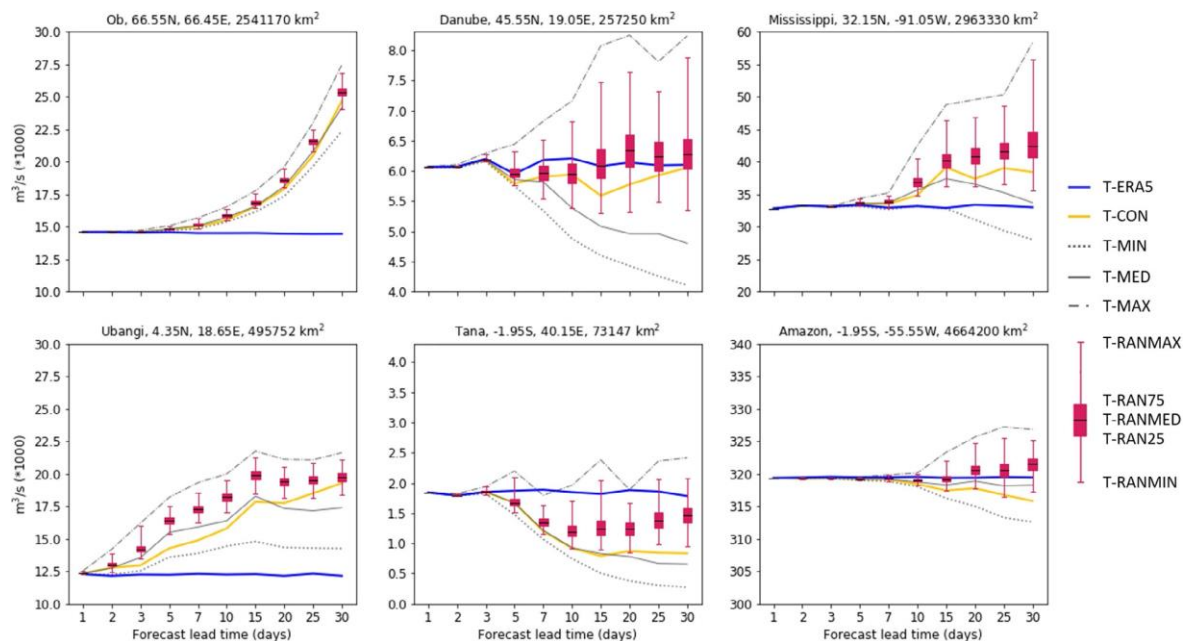


Figure 7-5. Flood thresholds of 10% AEP severity level based on the 1997–2016 period as function of forecast lead time for six contrasting catchments: benchmark T-ERA5 (blue), reforecast control T-CON (orange), minimum T-MIN (dotted grey), median T-MED (solid grey) and maximum T-MAX (dash-dotted grey) and also the extended reforecast T-RANMIN, T-RAN25, T-RANMED, T-RAN75, and T-RANMAX (red box and whiskers).

Figure 7-5 shows the flood thresholds for the 10% AEP severity level as a function of lead time for six contrasting catchments (consult Figure 7-2 for the catchment locations). The influence of the

ensemble reforecast sampling strategy on the flood threshold magnitudes gets larger with the increasing forecast lead time. For some catchments, such as the Ob and Amazon rivers, the impact is small (interquartile range of below 1% of T-ERA5 by day 30 as shown by the red boxes), but for some other catchments the difference could be as large as 10–20% of the T-ERA5 value at day 30 lead time (Tana and Mississippi rivers). Moreover, the flood thresholds, generated using the control member, are dominantly below the envelope of the ensemble reforecast (e.g., Ubangi and Tana rivers), confirming the general positive pattern already seen in Figure 7-4f.

Analysis on other flood threshold severity levels indicates that differences between reforecast- and reanalysis reanalysis-based thresholds and sampling strategies are generally increasing with both severity level and lead time (Figures 7-S4 - 7-S6).

7.3.2 How reliable are the forecast probabilities?

In this section, we investigate the match between the flood forecast probabilities and the flood occurrence frequencies, using the benchmark, the reforecast control, and the extended reforecast median flood thresholds, defined for the 10% AEP severity level (Figure 7-6).

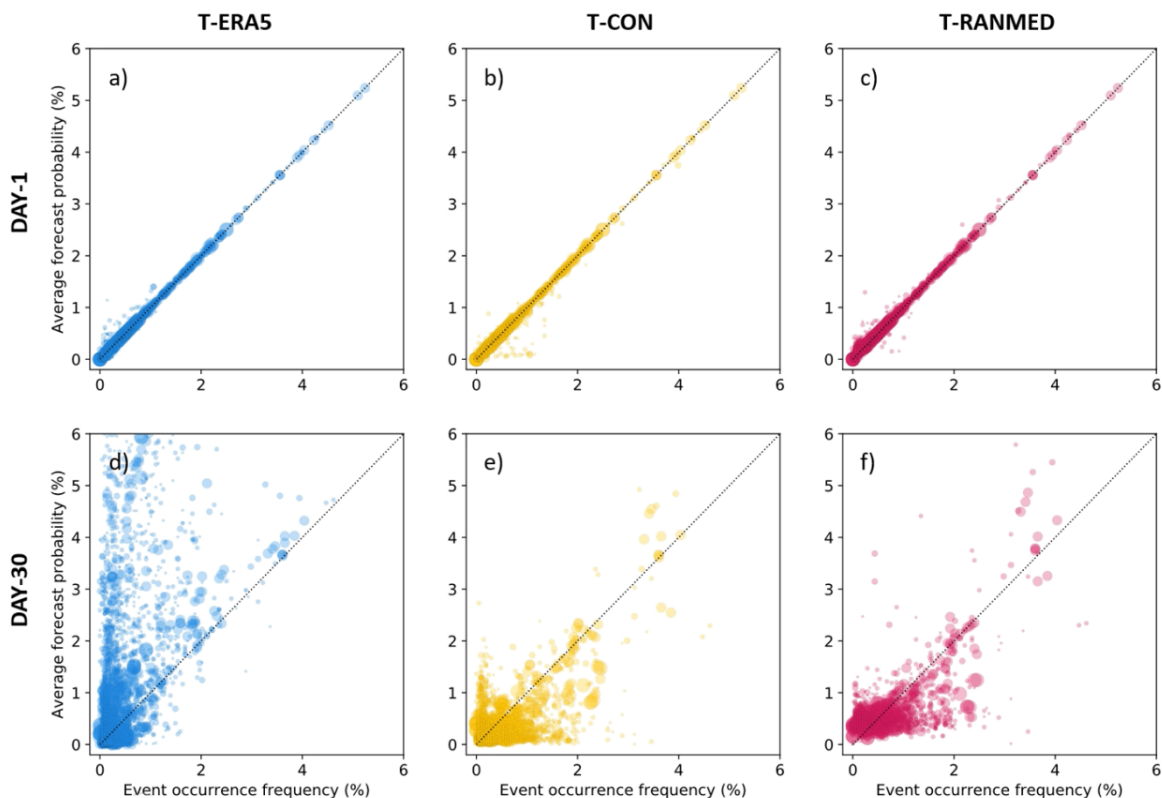


Figure 7-6. Scatter plot of day 1 (top) and day 30 (bottom) flood forecast probability (y-axis) against flood occurrence frequency (x-axis) using flood thresholds of T-ERA5 (a and d, blue), T-CON (b and e, orange), and T-RANMED (c and f, red) based on the 1997-2016 period. Dot size is proportional to catchment size.

At day 1 (Figure 7-6a–c), flood forecasts are very reliable regardless of the flood threshold generation used (points close to the diagonal line), but this is lost by day 30 (Figure 7-6d–f). The largest loss of reliability is found when using the benchmark flood threshold (T-ERA5), with many catchments showing too high flood forecast probability (points way above the diagonal line), suggesting that the T-ERA5 thresholds are too low. The performance using reforecasts-based thresholds shows a clear improvement over using T-ERA5, especially reducing the number of catchments with large flood forecast probability overestimation. Results based on T-RANMED are slightly better than those using T-CON with a larger cluster around the diagonal line (91 vs. 86% of the catchments with less than 0.5% absolute difference between forecast probability and occurrence frequency), showing a stronger, more linear relationship.

7.3.3 What is the impact on forecast reliability and skill?

Forecast reliability and skill were further examined using the reliability diagram and the Brier score for flood forecasts produced with the 10% AEP severity level.

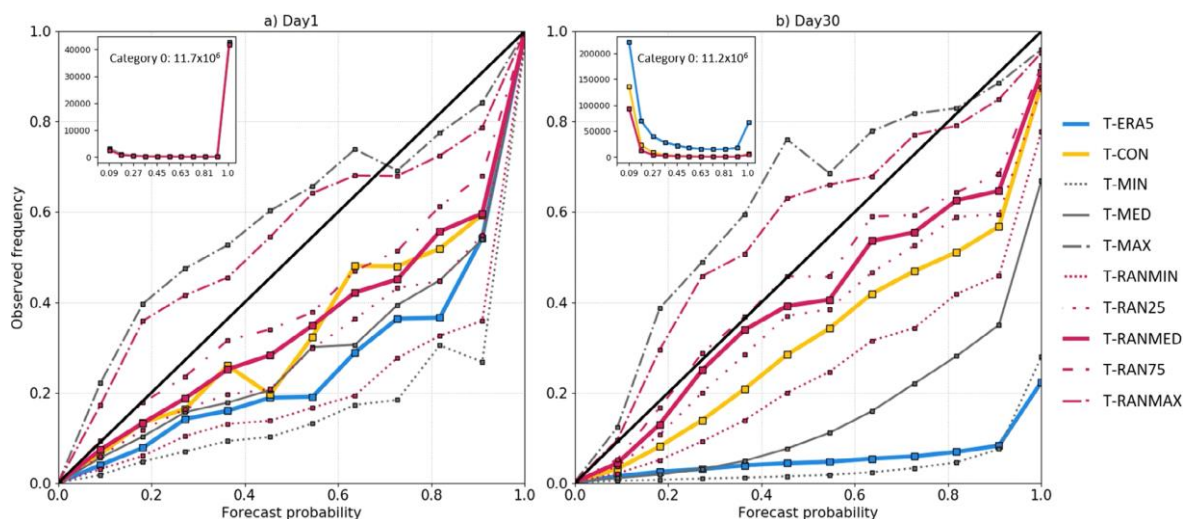


Figure 7-7. Reliability diagram for flood event forecast probabilities above 10% AEP based on the 1997–2016 period for (a) day 1 and (b) day 30 using flood thresholds based on the benchmark (T-ERA5), reforecast (T-CON, T-MIN, T-MED, and T-MAX), and extended reforecast (T-RANMIN, T-RAN25, T-RANMED, T-RAN75, and T-RANMAX) sets. The inset shows the distribution of number of cases in all 11 probability categories. The first category (0 ensemble member forecasting the event) is only indicated as a number.

As suggested by Figure 7-7, the reliability of the flood forecasts, based on the T-ERA5 thresholds, is low, especially for day 30 lead time, with an event frequency of less than 10% for almost all flood forecast probability categories (except the largest when it is just above 20%, see blue line close to x-axis). Using reforecast-based flood thresholds can greatly improve the flood forecast reliability, the only exceptions being T-MIN and T-RANMIN for day 1 lead time. T-MAX and T-RANMAX tend to systematically underestimate flood event frequency up to 70–80% forecast probability, whilst

overestimation of flood events is systematic for all other thresholds. The response is similar across all considered four severity levels except for flood thresholds of 50% AEP, where the reforecast thresholds become too high, making the flood forecast probabilities too low (points are above the diagonal; Figures 7-S7 - 7-S9). Generally, thresholds based on the full ensemble can provide better reliability than using T-CON, this is especially clear by day 30, when all T-RANMED, T-RAN25, and T-RAN75 are closer to the diagonal line.

Figure 7-8 shows the general skill of the flood forecasts for the 10% AEP severity level from day 1 to day 30 lead times, based on the Brier score. The benchmark flood thresholds (T-ERA5) can provide the lowest error only up to day 2 lead time (blue line below the other lines; for other flood severity levels this maximum lead time ranges from day 1 (5% AEP) to day 4 (50% AEP), Figures 7-S10 - 7-S12).

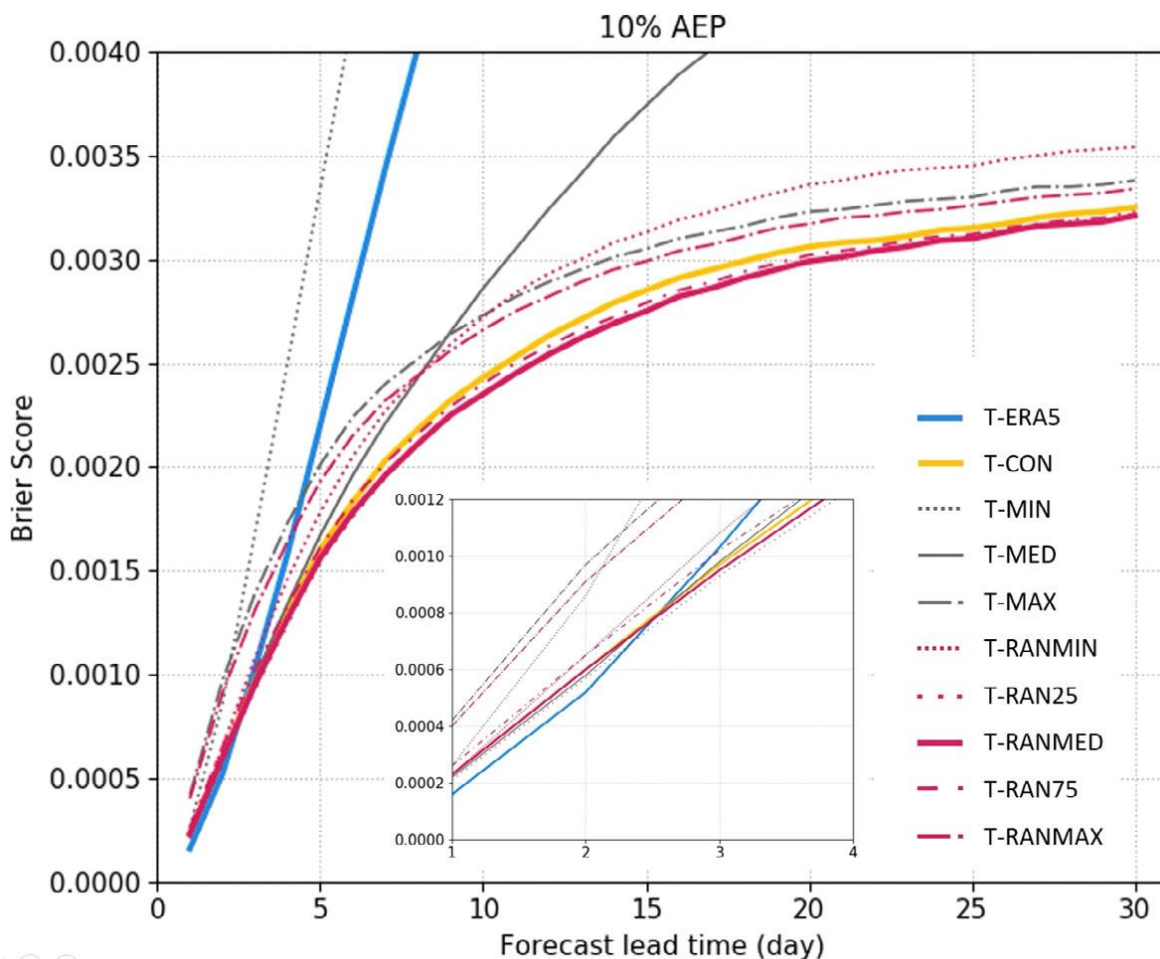


Figure 7-8. Brier score for flood event forecasts above 10% AEP over the 1997–2016 period for day 1 to day 30 using the benchmark (T-ERA5), reforecast (T-CON, T-MIN, T-MED, and T-MAX), and extended reforecast (T-RANMIN, T-RAN25, T-RANMED, T-RAN75, and T-RANMAX) flood thresholds. The inset shows the scores of the first 4 days only for better readability.

From day 3, the flood forecasts with reforecast-based thresholds become gradually more skilful than with T-ERA5 (the only exception is T-MIN), consistently with the conclusions of Figure 7-7b. In fact, both T-ERA5 and T-MIN show poor skill with multiple times higher Brier score values by day 30 (this higher section of the Brier score range is not shown in Figure 7-8 for better readability). The best performance is achieved by the median (T-RANMED) and the interquartile range boundaries (T-RAN25 and T-RAN75) of the extended reforecast threshold set and the reforecast control thresholds (T-CON), with skill slowly degrading with lead time. The skill improvement, using T-RANMED over T-CON, is statistically significant at the 99% level from day 5–6 lead time (tested by bootstrapping the dates in the verification sample, Figure 7-S13). The pattern is similar for flood events of higher severity (5% AEP), whilst for less severe floods (20 and 50% AEP), the highest skill is achieved using T-RAN25 or T-RAN75, but T-RANMED is still achieving high skill (Figures 7-S10 - 7-S12).

7.4 Discussion

The global analysis conducted here showed that using flood thresholds based on reforecasts improved substantially the forecast performance after the first 1-4 days of the forecast range (depending on the flood severity levels) compared with using thresholds based on reanalysis, as done operationally in most forecast systems. One of the key advantages is the lead time specific definition of thresholds, which accounts for the changing representation of extreme event frequencies in the forecasts. Overall, forecast errors are reduced by up to 2-4 times compared with reanalysis-based thresholds, depending on the flood severity and lead time. Results also showed that using the single unperturbed control member to define the thresholds is not sufficient and exploring the full ensembles of the reforecasts in the threshold derivation further increases the forecast reliability and skill.

7.4.1 Ensemble member independence

Using ensemble members allows a better representation of the extreme events in the forecast climatology by increasing the sample size. The ensemble members are correlated to some extent by sharing the same initial condition, especially at the beginning of the 30-day forecast horizon. Correlation between ensemble members reduces with increased lead time when each ensemble member drifts towards becoming an independent and identical random sample from the mode climate. In addition, for each of the reforecast-based threshold methods, only one member was chosen from all the twice weekly reforecasts in the 20-year period in order to increase independence. This guaranteed that the correlation between the individual reforecast values in the climatological sample remained very small. This made them an effectively independent realisation

of the true underlying model climate distribution, ultimately providing an appropriate basis for the extreme value distribution fitting in the flood threshold computation.

7.4.2 Best performing thresholds

The median of the extended reforecast threshold set, produced by using one random ensemble member from the reforecasts, provides the best overall performance, however, for lower severity levels some other reforecast-based thresholds can be slightly better. This can be related to the nonlinear response between reforecast ensemble time series and flood quantile estimation. In particular, with increasing lead time, outliers associated with very high forecasted river discharge become more likely within the 11 ensemble members. The annual maximum selection then will over-represent the high outliers through the random member selection process, as even if only one very high forecast value is selected in 1 or 2 years, it is likely to shift the estimate of the 5–10% AEP flood quantile to a high value. This potential increase of flood threshold value with lead time does not affect the forecast probabilities of flood event occurrence to the same extent, as the probabilities are calculated considering the full ensemble and are influenced much less by these relatively rare outliers in some of the reforecast members. This different effect of outliers on flood thresholds and flood forecast probabilities will translate into inconsistent reliability and skill impact associated with the various ways to sample the reforecasts to produce flood threshold, and could result, in some cases, in favouring a different sampling strategy than picking up the median.

7.4.3 Biases in the forecasts

Using range-dependent flood thresholds, based on ensemble reforecasts, can account for the evolving biases in the forecasts across the forecast range. This study demonstrated that biases can grow large, affecting the extreme event representation and the use of flood thresholds in medium to extended range hydrological forecast systems like GloFAS 30-day. These biases can originate from the meteorological forcing and impact the hydrological simulations, mainly through precipitation and marginally also temperature, humidity, wind, and radiation, as shown by Zsoter et al. (2016) for the first 10 days of the forecasts. Another likely source for the biases is the land data assimilation (LDAS) impact documented by Zsoter et al. (2019). The LDAS can result in not conserving the water budget in coupled land surface models such as used in GloFAS, possibly contributing to biases seen in the GloFAS-ERA5 reanalysis across large parts of the world (Harrigan et al., 2020b). In GloFAS, the reforecasts are initialised from GloFAS-ERA5, but with increasing lead time, the influence of LDAS on reforecast gradually decreases. This means that biases coming from the LDAS impact will remain present in any reanalysis-based flood threshold (in our case GloFAS-ERA5) but will slowly disappear with lead time in reforecasts-based flood thresholds. This

inconsistency is likely to contribute to the large differences between the GloFAS-ERA5- and the ensemble-forecast-based thresholds shown in this study.

7.4.4 Forecast post-processing

Post-processing of the forecasts against the reference dataset used to derive the flood thresholds (i.e., in our case, GloFAS-ERA5) is an alternative to ensemble reforecast-based thresholds. By removing biases in the forecasts (e.g., linear regression or quantile mapping; see Wentao et al., 2017 for a review of methods), the extreme event representation of the forecasts would be expected to become similar to that of the reference dataset or climatology. The use of post-processing techniques to create a consistent system between forecasts and flood thresholds was beyond the scope of this paper but could be pursued in the future.

7.4.5 Modelling system independence

Whilst the research was conducted on the GloFAS flood forecasting system (based on the HTESSEL land surface model), the main findings of this work are expected to be independent of the modelling system, the extreme value fitting method or the sampling period length used. Although a different fitting method or sampling length could inevitably change the flood thresholds locally, in the global context, they are expected to have a neutral impact on the relation of the threshold magnitudes amongst the different annual maxima sampling methods. In addition, the forecast biases are bound to be modelling system related, which will inevitably change the flood threshold behaviour across the forecast lead times. However, the benefit of using ensemble-based, lead-time-specific thresholds is expected to be general and not dependent on the actual underlying bias behaviour. This is supported by the consistent results found using the Lisflood hydrological model in Alfieri et al. (2019), where the reforecast-control-member-based flood thresholds showed significant biases compared with the ERA5-based thresholds, confirming the benefit of using ensemble reforecasts.

7.4.6 Practical recommendations for flood applications

Severe problems can arise in flood forecasting because of the potential issue with inconsistencies between the representation of extreme event frequencies in the thresholds and the forecasts, due to the biases that might be present especially for longer lead times. We recommend that forecast system developers should evaluate these potential inconsistencies for themselves using the methodology presented in this paper. We further recommend that this should be carried out with the use of reforecasts where they are available. But even where this is not the case, attempts should be made to diagnose the biases in the climatological data and the available historical forecasts for potential inconsistencies. Without addressing this inconsistency issue, the reliability and skill of the forecast flood events, and thus the quality of the flood warnings, could be substantially reduced,

which could strongly impact on the decision-making process and ultimately lead to loss of confidence in the products.

In addition, even though the reanalysis-based flood thresholds are proven to be preferred in the first days of the forecast range, the difference in forecast skill to the reforecast-based thresholds is small. We recommend that it is both sensible and practical using the ensemble reforecasts for computing the flood thresholds for all forecast lead times and flood severity levels. Similarly, the best performing thresholds for the more impactful high floods (below 20% AEP) were generated from the median of a large number of random ensemble member selections from each reforecast. Although they are not necessarily the most favourable thresholds for smaller floods, they are the best overall choice and are recommended to be used for flood predictions across all flood severities and forecast lead times.

This study highlighted that flood forecasting applications, such as GloFAS, which use flood thresholds generated from a single time series (reanalysis or observation), can greatly benefit from using ensemble-forecast-based thresholds instead, as a practical and effective way to resolve inconsistencies between forecasts and flood thresholds, and therefore increasing the flood forecast skill.

7.5 Conclusions

Using reliable thresholds in global flood forecasting, that truly reflect the flood event frequencies of the real-time ensemble forecasts across all forecast lead times, is very important. The generation of flood signals with such thresholds can provide the highest forecast reliability and skill, which then gives the best chance to create trust in the users for the application.

In this paper, different annual maxima sampling methods were analysed to generate flood thresholds, using both GloFAS-ERA5 river discharge reanalysis and ensemble reforecasts. The flood thresholds were compared and their impact on the forecast reliability and skill was evaluated.

Reanalysis-based thresholds were found appropriate for the first 1–4 days of the 30-day (depending on the flood severity level) forecast range only. For longer lead times, both global average forecast reliability and skill deteriorate, effectively due to the increasing forecast biases over large parts of the world not accounted for in the reanalysis-based thresholds. The ensemble-forecast-based thresholds provide increasing improvement over the reanalysis-based thresholds for up to the evaluated day 30 lead time. Additionally, using flood thresholds that sample the full ensemble in the reforecast, was found to be overperforming a simple, single member sampling strategy (e.g., using the control reforecast), with generally better reliability and higher skill of the forecast probability.

The results of this study suggest that acknowledging the large uncertainty coming from the data sampling method in flood threshold generation is a crucial step in understanding and improving forecast skill, so that the system configuration that provides the highest reliability and lowest error globally can be found. In turn, better flood forecasts and better flood warnings could be delivered to the public, increasing the confidence and uptake of these products. Ultimately, the increase in confidence in the flood forecasts should result in better flood preparedness for humanitarian and civil protection partners, potentially reducing damages and casualties world-wide.

Acknowledgements. Ervin Zsoter's PhD is supported by the Wilkie Calvert Co-Supported PhD Studentships at the University of Reading. Ervin Zsoter and Christel Prudhomme were supported by the Copernicus Emergency Management Service—Early Warning Systems (CEMS-EWS [EFAS]). Hannah Cloke is supported by the TENDERLY project: Towards END-to End flood forecasting and a tool for Real-time catchment susceptibility UK NERC Flooding From Intense Rainfall (FFIR) programme, NE/K00896X/1. Elisabeth Stephens and Hannah Cloke are supported by the FATHUM project: Forecasts for Anticipatory Humanitarian Action funded by UK NERC as part of their Science for Humanitarian Emergencies & Resilience (SHEAR) programme, NE/P000525/1.

Contribution of this chapter to the thesis. This chapter addressed the objective: "Assess the impact of innovative ways of generating flood thresholds on the skill of global flood forecasts, using hydrological ensemble reforecasts." This work demonstrated that the impact of nonstationary forecast biases can be large in flood forecasting systems. It also showed that the use of lead-time-dependent thresholds, generated from ensemble reforecasts, can provide a superior alternative over the traditional reanalysis-based thresholds, and can deliver large increase of flood forecast reliability and skill.

7.6 Supplementary figures

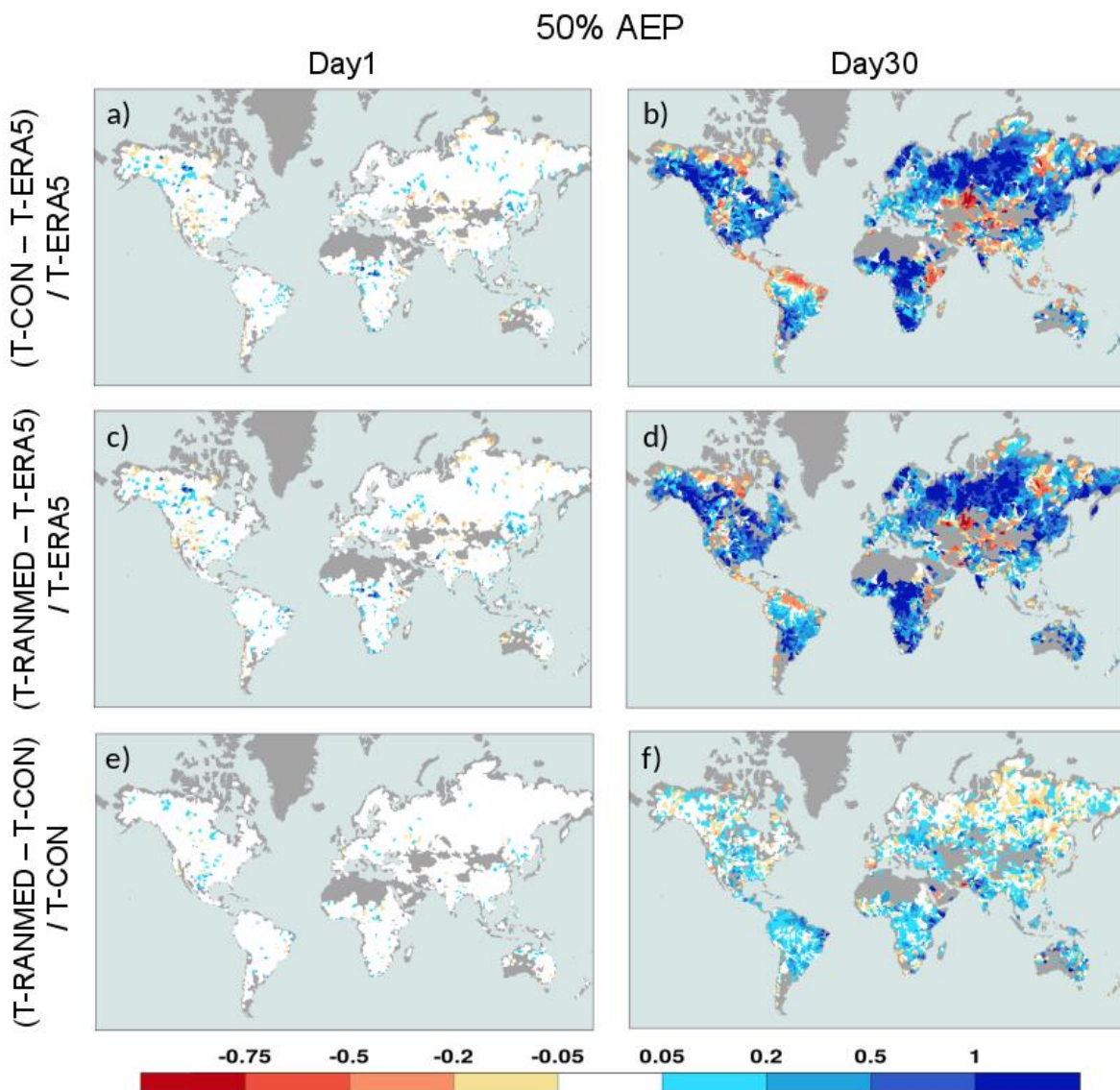


Figure 7-S1. Percentage difference of 50% AEP flood thresholds between a-b) T-CON and T-ERA5, c-d) T-RANMED and T-ERA5 and e-f) T-RANMED and T-CON based on the 1997-2016 period. The left column (a-c-e) is for day1 while the right one (b-d-f) is for day30 lead time. Percentage differences of orange (blue) colour palette mean lower (higher) flood thresholds respectively in T-CON (vs T-ERA5) and in T-RANMED (vs T-ERA5 and T-CON).

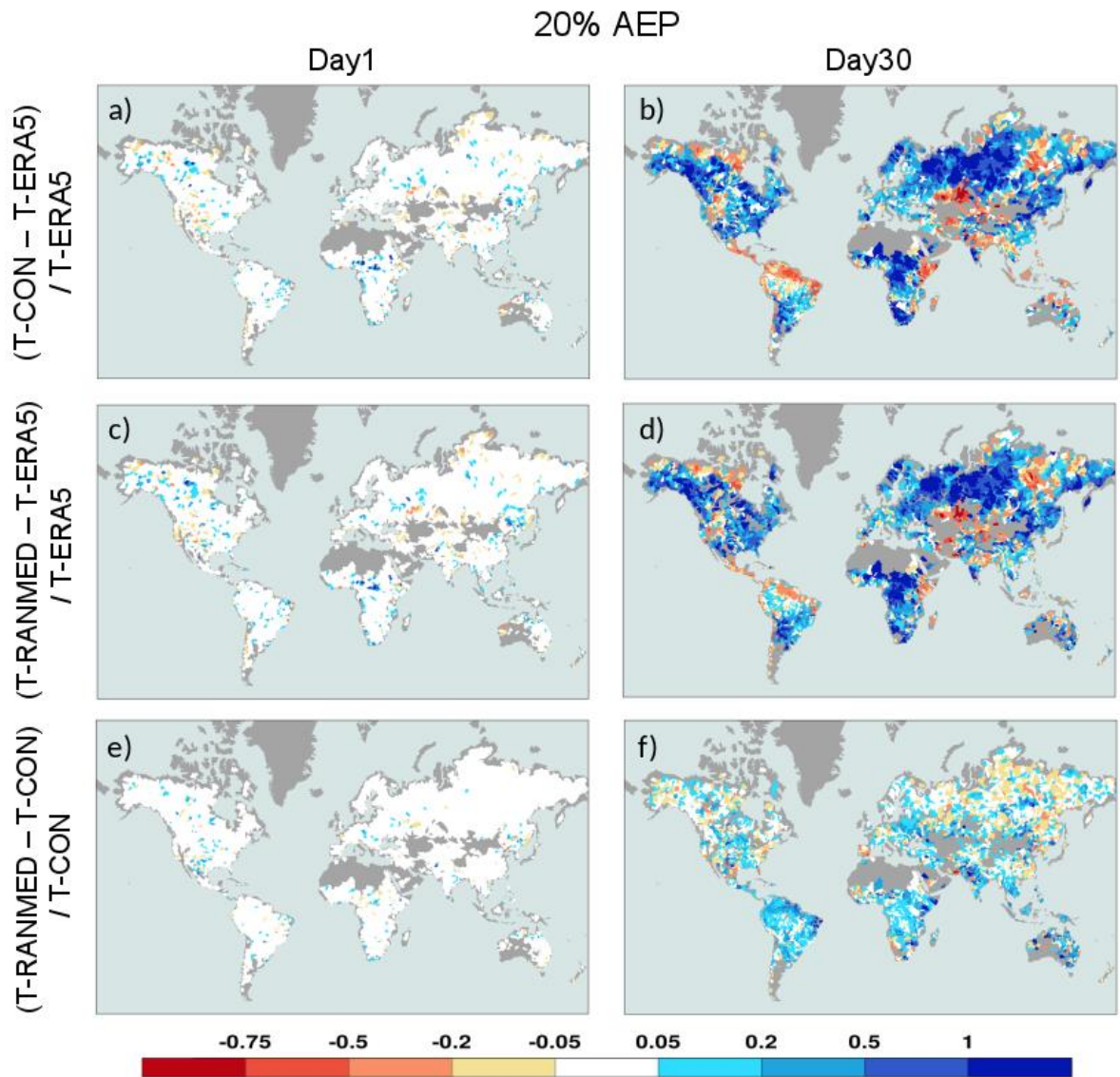


Figure 7-S2. Percentage difference of 20% AEP flood thresholds between a-b) T-CON and T-ERA5, c-d) T-RANMED and T-ERA5 and e-f) T-RANMED and T-CON based on the 1997-2016 period. The left column (a-c-e) is for day1 while the right one (b-d-f) is for day30 lead time. Percentage differences of orange (blue) colour palette mean lower (higher) flood thresholds respectively in T-CON (vs T-ERA5) and in T-RANMED (vs T-ERA5 and T-CON).

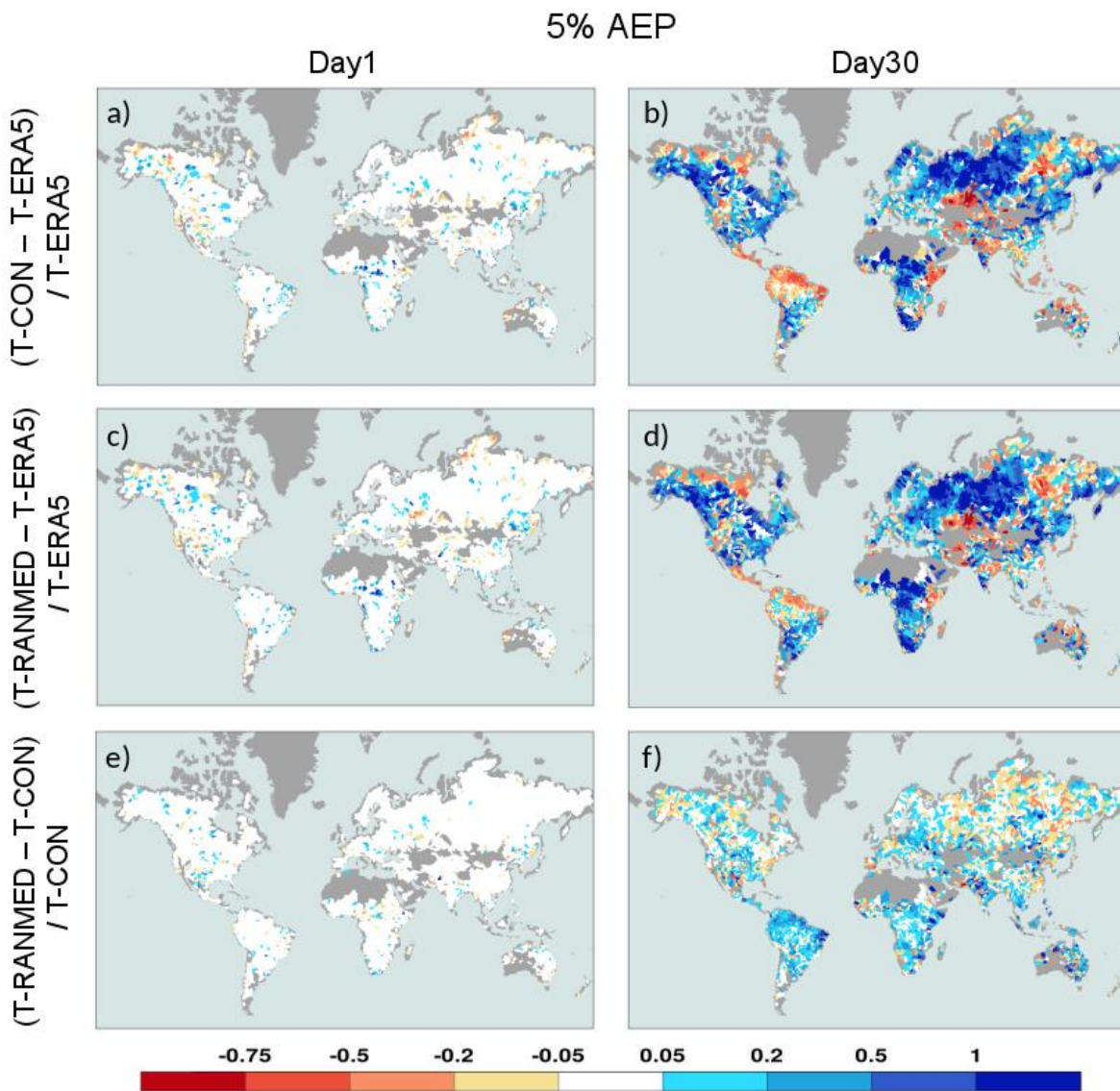


Figure 7-S3. Percentage difference of 5% AEP flood thresholds between a-b) T-CON and T-ERA5, c-d) T-RANMED and T-ERA5 and e-f) T-RANMED and T-CON based on the 1997-2016 period. The left column (a-c-e) is for day1 while the right one (b-d-f) is for day30 lead time. Percentage differences of orange (blue) colour palette mean lower (higher) flood thresholds respectively in T-CON (vs T-ERA5) and in T-RANMED (vs T-ERA5 and T-CON).

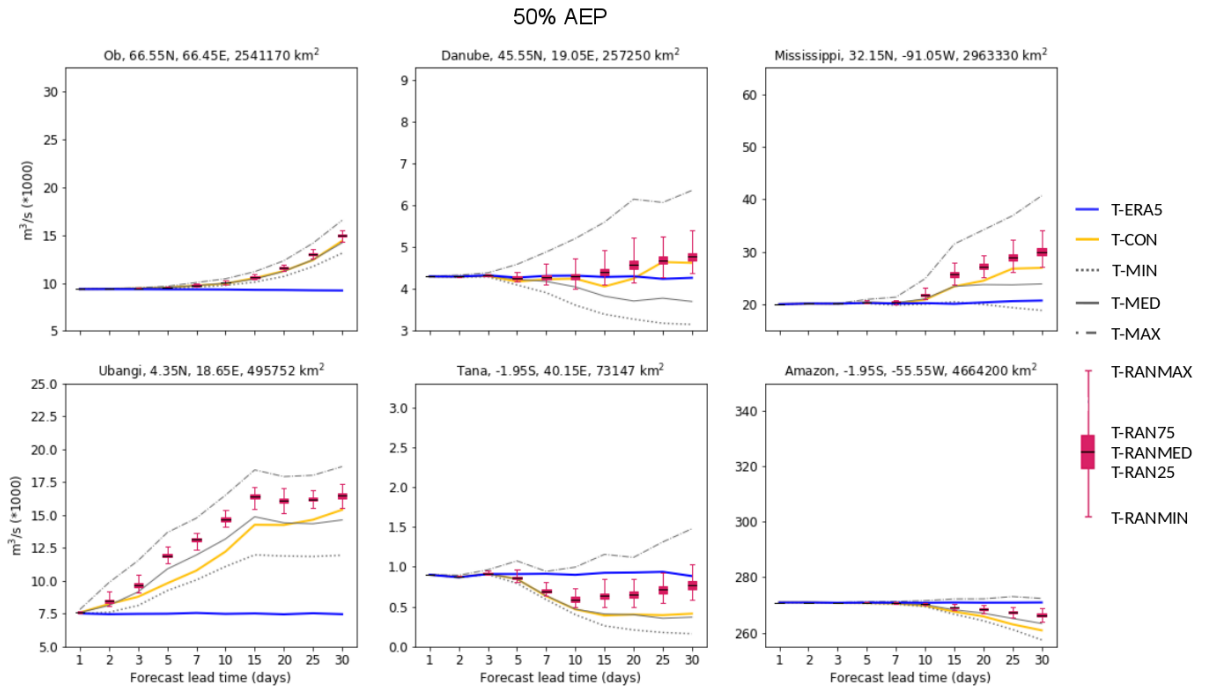


Figure 7-S4. Flood thresholds of 50% AEP severity level based on the 1997-2016 period as function of forecast lead time for six contrasting catchments: benchmark T-ERA5 (blue), reforecast control T-CON (orange), minimum T-MIN (dotted grey), median T-MED (solid grey) and maximum T-MAX (dash-dotted grey) and also the extended reforecast T-RANMIN, T-RAN25, T-RANMED, T-RAN75 and T-RANMAX (red box-whiskers).

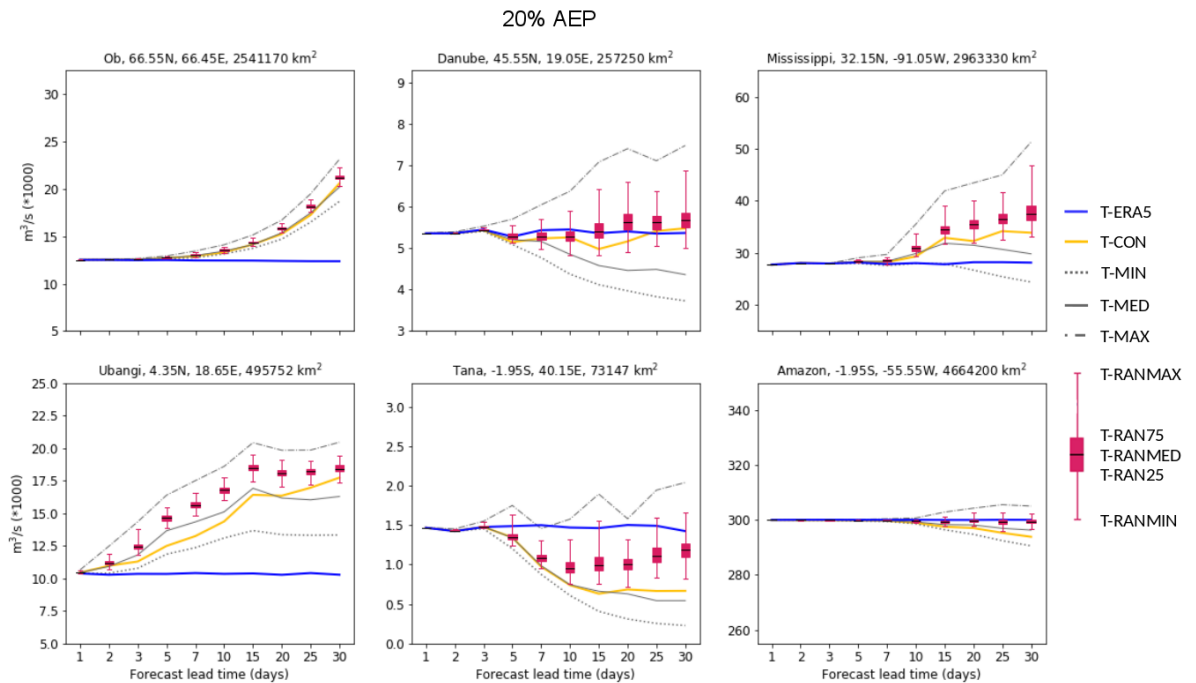


Figure 7-S5. Flood thresholds of 20% AEP severity level based on the 1997-2016 period as function of forecast lead time for six contrasting catchments: benchmark T-ERA5 (blue), reforecast control T-CON (orange), minimum T-MIN (dotted grey), median T-MED (solid grey) and maximum T-MAX (dash-dotted grey) and also the extended reforecast T-RANMIN, T-RAN25, T-RANMED, T-RAN75 and T-RANMAX (red box-whiskers).

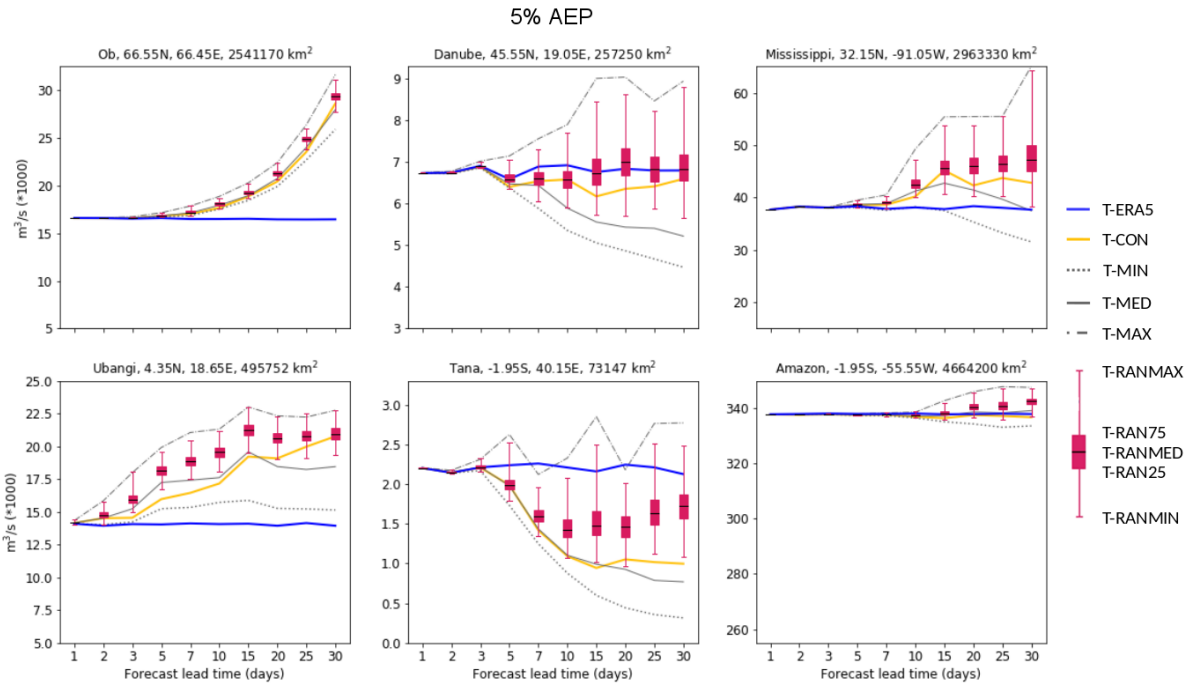


Figure 7-S6. Flood thresholds of 5% AEP severity level based on the 1997-2016 period as function of forecast lead time for six contrasting catchments: benchmark T-ERA5 (blue), reforecast control T-CON (orange), minimum T-MIN (dotted grey), median T-MED (solid grey) and maximum T-MAX (dash-dotted grey) and also the extended reforecast T-RANMIN, T-RAN25, T-RANMED, T-RAN75 and T-RANMAX (red box-whiskers).

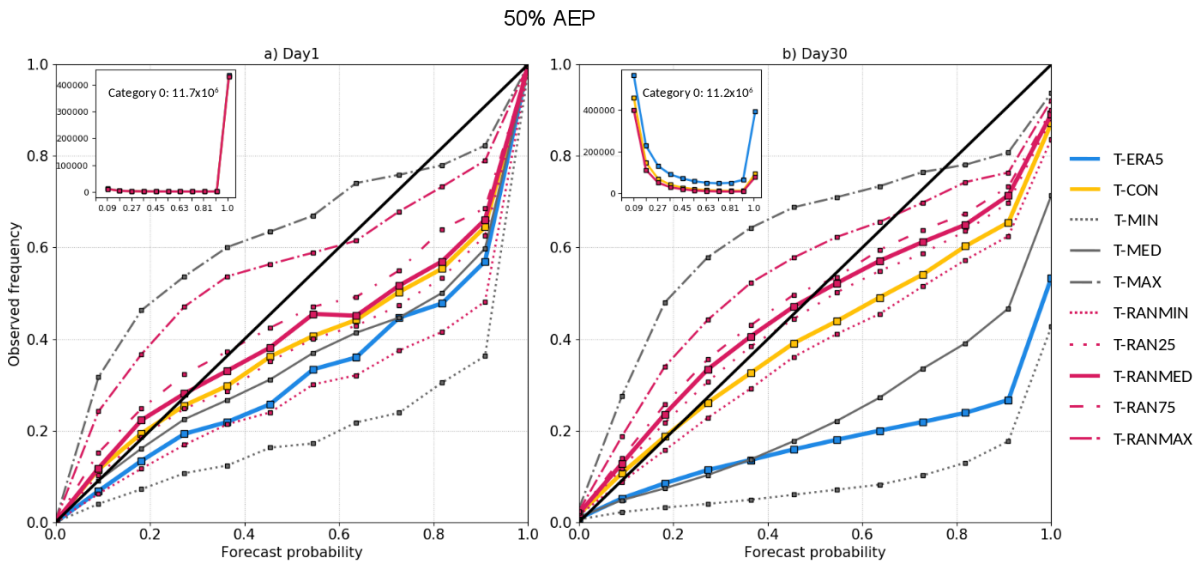


Figure 7-S7. Reliability diagram for flood event forecast probabilities above 50% AEP based on the 1997-2016 period for a) day1 and b) day30 using flood thresholds based on the benchmark (T-ERA5), reforecast (T-CON, T-MIN, T-MED and T-MAX) and extended reforecast (T-RANMIN, T-RAN25, T-RANMED, T-RAN75 and T-RANMAX) sets. The insert shows the distribution of number of cases in all 11 probability categories. The first category (0 ensemble member forecasting the event) is only indicated as a number.

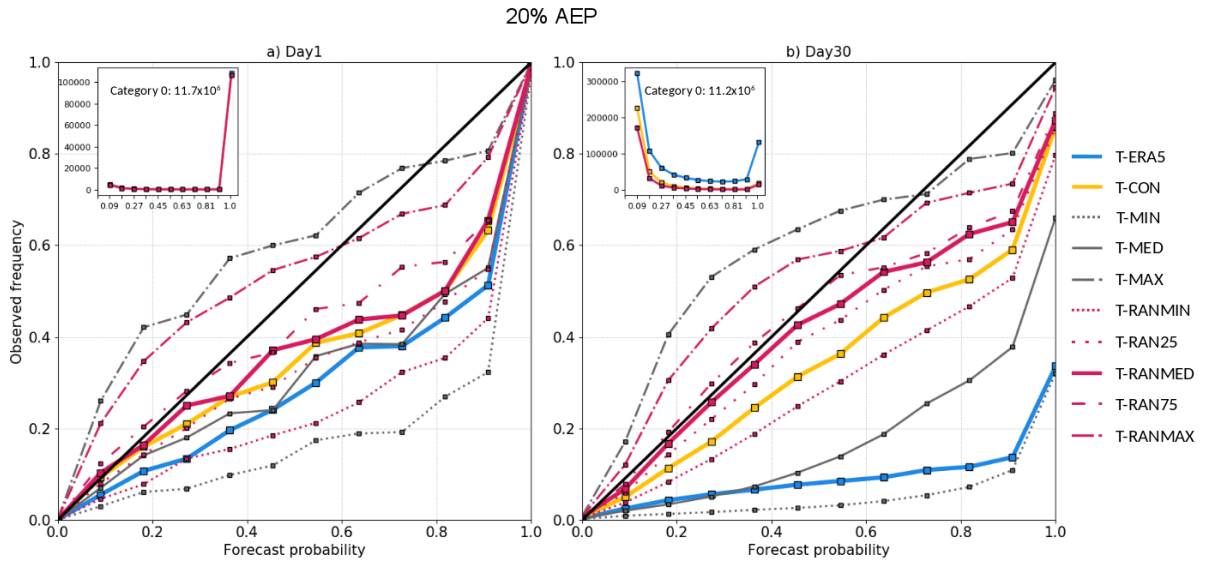


Figure 7-S8. Reliability diagram for flood event forecast probabilities above 20% AEP based on the 1997-2016 period for a) day1 and b) day30 using flood thresholds based on the benchmark (T-ERAS), reforecast (T-CON, T-MIN, T-MED and T-MAX) and extended reforecast (T-RANMIN, T-RAN25, T-RANMED, T-RAN75 and T-RANMAX) sets. The insert shows the distribution of number of cases in all 11 probability categories. The first category (0 ensemble member forecasting the event) is only indicated as a number.

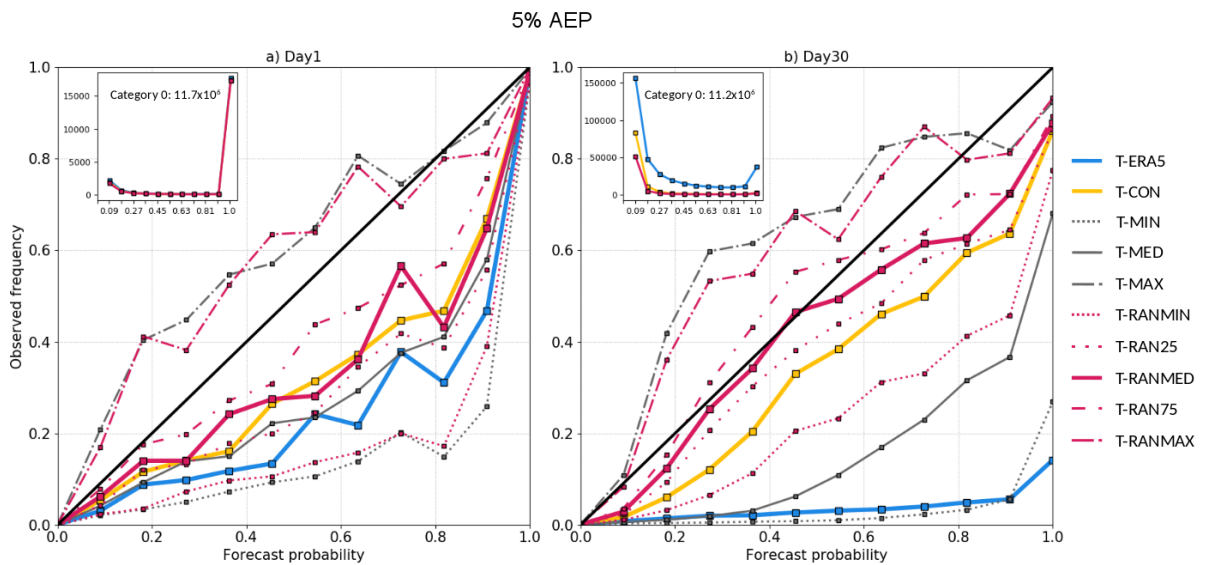


Figure 7-S9. Reliability diagram for flood event forecast probabilities above 5% AEP based on the 1997-2016 period for a) day1 and b) day30 using flood thresholds based on the benchmark (T-ERAS), reforecast (T-CON, T-MIN, T-MED and T-MAX) and extended reforecast (T-RANMIN, T-RAN25, T-RANMED, T-RAN75 and T-RANMAX) sets. The insert shows the distribution of number of cases in all 11 probability categories. The first category (0 ensemble member forecasting the event) is only indicated as a number.

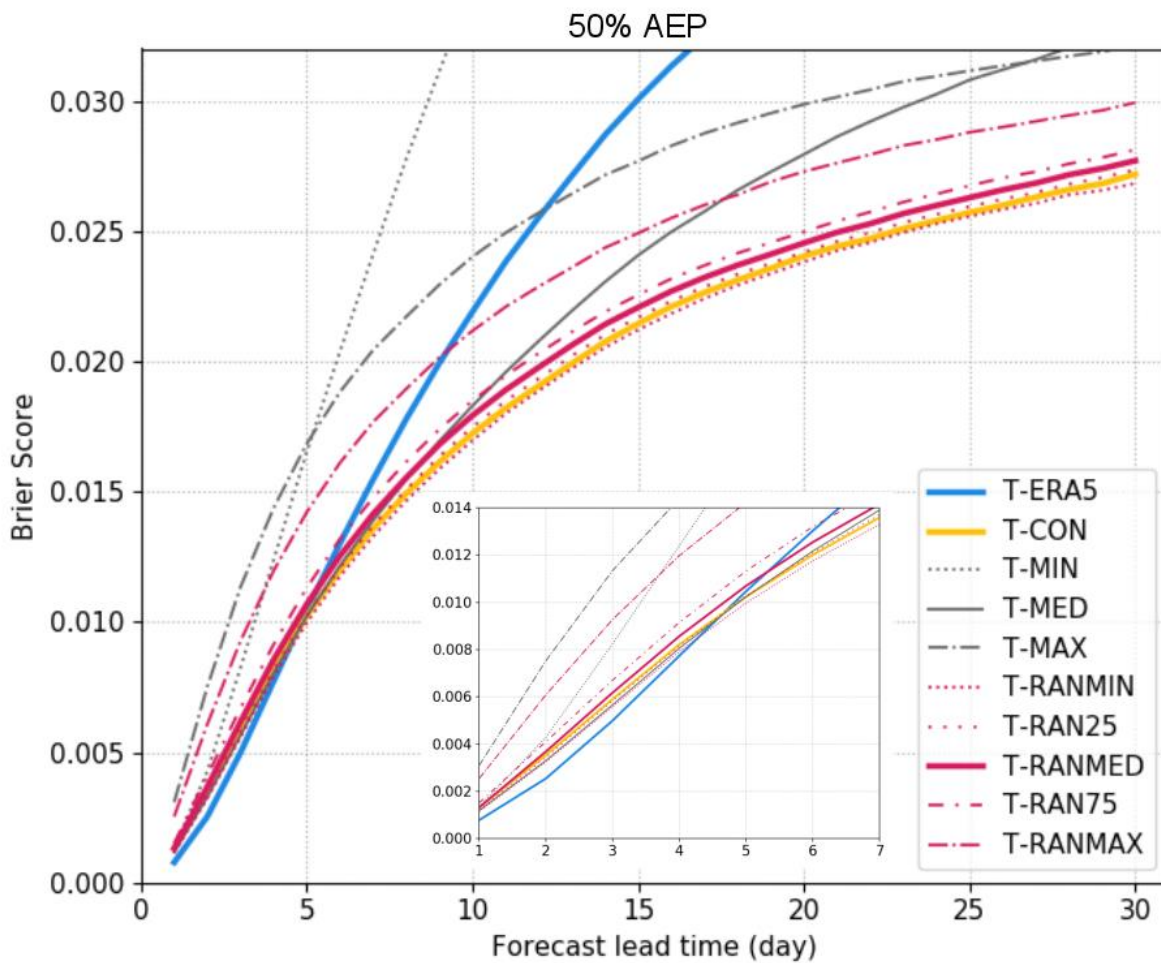


Figure 7-S10. Brier score for flood event forecasts above 50% AEP over the 1997-2016 period for day1 to day30 using the benchmark (T-ERA5), reforecast (T-CON, T-MIN, T-MED and T-MAX) and extended reforecast (T-RANMIN, T-RAN25, T-RANMED, T-RAN75 and T-RANMAX) flood thresholds. The inset shows the scores of the first seven days only for better readability.

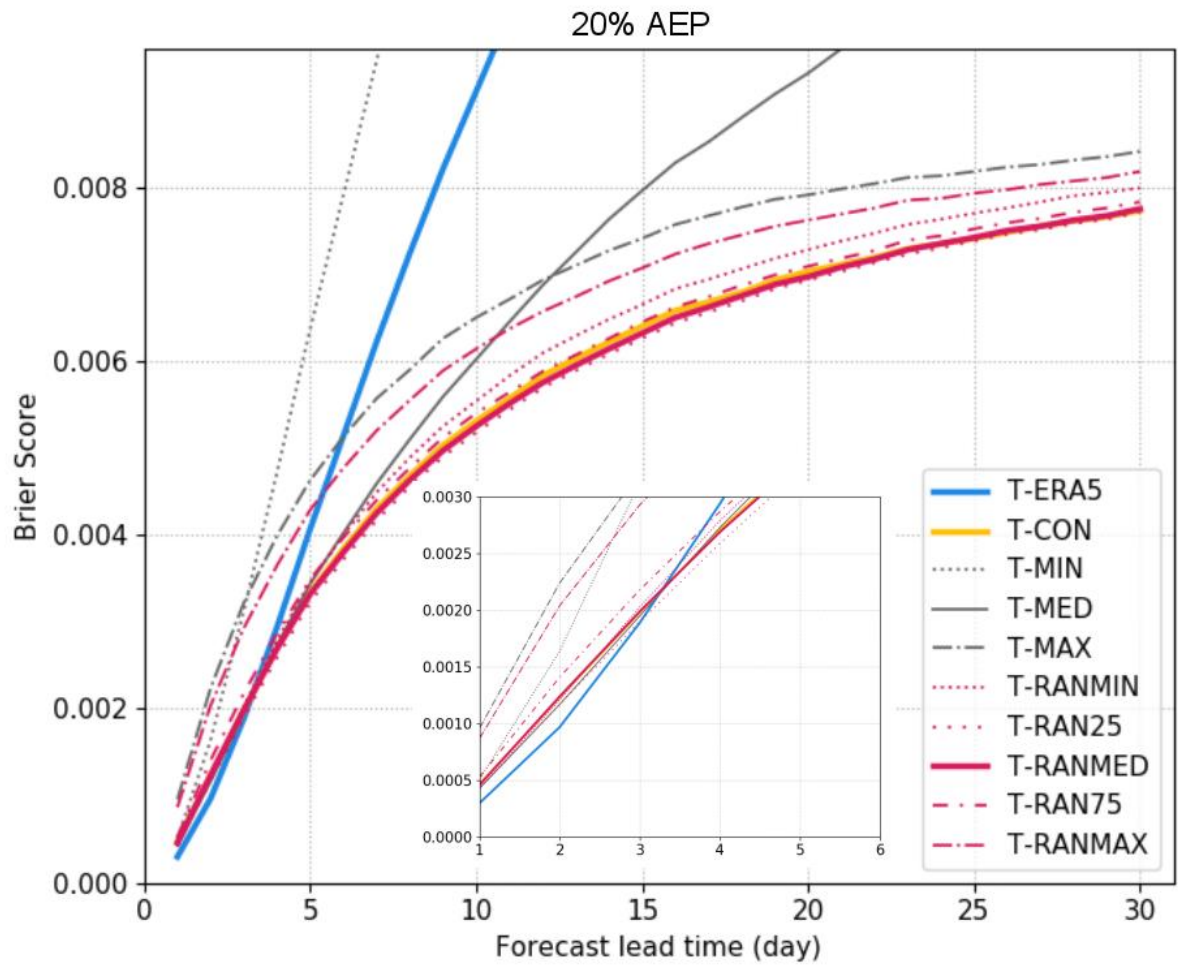


Figure 7-S11. Brier score for flood event forecasts above 20% AEP over the 1997-2016 period for day1 to day30 using the benchmark (T-ERA5), reforecast (T-CON, T-MIN, T-MED and T-MAX) and extended reforecast (T-RANMIN, T-RAN25, T-RANMED, T-RAN75 and T-RANMAX) flood thresholds. The inset shows the scores of the first six days only for better readability.

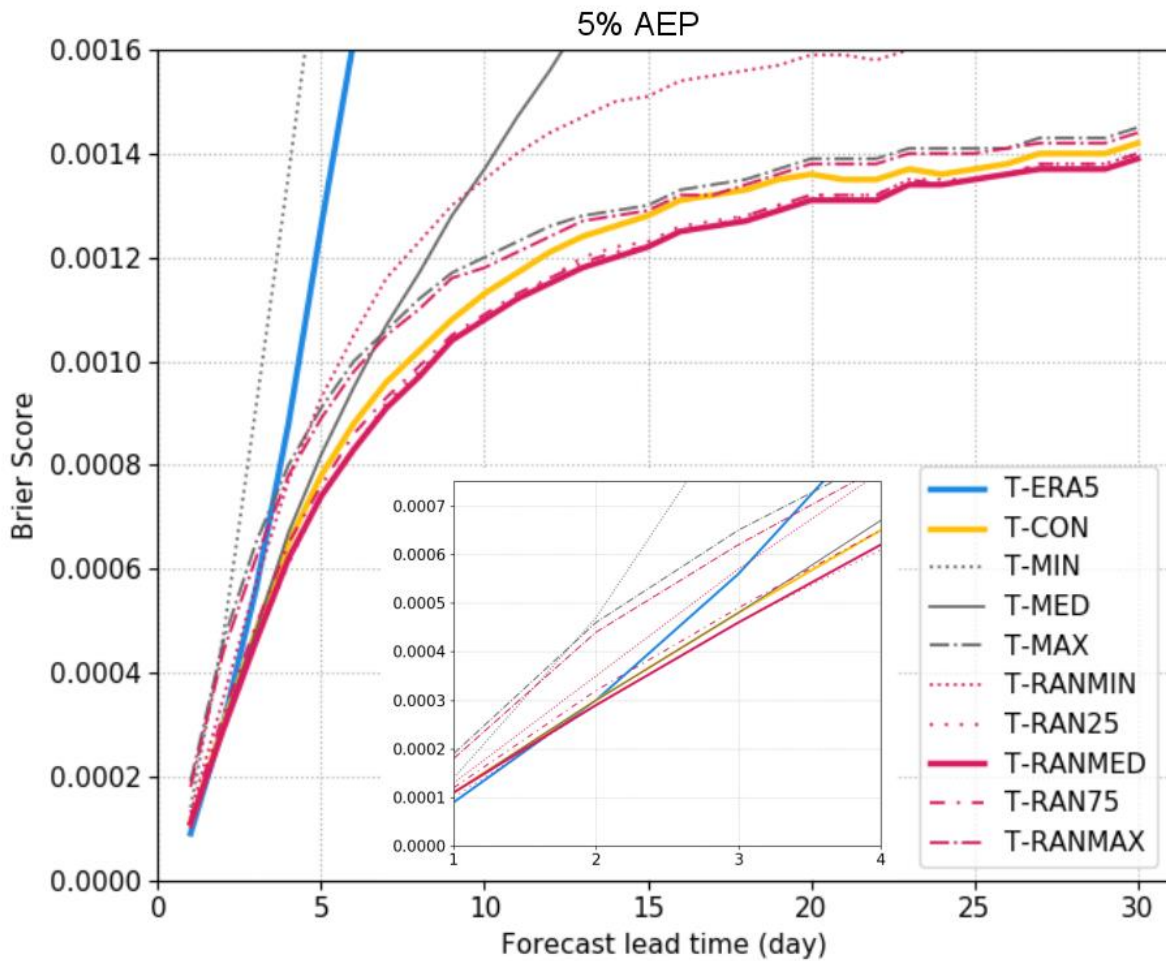


Figure 7-S12. Brier score for flood event forecasts above 5% AEP over the 1997-2016 period for day1 to day30 using the benchmark (T-ERA5), reforecast (T-CON, T-MIN, T-MED and T-MAX) and extended reforecast (T-RANMIN, T-RAN25, T-RANMED, T-RAN75 and T-RANMAX) flood thresholds. The inset shows the scores of the first four days only for better readability.

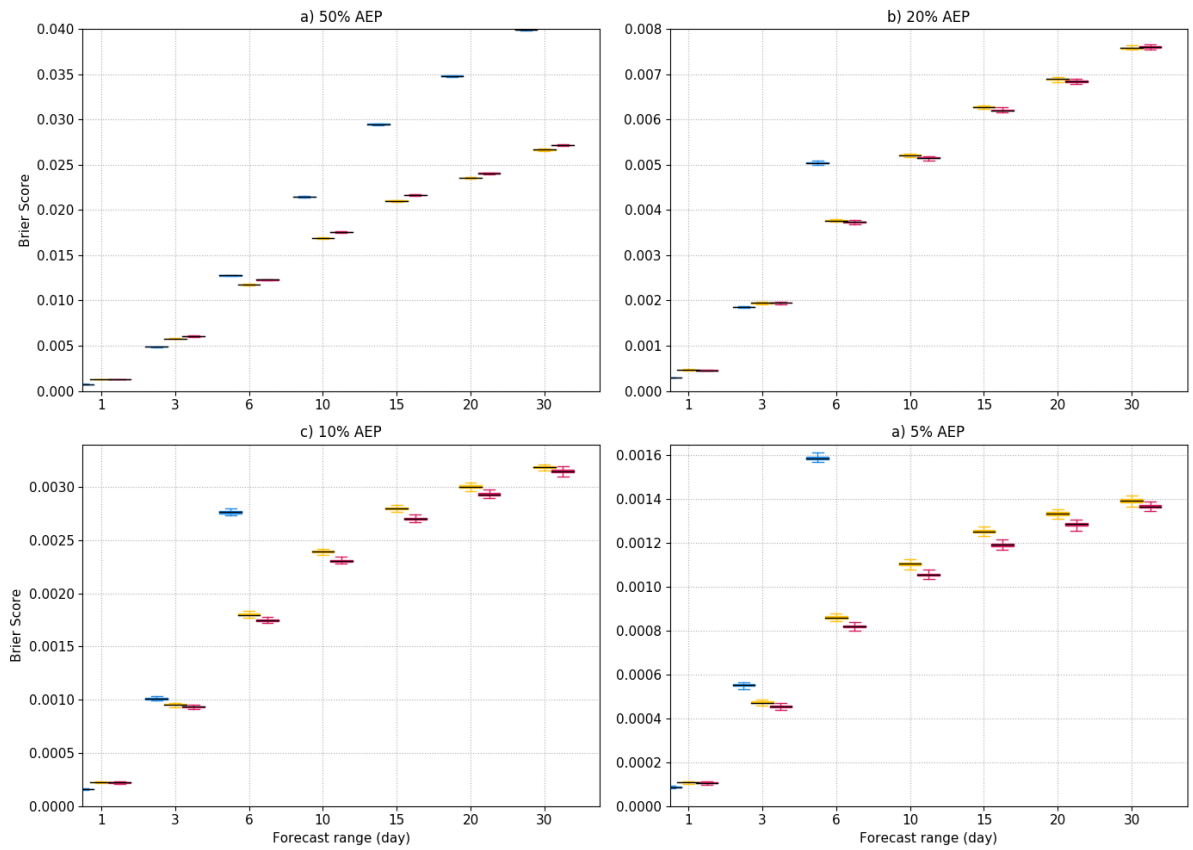


Figure 7-S13. Brier score with significance after bootstrapping the dates in the verification sample, for flood event forecasts above a) 50% AEP, b) 20% AEP, c) 10% AEP and d) 5% AEP, over the 1997-2016 period, for day1 to day30, using the benchmark (T-ERA5 in blue), reforecast control (T-CON in orange) and extended reforecast (T-RANMED in red) flood thresholds.

Chapter 8 Summary of co-authored PhD publications

This PhD has provided a good opportunity to collaborate with numerous scientists in hydrometeorology, mostly around GloFAS and more generally in global hydrology. These papers have extended the horizon of the topics covered in the four main papers (and four objectives), as they cover various aspects of global hydrological simulations. They explore the capabilities for simulating several hydrological aspects in the various reanalysis and forecast data sets, produced during the six-year period of the PhD (listed in Table 3-1). In addition, they include studies about the quality of meteorological forcing data sets and about different methodological aspects of the flood predictions.

In the following, these papers are summarised, presenting the main findings and the paper-specific contributions to the PhD.

8.1 Exploring hydrological reanalysis data sets

8.1.1 *Cao et al. (2022)*

<https://centaur.reading.ac.uk/106794/>: This paper contributes to the PhD with additional understanding of the impact of the new ECMWF multi-layer snow scheme specifically on soil temperature in permafrost areas. This work is strongly related to the study presented in Chapter 5 with the impact of the multi-layer snow scheme on hydrology, concentrating on river discharge. It evaluates some of the same experiments explored in Chapter 5 and compares them for improvements to the soil temperature biases in the cold permafrost areas. Ervin Zsoter contributed to the experimental design, the interpretation of results and the writing of the manuscript.

8.1.2 *Winkelbauer et al. (2022)*

<https://centaur.reading.ac.uk/106795/>: This paper contributes to the PhD with additional understanding of the river discharge behaviour produced from land surface models. Different reanalysis products were compared, including ERA5, ERA5-Land, GloFAS-ERA5-v2.1 and the latest GloFAS-ERA5-v3.1, for runoff and river discharge seasonal cycle and annual trends in the Arctic region and water volumetric budget analysis for the atmosphere, land and ocean. It was shown that GloFAS-ERA5-v3.1 outperforms the other data sets in the river discharge volume terms and trends in the Arctic region. Ervin Zsoter contributed to the production and extraction of the GloFAS-ERA5v3.1 data set and to the interpretation of results and the writing of the manuscript.

8.1.3 *Ficchi et al. (2021)*

<https://centaur.reading.ac.uk/99016/>: This work contributes to the knowledge about the GloFAS river discharge reanalysis. It compares different climate modes of ENSO, the Indian Ocean Dipole and the

Tropical South Atlantic SST mode for their role in driving flood hazard over sub-Saharan Africa. For this, the ERA-Interim/Land-forced river discharge reanalysis (GloFAS-ERA-Land-v1.0) is used, through flood frequency analysis and comparing the flood event probabilities for 5-year return period floods. The results show that Indian and Atlantic Ocean modes of climate variability are equally as important as ENSO for driving changes in the frequency of impactful floods across Africa. Ervin Zsoter contributed to data curation, methodology, investigation and writing the original draft.

8.1.4 Titley et al. (2021)

<https://centaur.reading.ac.uk/97991/>: This study contributes to the knowledge about the GloFAS river discharge reanalysis with the relation between tropical cyclones and flood hazard. It examines landfalling cyclones to identify characteristics, including size, intensity, speed or antecedent weather conditions, that influence the severity of flood hazard arising from these tropical cyclones, based on the GloFAS-ERA5-v2.1 data set. Ervin Zsoter contributed to the GloFAS data curation, interpretation of the results and writing of the manuscript.

8.1.5 Muñoz-Sabater et al. (2021)

<https://centaur.reading.ac.uk/106796/>: This paper contributes to the knowledge about the GloFAS river discharge reanalysis with analysing the impact of the offline simulation methodology (without the atmosphere-land coupling and land data assimilation) and the temperature downscaling on the river discharge quality by using the GloFAS-ERA5-Land-v2.0 data set. Ervin Zsoter contributed by generating the river discharge data sets and the skill evaluation.

8.1.6 Harrigan et al. (2020b)

<https://centaur.reading.ac.uk/106797/>: This paper contributes to the PhD by introducing and evaluating the GloFAS-ERA5-v2.1 river discharge reanalysis as a global gridded dataset with a horizontal resolution of 0.1 degree at a daily time step. The reanalysis was evaluated against a global network of 1801 daily river discharge observation station and was found to be skilful in 86% of catchments according to the modified Kling–Gupta efficiency skill score. Ervin Zsoter contributed to writing the suite to produce the dataset and to the editing of the paper and to the discussion and interpretation of results.

8.1.7 Alfieri et al. (2020)

<https://centaur.reading.ac.uk/106801/>: This paper contributes to the PhD by presenting a new and improved calibration with evaluation of the GloFAS model for version 3. The calibration was performed with the reanalysis period over 1980-2018 using the ERA5 meteorological forcing. GloFAS v3.1 uses the operationally implemented version of the Lisflood model, whereas this paper is based on an earlier

research version of the same Lisflood model. The paper presented the methodology of the calibration and analysed the impact on the GloFAS reanalysis (GloFAS-ERA5-v3.1) skill by comparing it to the previous model version v2.1. Ervin Zsoter contributed with reviewing and editing the manuscript.

8.1.8 Towner et al. (2019)

<https://centaur.reading.ac.uk/85114/>: This study contributes to the PhD by comparing various river discharge reanalyses, including GloFAS v1.0 and v2.0 and CaMa-Flood data sets (GloFAS-ERA5-Land-v1.0, GloFAS-ERA5-v1.0, GloFAS-ERA5-Land-v2.0, GloFAS-ERA5-v2.0, GloFAS-RECF-v2.0 and CAMA-ERA5-Land), for their ability to simulate floods in the Amazon basin. The paper presents an intercomparison of eight different global hydrological models from the collaborators of the Global Flood Partnership using gauged observations as truth. As well as highlighting regional variability in the accuracy of simulated streamflow, these results indicate that the meteorological input is the dominant control on the river discharge accuracy and that groundwater and routing calibration of Lisflood has no impact on the ability to simulate flood peaks in the Amazon basin. Ervin Zsoter provided data and information for all simulations incorporating Lisflood and for the ERA5-Land/H-TESSEL/CaMa-Flood set-up and also worked on producing the simulations. He was also involved in discussions throughout the development and commented on the manuscript.

8.1.9 Hersbach et al. (2018)

<https://centaur.reading.ac.uk/106838/>: This ECMWF report contributed to the PhD by providing an overview of ECMWF's atmospheric, ocean and land reanalysis activities, in particular the latest reanalysis system, ERA5, which is the basis for the GloFAS-ERA5 reanalysis product. It presents various aspects of ERA5's performance, including for simulating river discharge by ERA5-Land (using GloFAS-ERA5-Land-v2.0). Ervin Zsoter contributed by generating river discharge simulations based on ERA5-Land and analysing its quality.

8.1.10 Hirpa et al. (2018b)

<https://centaur.reading.ac.uk/106869/>: This paper contributes to the PhD by presenting the first calibration and evaluation of GloFAS for version 2.0, with the related evaluation of the river discharge simulations (based on GloFAS-RECF-v2.0). Here, the Lisflood routing and groundwater model parameters are calibrated with ECMWF reforecasts as forcing, from 1995 to 2015, using daily streamflow data from 1287 stations worldwide. The optimisation and the verification are both based on the KGE metric. Results show that the calibration could improve the simulation skill for the large majority of catchments. Ervin Zsoter helped with generating the simulations and reviewing and editing of the manuscript.

8.1.11 *Emerton et al. (2017)*

<https://centaur.reading.ac.uk/68946/>: This study contributed to the PhD by exploring the river discharge reanalysis for the presence of ENSO teleconnections, which can potentially help with reducing flood risk. The analysis was undertaken using the CAMA-ERA20CM-R river flow reconstruction for the twentieth century. Results show that the likelihood of increased or decreased flood hazard during ENSO events is much more complex than is often perceived and reported; probabilities vary greatly across the globe, with large uncertainties inherent in the data and clear differences when comparing the hydrological analysis to precipitation. Ervin Zsoter produced the centurial river discharge reanalysis data set and also commented on the manuscript.

8.2 Exploring hydrological forecast data sets

8.2.1 *Harrigan, et al. (2023)*

<https://centaur.reading.ac.uk/106766/>: This work contributed to the PhD by introducing the GloFAS model components and configuration used to generate the v2.1/v2.2 real-time river discharge forecasts and the v2.1/v2.2 reforecasts (GloFAS-ERA5-RFC-v2.2) with 2019 as reference year. It also provides forecast skill information using the reforecasts and the persistence and climatology as verification benchmarks, based on the v2.1/v2.2 reforecasts with 2019 as reference year. Ervin Zsoter contributed to developing the GloFAS suites to produce the reforecasts and real-time forecasts and to the editing of the manuscript and to the discussion and interpretation of results.

8.2.2 *Bischirotis et al. (2020)*

<https://centaur.reading.ac.uk/106803/>: This study contributed to the PhD by information on the usability of the GloFAS reforecasts in forecast-based financing (FbF) decisions. It presents a methodology that compares permanent with forecast-based flood-prevention measures on the bases of a 23-year period in a case study area in Malawi, based on the GloFAS-ERA5-RFC-v2.0, the v2.0 GloFAS reforecasts with 2017 as reference year. The results indicated that the choice between permanent and temporary measures is affected by the cost of measures, climatological flood risk, and forecast ability to produce accurate flood warnings. Results also showed that a combination of the two types of measures can be the most cost-effective solution, particularly when the forecast is more skilful in capturing low-frequency events. Ervin Zsoter contributed to the generation of the GloFAS simulations and forecasts and the writing of the manuscript.

8.2.3 *Passerotti et al. (2020)*

<https://centaur.reading.ac.uk/106804/>: This paper contributed to the PhD with knowledge about the local usability of GloFAS for early warnings in an important catchment in the Sahel. It has compared the performance of two glofas versions with reforecasts, the daily reforecasts from GloFAS-ERA5-Land-RFC-v1.0

and the twice-weekly reforecasts from GloFAS-ERA5-RFC-v2.0. The comparison considered the raw and the post-processed reforecasts with a simple linear regression. The results highlighted the limitations of the raw GloFAS forecasts and confirmed that a simple optimization could improve the biases and improve the overall performance. It was concluded that the post-processed GloFAS v2.0 forecasts are acceptable for the regional early warning system. Ervin Zsoter contributed to the methodology, interpretation of the results, reviewing of the manuscript and producing the GloFAS reforecast data sets.

8.2.4 *Bischiniotis et al. (2019)*

<https://centaur.reading.ac.uk/106837/>: This study contributed to the PhD by evaluating the GloFAS system skill in Peru, using both the ERAI-Land reanalysis and the daily reforecasts produced with the preoperational version of GloFAS, available at the time. It included verification of the reanalysis (GloFAS-ERAI-Land-v0) against gauged observations, verification of the reforecasts (GloFAS-ERAI-Land-RFC-v0) against reanalysis out to day 15 lead time on each river pixel, including quantile mapping post-processing and it also included the comparison with reported flood events in the verification period of 2009-2015. Results showed that GloFAS is able to predict the majority of the observed flood events and even though the biases are substantial in the raw forecast, the quantile-mapped version is able to provide much higher skill for most of the rivers in Peru. Ervin Zsoter contributed by the generation of the GloFAS reanalyses and reforecasts and to the reviewing and editing of the manuscript.

8.2.5 *Emerton et al. (2018)*

<https://centaur.reading.ac.uk/78570/>: This paper contributed to the PhD by introducing and evaluating the first operational global-scale seasonal hydrometeorological forecasting system, GloFAS-Seasonal. This system was introduced as an extension of the preoperational GloFAS. The paper described the key hydrometeorological components and computational framework of GloFAS-Seasonal, alongside the forecast products available and some initial evaluation of the forecasts (GloFAS-ERA5-SRFC-v1.0). Ervin Zsoter contributed by generating the reanalysis and reforecast data sets and building the operational production system. He was also involved in discussions throughout development and gave comments on the manuscript.

8.3 Exploring meteorological forcing data sets

8.3.1 *Lavers et al. (2018)*

<https://centaur.reading.ac.uk/106840/>: This paper contributed to the PhD by exploring the predictability related to precipitation and water vapour transport in the meteorological forcing data. It analysed the skill of the Extreme Forecast Index (EFI) for the integrated water vapour transport (IVT) over the Iberian Peninsula and compared it to the precipitation EFI using real time 15-day ensemble forecasts. Results show that IVT EFI has slightly more skill than the precipitation EFI in discriminating extreme precipitation

anomalies across the western Iberian Peninsula at the extended range of day 11 onwards. Ervin Zsoter contributed to the preparation of the data sets used and the reviewing and editing of the manuscript.

8.3.2 Coughlan de Perez et al. (2018)

<https://centaur.reading.ac.uk/77889/>: This study contributed to the PhD by exploring the predictability of temperature extremes in the reforecast meteorological forcing data. In this study, the areas are assessed at global level, where there is a potential to reduce risk from temperature extremes. For this, the climatologies of heatwaves and cold waves with the seasonality of these extremes are established and the medium-range predictability of these extreme events are explored. Results showed that in fact, almost 5 billion people live in regions that have seasonality and evidence of predictability for heatwaves and/or cold waves, which opens the possibility for climate adaptation investments to reduce risks to vulnerable populations. Ervin Zsoter contributed by generating the daily extreme temperature data from the ECMWF data sets, helped with interpretation of the results and commented on the manuscript.

8.3.3 Lavers et al. (2017)

<https://centaur.reading.ac.uk/106883/>: This paper contributed to the PhD by exploring the predictability of precipitation and water vapour transport in the meteorological forcing data. It analysed the skill of the IVT EFI globally and compared it to the precipitation EFI using real time 15-day ensemble forecasts over two winters. Results show that the IVT EFI is more skilful than the precipitation EFI in forecast week 2 over Europe and western North America. This is related to the large-scale nature of the IVT, its higher predictability and its relationship with extreme precipitation, which can potentially lead to earlier awareness of extreme precipitation in these areas. Ervin Zsoter contributed to the preparation of the data sets used, helped with the design of the study, the interpretation of the results and the reviewing and editing of the manuscript.

8.3.4 Lavers et al. (2016)

<https://centaur.reading.ac.uk/106884/>: This work contributed to the PhD by exploring the predictability of precipitation and water vapour transport in the meteorological forcing data. It analysed the skill of the IVT EFI and compared it to the precipitation EFI using real time 10-day ensemble forecasts over three winters over Europe. We show that the IVT EFI is more able (than the precipitation EFI) to capture extreme precipitation in forecast week 2 during positive North Atlantic Oscillation (NAO) phase; conversely, the precipitation EFI is better during the negative NAO phase and at shorter leads. This can potentially lead to earlier awareness of extreme precipitation in the extended-range forecasts. Ervin Zsoter contributed to the preparation of the data sets used and the reviewing of the manuscript.

8.4 Exploring global flood prediction methodologies

8.4.1 *Baugh et al. (2020)*

<https://centaur.reading.ac.uk/106836/>: This study contributed to the PhD by exploring the impact of soil moisture assimilation on the GloFAS river discharge quality. It analysed the impacts of Soil Moisture and Ocean Salinity (SMOS) soil moisture data assimilation in the ECMWF IFS by coupling the surface and subsurface runoff outputs to Lisflood (using the Lisflood-routing model as in GloFAS-ERA5-v2.1). Results confirmed that GloFAS river discharge skill is generally affected by only a small amount, larger impact can be expected only during some high flow events on surface runoff. Ervin Zsoter contributed to the data curation, result interpretation and the writing of the manuscript.

8.4.2 *Alfieri et al. (2019)*

<https://centaur.reading.ac.uk/106839/>: This paper contributed to the PhD by introducing a novel approach to estimate flood thresholds from using hydrological reforecasts (similar to GloFAS-ERA5-RFC-v3.1). It was based on a 21-year reforecast-based dataset to derive flood thresholds with six-week lead time and compared these with thresholds derived from the ERA5 reanalysis (similar to GloFAS-ERA5-v3.1) in order to check consistency. This study used an earlier research version of the same Lisflood model that was implemented in GloFAS v3.1. Ervin Zsoter helped with generating the reanalysis and reforecast simulations and reviewing and editing of the manuscript.

8.4.3 *Santoro et al. (2018)*

<https://centaur.reading.ac.uk/106873/>: This paper contributed to the PhD by presenting the technological complexities for building a use-case scenario of providing river discharge ensemble simulation based on the TIGGE (THORPEX Interactive Grand Global Ensemble) meteorological archive and observation data in an interoperable web system. The paper describes the work performed to address multidisciplinary interoperability challenges related to river discharge modelling, validation and display. This includes definition and standardization of domain specific interoperability standards for hydrological data sharing and their support in global frameworks such as GEOSS (Global Earth Observation System of Systems). Ervin Zsoter contributed by generating the river discharge simulations and in general to the methodology behind the hydrological data used, as well as reviewing the manuscript.

8.4.4 *Hirpa et al. (2018a)*

<https://centaur.reading.ac.uk/81040/>: This book chapter contributed to the PhD by summarising the recent advances in large-scale flood forecasting with a focus on already existing global and continental flood forecasting systems in operation, such as GloFAS. It presented a review of scientific methodologies used for evaluating and improving the forecast skill, such as evaluation methods, precipitation bias corrections, multi-model approaches and data assimilation. Ervin Zsoter contributed by writing the section

on reducing forecast uncertainty through a multi-model environment.

8.4.5 Alfieri et al. (2018)

<https://centaur.reading.ac.uk/106882/>: This paper contributed to the PhD by summarising the global flood products developed by the institutions participating in the Global Flood Partnership (GFP). It introduces the emerging role of the GFP, a global network of scientists, users, private and public organizations active in global flood risk management, by sharing products in near-real time, developed to predict and monitor where and when floods are taking place. Ervin Zsoter contributed with description of the Extreme Forecast Index and comments on the manuscript.

8.4.6 Zsoter et al. (2016)

<https://centaur.reading.ac.uk/106885/>: This paper presents a study about hydrological uncertainties and skill in reanalysis and ensemble forecast simulations produced using the HTESSSEL and CaMa-Flood models. It can be considered as a preliminary study I have worked on that lead to the start of this PhD. It presents a multi-model approach to producing global flood predictions using multiple reanalysis datasets for river initial conditions and multiple ensemble forcing inputs from the TIGGE data archive. It analysed the sensitivity to the forcing variables and the potential for improving the forecast skill by bias correction. The results highlighted that precipitation is by far the most dominant amongst the forcing variables in producing river discharge. The results also highlight that the three applied reanalysis datasets have different error characteristics that allow for large potential gains with a multi-model combination. It is shown that large improvements to the forecast performance for all models can be achieved through appropriate statistical postprocessing. Ervin Zsoter generated the river discharge simulations, carried out the computations and the analysis and lead the writing of the manuscript.

Chapter 9 Discussion

The aim of this thesis was to increase the understanding of global flood hazard climatologies, to analyse how well they can represent the Earth system and to assess how to improve their relevance for global flood forecasting applications.

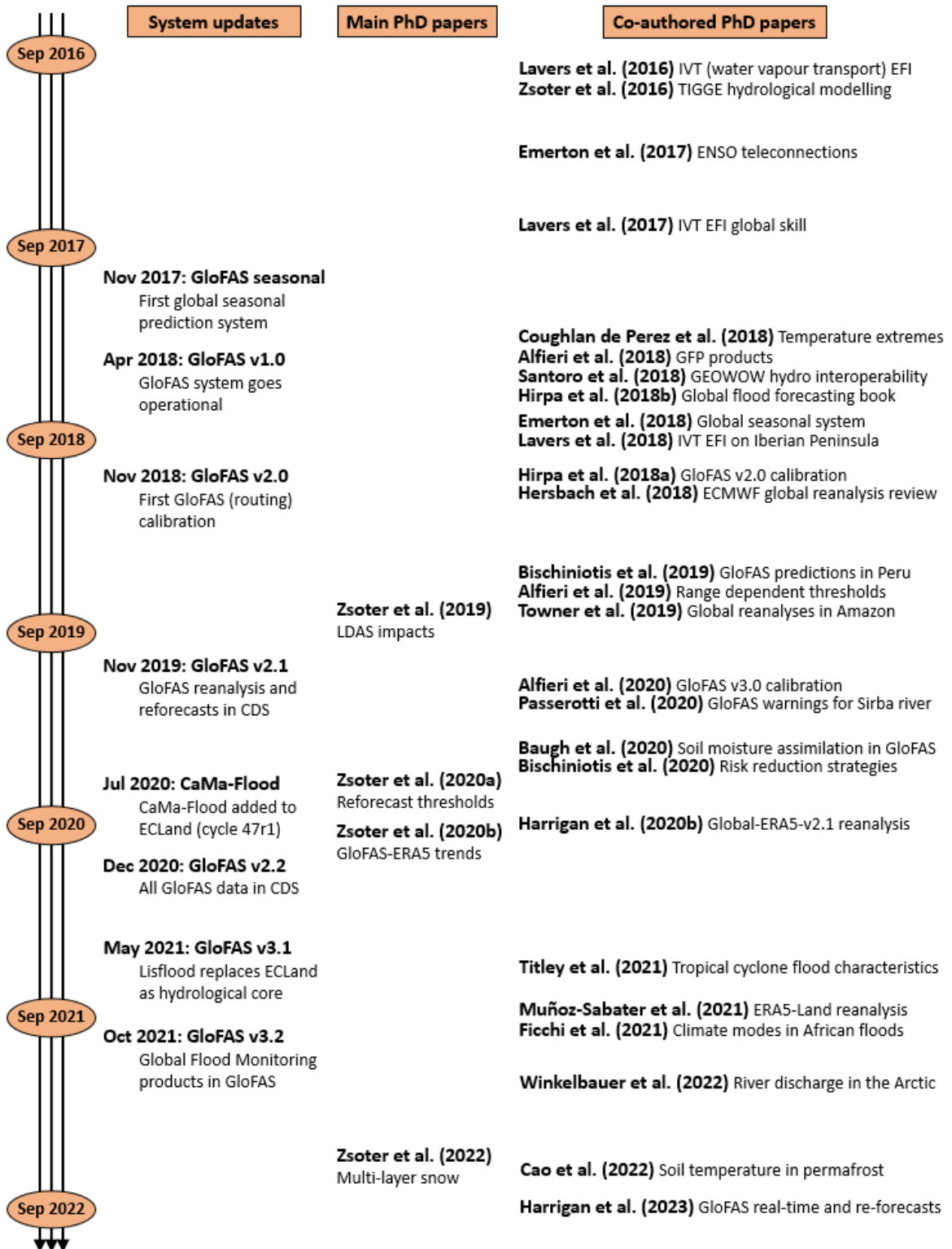


Figure 9-1. Timeline of the major hydrological system updates and scientific publications during the PhD's 6-year period.

This work evaluated some of the crucial characteristics of the reanalysis data sets used to produce the climatologies. Some of the limitations of these input data sets were identified, and suggestions presented to further improve the modelling methodologies. This could improve climatologies and flood thresholds, which are crucial for delivering higher quality flood warnings.

To illustrate the wealth of the applied hydrological systems (to produce the hydrological data sets listed in Chapter 3) and the published papers, either as the four main ones in Chapter 4 to Chapter 7 or as additional co-authored papers (Chapter 8), Figure 9-1 gives a graphical representation of the PhD's timeline during the 6-year period from September 2016 to September 2022.

9.1 Key results and discussion points

The thesis is structured around four main objectives concentrating on specific areas, which resulted in the publication of one scientific paper for each of these areas:

1. Analyse how well the land-surface modelling approach in Earth systems is able to support hydrological applications, in particular focussing on the impact of land-data assimilation of snow and soil moisture on the hydrological cycle in reanalysis simulations.
2. Evaluate the hydrological impact of the complexity of the snow scheme in the land-surface models in reanalysis simulations, with special focus on cold climate areas in permafrost.
3. Evaluate the relevant trends in hydrological reanalyses for river discharge and other related land-surface variables and analyse how the interactions amongst these variables contribute to the trends.
4. Assess the impact of innovative ways of generating flood thresholds on the skill of global flood forecasts, using hydrological ensemble reforecasts.

The first two objectives looked at land-surface modelling aspects of the Earth system models, focussing on the impacts of the land data assimilation and the snow scheme physical complexity on global hydrology. Both were investigated in the context of running ERA5-forced reanalysis simulations, which provided a long enough period of around 40 years to analyse using river discharge observations. These two studies focussed on identifying limitations in the state-of-the-art Earth system modelling capabilities for global flood forecasting.

It was found in Chapter 4 that using land data assimilation in the coupled land-atmosphere reanalysis simulation of ECLand can largely improve the ERA5 weather conditions, and can for example provide better 2-m temperature or snow depth. However, after comparing experiments with and without land data assimilation (LDAS), it could be concluded that the current implementation of LDAS in the ECLand land surface system, tends to produce increments that remove or add water even on an annual average scale. This inevitably leads to systematic water

budget errors that can subsequently contribute to significant errors in the river discharge during times of peak flow downstream, which is a crucial aspect of any flood prediction. The sign of the average increments in the LDAS experiments appears to be negative and seems to be coming mainly from the snow-dominated areas, where the assimilation removes snow. In soil moisture-dominated areas the river discharge seems to be less impacted by the increments, as the evaporation rate plays a more important role.

The snow assimilation impact on the water content is not specific to the ECLand land surface modelling system. Other studies also highlighted significant negative impacts of the snow assimilation on the water balance, for example as described in De Lannoy et al. (2012) on a small catchment in Colorado, United States, or in Arsenault et al. (2013) in two areas in Colorado and Washington states in the United States.

The impact of LDAS on the water budget is an important aspect of the Earth system models, as it can cause potential issues for flood forecasting systems, in case the climatologies (i.e. the flood thresholds) behave differently to the real time forecasts. Prime examples of this problem are the early GloFAS systems (v0, v1.0 and v2.0), which had used ECLand as their hydrological modelling core before v3.1 (with Lisflood) was introduced in May 2021 (Alfieri et al., 2020).

These GloFAS versions used climatologies generated from offline, surface-only reanalysis data sets, which by design did not include the impact of LDAS (data sets such as: GloFAS-ERA-Land-v0, GloFAS-ERA-Land-v1.0, GloFAS-ERA5-Land-v2.0 and GloFAS-ERA-Land-v2.0; see Table 3-1). However, the real-time GloFAS forecasts of these versions and the reforecasts in v0 and v1.0, did include the impact of LDAS, as they used the ensemble runoff forecasts from ENS, that are initialised from the operational 4D-Var analysis of ECMWF, which assimilates millions of observations (Wedi et al., 2015).

This thesis demonstrated that the inclusion of LDAS can create large differences in river discharge. For these early GloFAS systems, this created large inconsistencies within the system, with the climatologies (i.e. the flood thresholds) behaving differently to the real-time forecasts, resulting in a different representation of the flood event frequencies. Model developers should be aware of this kind of possible inconsistencies, originating from the use of LDAS, and should analyse their systems for the severity of the potential impacts. Addressing this potential inconsistency in the Earth-system-model-based flood forecasting systems is of great importance.

It is well-known that data assimilation is a very important component of any Earth system model, because it corrects random (day-to-day) errors by making optimal use of a large set of Earth observations. However, data assimilation systems are not designed to account for systematic

biases. The fact that LDAS produces consistently negative increments in snow-covered areas in this study points towards an apparent snow model bias. In contrast, a model affected by random errors only, would lead to data assimilation increments of both signs with close to zero annual mean values.

As Dutra et al. (2012) highlighted, the single-layer snow scheme in ECLand (still operationally used in 2022) melts the snow too slowly. This makes the LDAS snow increments compensate for the systematic snow depth overestimation bias, shown in Chapter 4. The use of physically more complex multi-layer snow schemes is one of the suggested ideas that could provide potential improvements.

Chapter 5 explored this idea, by evaluating the hydrological impact of the new ECMWF multi-layer snow scheme in reanalysis simulations. The physically more complex snow scheme was shown to be able to provide an improved representation of the hydrological processes, demonstrated mainly through reduced bias and variability errors of river discharge. However, the performance was proved to be suboptimal in large parts of the high latitude permafrost regions, where the current operationally used single-layer snow scheme of ECLand (that was also used in the LDAS study in Chapter 4) is superior, producing smaller river discharge errors.

To improve the multi-layer scheme's hydrological representation in permafrost, modifications of the ECLand snow and soil freezing parametrisations were evaluated. It was shown that a series of incremental changes could noticeably improve the quality of the river discharge simulation over a large area in permafrost, primarily through decreasing the soil temperature and thus increasing the amount of surface runoff in the critical spring snowmelt period.

Even though it was not possible to carry out a full-blown sensitivity study, which would have required testing hundreds of parameter combinations, it already largely contributed to the understanding of the hydrological importance of each of the modified aspects of the ECLand parametrization, by ranking these parametrisation changes from least contributing to most contributing in hydrological improvements, based on the verification results for permafrost.

In addition, this work had a direct impact on the development of the snow model applied in the ECMWF IFS. Based on this study, the slightly modified version of ML-Meta1 was selected for operational implementation in the next IFS cycle of 48r1, expected sometime during 2023. This version of the snow model will include the snow vertical discretization and the first snow metamorphism changes, which has been tested for permafrost, on top of the default ML configuration.

By reporting on the successful impact of the ECLand model's snow component on the operational development, this study demonstrated that hydrological diagnostic studies, such as the work presented in Chapter 4 and Chapter 5, have great potential to help improve the land-surface realism in Earth system models and contribute to improvements of the whole Earth system, not just the hydrological variables, such as river discharge.

After analysing the hydrological contributions of the crucial aspects of LDAS and snow modelling in Earth system models in the context of reanalysis simulations, the two remaining objectives of the thesis considered the applicability of these hydrological reanalysis simulations to the generation of flood hazard climatologies with related flood warning thresholds.

These objectives focussed on identifying limitations in the current use of flood climatologies and thresholds, either with the presence of non-stationarity in the reanalysis time series or the suboptimal use of the reanalysis-based thresholds in flood forecasting. The findings for these two objectives are likely to be transferable to forecasting systems based on global hydrological models and are not expected to be specific to the earth-system modelling approach, evaluated in this PhD.

It was found in Chapter 6 that non-stationarity (measured by linear trends) is widespread in the global v2.1 GloFAS-ERA5 reanalysis time series. Moreover, the river discharge trends are closely linked to changes and discontinuities in the underlying land surface variables, most notably precipitation and snowmelt. The linear trends in GloFAS-ERA5 seem to be driven by changes in precipitation over tropical and subtropical areas of the world, while snowmelt changes show a very strong influence in determining the river discharge trends in northern latitudes. The snowmelt behaviour in northern latitudes is likely to be related to changes in the snowmelt processes, especially the snow assimilation combined with inhomogeneities in the use of available observations. This is confirmed by more realistic-looking trends in ERA5-Land, which is not impacted by land data assimilation.

The linear trends, analysed in this thesis, highlighted very noticeable issues with the non-stationarity of the ERA5 (and in many aspects also ERA5-Land) land-surface related variables. The majority of the demonstrated, dominantly negative trends in ERA5 could not be confirmed by corresponding observed linear trends. Regardless of the often-shorter period of available observations there have to be some fundamental reasons behind this discrepancy.

Some of the trends, in particular the very negative ones in the higher latitudes, could be attributed to the land data assimilation impact on the snowmelt, which appears to show a large shift related to the use of the IMS satellite-based snow product from 2004. Another potential driver of the ERA5 trends could be the several streams that were run in the ERA5 production and later merged into

one consolidated data set. However, even though this was shown to have produced some discontinuities in the deep soil, where spin-up can take several years (Hersbach et al., 2020), it is not expected to be a noticeable contributor to the large trends, highlighted in this thesis.

The findings of the thesis suggest that there have to be other underlying reasons, possibly significant observation system inhomogeneities (similar to the IMS data), which could impact the land-surface variables and contribute to the large trends in variables like precipitation or snow. The future versions of ECWMF reanalysis data sets, such as ERA6, will undoubtedly benefit from studies like the one presented in this thesis. These hydrology-focussed diagnostic studies can help the research teams by signposting where further developments are needed in order to improve the stationarity aspect of the land-surface variables in the reanalyses.

The presence of large trends can have a very negative impact on flood forecasting systems that use reanalysis data sets to generate their flood thresholds, such as GloFAS. The thesis highlighted this for an example catchment in the Congo basin in Africa, which suffers a very large drop in river discharge from about year 2000. When flood thresholds are computed based on a time series with a large decrease in river discharge, the system does not produce flood warnings at all. This is actually an existing problem in the operational GloFAS version, which is visible in the missing flood reporting points in that area for all forecast dates on the GloFAS forecast website (www.globalfoods.eu). Based on the analysis presented in this thesis, it is recommended that flood forecasting system developers should carefully analyse the climatologies for the occurrence of large trends, such as shown here for the state-of-the-art ERA5 reanalysis and the related GloFAS-ERA5 river discharge reanalysis.

As the above example suggests, the way flood thresholds are generated, is expected to have a significant impact on the behaviour of the flood forecasting systems and on the related reliability and skill of the system. Moreover, the usability of flood thresholds and the related forecast performance are not only impacted by large trends in the river discharge reanalysis, but also by biases in the forecast simulations that evolve with the increasing forecast range. This can also be an issue, as different biases across lead times are likely to lead to different occurrence frequencies of extreme events (i.e. floods), which cannot any more be represented by the same set of flood thresholds at all those different lead times.

To address this issue, a methodology of deriving range-dependent flood thresholds from ensemble reforecasts was developed in this thesis in Chapter 7. After different annual maxima sampling methods were analysed to generate flood thresholds, the ensemble-forecast-based thresholds could be shown to deliver much improved forecast reliability and skill, compared with the use of

reanalysis-based nonvarying thresholds, that are usually applied operationally in most forecasting systems. Overall, forecast errors could be reduced by up to 50-90% when using the ensemble-forecast-based flood thresholds, compared with the traditional reanalysis-based thresholds, depending on the flood severity and lead time. Results also showed that using the single unperturbed control member to define the thresholds is not sufficient and exploring the full ensemble of the reforecasts in the threshold derivation can often further increase the forecast performance.

The thesis highlighted, that using the median of hundreds of threshold sets, generated by taking one randomly selected ensemble member-from each reforecast, can deliver the best overall results. However, this is not always the case, as for lower flood severity levels there were slightly better performing options. In addition, it has to be acknowledged that the methodology presented in this study, will become relatively expensive when performed not only for a selection of few thousand catchments, but for the whole GloFAS river network (with millions of river pixels). Therefore, in case the computational resources are limited, the use of the single ensemble control member, which does not require any randomised threshold generation, could provide a good compromise with high overall performance.

Similarly, it was argued in Chapter 7, that the post-processing of the forecasts against the reanalysis data set (e.g. by quantile mapping), could be an alternative method to account for the range-dependent biases in the system. However, this will naturally alter the forecast values, which might not be preferable to some users, regardless of how well this method handles skill and reliability.

Therefore, it is recommended that every hydrological forecasting system should be evaluated for the underlying biases, following the method described in this thesis. This will highlight possible inconsistencies in the system between the flood hazard climatology and the forecasts. Using ensemble reforecasts could provide the best platform to do this, but even if they are not available, the consistency between the climatology and the historical forecasts (even if produced with different model versions), should be diagnosed.

This thesis demonstrated that using flood thresholds in flood forecasting systems, which do not represent the behaviour of the real-time forecasts well for all lead times (for example in GloFAS), will inevitably decrease the skill and reliability of the flood warnings and jeopardise the confidence of the users in the system. These systems can greatly benefit from using ensemble-forecast-based thresholds instead, as a practical and effective way to resolve inconsistencies between forecasts and flood thresholds, thus increasing the flood forecast skill.

9.2 Next steps

Each of the four main papers presented in this thesis contributed to improving the knowledge about flood hazard climatologies. While significant progress was made in exploring possibilities for improved quality as well as using these climatologies in global flood predictions, the work also raised several questions and provided motivation for further research. While each of the Chapter 4 to Chapter 7 highlighted some aspects for further studies, this section identifies some key areas for future work with examples where work could be extended and built upon.

Chapter 4 highlighted the hydrological impact of a crucial land surface modelling aspect, demonstrating that LDAS can cause significant water budget issues and introduce biases in flood predictions. When the study was conducted, there was no feasible option to run the land–atmosphere coupling and LDAS separately in the ECMWF NWP system. Either both were active as in ONLINE (in the coupled modelling system with LDAS), or neither of them as in OFFLINE (uncoupled land-surface only simulation without LDAS, using meteorological forcing). Separating these two contributing modelling options by running and checking reanalysis simulations with only one of them being active would provide an interesting insight into the land–atmosphere coupling’s contribution in the future, in addition to the impact of LDAS.

Even though it was demonstrated that LDAS can cause issues in downstream hydrological applications by opening the water budget, the use of data assimilation in hydrological simulations is undoubtedly crucial. The inclusion of LDAS in reanalysis simulations is expected to be made possible by the stand-alone surface analysis (SSA) configuration, which is a recent, ongoing area of developments at ECMWF. The SSA skips the very expensive upper atmosphere 4D-Var part of the assimilation by replacing it with the upper-air analyses from another existing experiment, while the land data analysis is still performed in the standard way. This offers the possibility of carrying out offline reanalysis runs, similar to the experiments evaluated in Chapter 4 and Chapter 5, but this time also including land data assimilation. In the future, the applicability of the SSA to generate surface-improved hydrological reanalysis data sets should be evaluated. This new methodology promises the generation of improved, forecast-consistent climatologies, which include not just hydrological modelling improvements and externally enhanced meteorological forcing, similar to ERA5-Land, but also the crucial land data assimilation.

An area where improvements are needed in the future, is to directly address the LDAS’s limitation to open the water budget by the potentially systematic increments. For this, different approaches (e.g. Zaitchik and Rodell, 2009 or Pan and Wood, 2006) could be explored in ECLand that include special handling of the snow and soil moisture increments, in order to retain the water in the hydrological cycle during the data assimilation cycle. As a first step, this could be investigated in

uncoupled systems, where the closure of the water balance could be easier to achieve when assimilating hydrological observations in an offline environment. In the long term, however, a fully coupled land–atmosphere NWP system’s hydrological performance is to be explored, with a coupled land–hydrology data assimilation at its core. This would provide a true Earth system approach, where the joint assimilation of land surface and river discharge observations would consistently correct the different components of the Earth system.

Another area where the Earth system modelling approach could be improved at ECMWF, is the handling of the parametrisation in ECLand, as highlighted in Chapter 5, with the evaluation of the new multi-layer snow scheme. The results demonstrated that the uniform parameters used currently in ECLand are too simplistic and cannot work efficiently for both permafrost and non-permafrost areas in snow-impacted climate. It was suggested that spatially variable parametrization of hydrologically sensitive variables in permafrost (such as the snow and soil freezing parameters explored in Chapter 5), could bring a more balanced approach and would deliver an improved hydrological process representation. This idea should be tested in the future.

The first building blocks of this work have already been implemented at ECMWF, with the refactorization of ECLand (by removing hard-coded parameter values) and the implementation of Multiscale Parameter Regionalization (MPR; Scheppe et al., 2021) for estimation of spatially varying parameters. The ambitious attempt to calibrate the parameters in ECLand for an improved land–atmosphere–coupled hydrological performance is going to be one of the most important future developments in the Earth system modelling.

The results of this thesis proved that further development of the snow and soil parametrizations in ECLand are necessary in order to achieve a better hydrological performance, not just in permafrost, but everywhere globally.

Future modelling work, on the one hand, should focus on improving the hydrological response of the existing snow and soil parametrisation areas, for example by analysing the sensitivity to the snow cover fraction parametrisation by exploring more complex relationships between snow depth and snow cover. On the other hand, there are still some poorly represented or even non-represented hydrological processes in ECLand, which promise to bring potential hydrological improvements in snow-dominated areas. Areas that could be considered include a deeper soil with additional vertical layers, wind-driven snow sublimation, snow interception by forests or a more physically complex phase representation of the water in the soil.

Another area that could have a potential impact on the river-discharge-reanalysis-derived flood climatologies is the presence of striking and dominantly negative trends, as highlighted in Chapter

6, by evaluating the state-of-the-art GloFAS-ERA5-v2.1 reanalysis. Although the presented analysis managed to provide a preliminary exploration of the underlying causes for these pronounced linear trends, further investigations of the reasons could be crucial for future reanalysis developments at ECMWF. This should include further analysis of alternative meteorological reanalyses data sets, such as MERRA2 or JRA55, and other gauge- and/or satellite-based climate data sets, such as GPCP, GPCC, TRMM 3B42, CMORPH, MSWEP etc. These could be analysed for similarities and differences in the trend patterns vs trends in ERA5/ERA5-Land variables, which could highlight further explanations behind the demonstrated ERA5 trends presented in Chapter 6. This is especially important for areas such as the central region of Africa and the southwestern part of the United States, where large negative trends are present across both precipitation and river discharge and also ERA5 and ERA5-Land, pointing to a potential observation system inhomogeneity issue.

It would be beneficial to compare the newer version of the GloFAS-ERA5-v3.1 river discharge reanalysis with GloFAS-ERA5-v2.1 (the one analysed in Chapter 6). Winkelbauer et al. (2022) already looked at the trends in GloFAS-ERA5-v3.1 (including other reanalysis products), specifically in the Arctic regions, and pointed out that v3.1 shows a marked improvement over v2.1 for this area of the world. As GloFAS-ERA5-v3.1 uses Lisflood for the hydrological modelling core instead of ECLand, the globally extended v3.1 vs v2.1 comparison could highlight differences caused by the hydrological modelling and not by the meteorological forcing. For example, land data assimilation will impact v2.1, as ERA5 runoff is used in there directly, but will not impact v3.1, as runoff is computed by Lisflood from the ERA5 meteorological input.

Moreover, it is expected that large trends in the river discharge reanalysis will potentially cause problems when flood thresholds are derived from these non-stationary time series, as the trends make them less representative of the forecast behaviour, ultimately causing unreliable warnings when related forecast probabilities are used. Future versions of the flood climatologies could build upon studies that analyse the extent of the trend impact on the actual forecast reliability and skill. This should preferably be done in a combined evaluation with the ERA5 back-extension to 1959, by rechecking the non-stationarities in the longer period, and the alternative threshold generation methodologies, such as suggested in Chapter 7, using ensemble reforecasts over a shorter period (i.e. 20 years), simply because shorter periods could naturally be less prone to the presence of impactful trends.

The underlying cause for the drop of reliability and skill, when using non-lead-time-dependent flood thresholds for longer lead times, as presented in Chapter 7, is related to the non-stationary, evolving biases across lead times in the forecasts. As different hydrological forecast systems are expected to show different bias structures, applying the methodology of this thesis to other

systems could lead to different results. It would be beneficial to test this in the future by repeating the study with the most recent modelling version of GloFAS (currently GloFAS-ERA5-v3.1 reanalysis and GloFAS-ERA5-RFC-v3.1 reforecasts). The behaviour of the biases and the impact on forecast skill and reliability could well be different, for example in the case of v3.1 vs v2.1, as it could possibly originate from the contrasting hydrological modelling cores of Lisflood vs ECLand/Lisflood-routing.

It was also discussed in Chapter 7 that the post-processing of the river discharge forecasts against the reference dataset (in our case the reanalysis) could be an alternative route to using range-dependent flood thresholds based on ensemble reforecasts. This could be explored in a future study by applying e.g. a simple linear regression or quantile mapping to the forecasts. The use of ensemble reforecasts, presented in this thesis, and the alternative forecast post-processing could be compared and the pros and cons evaluated for both methods.

Moreover, future studies should also focus on investigating the possibility of an operational implementation in the GloFAS system of the method recommended in Chapter 7, or potentially the alternative forecast bias correction that is to be explored in the future. The implementation procedure should also include some careful consideration of the users' needs by collecting their suggestions and feedbacks. The potential benefits and drawbacks of changing the operational system's one-value-for-all-lead-times thresholds should be evaluated and the most beneficial method, which can most effectively resolve inconsistencies between forecasts and flood thresholds and deliver increased flood forecast reliability and skill, should be identified and carefully implemented.

9.3 Limitations of the thesis

The findings in this thesis are specific to the Earth system model used at ECMWF, with the land-surface scheme of ECLand and the processes associated with this scheme, which provided the hydrological modelling core for all the main studies presented in Chapter 4 to Chapter 7. However, different hydrological modelling systems that include LDAS, should all be susceptible to water budget issues through the increments, regardless of the models and processes included, even though the impacts are obviously not expected to be entirely the same.

Similarly, the hydrological impact of the snow model complexity will be expected to be specific to the hydrological modelling and the various land-surface process interactions in ECLand, but the individual analysed processes, such as the snow vertical discretisation or the change in the snow metamorphism, should represent a more general behaviour independent of the model system.

In addition, the choice of modelling system could also be a potential limitation for the threshold impact study, presented in Chapter 7. However, the methodology was designed in a way to ensure

that the results are expected to be generally independent of factors like the choice of model, the extreme value fitting method or the sampling period length.

In Chapter 4, the limitations also include the missing option of separating the impact of land–atmosphere coupling and LDAS in the ECMWF Earth system model. Either both were active as in ONLINE, or neither of them as in OFFLINE. Future studies could potentially evaluate, whether the land–atmosphere coupling noticeably contributes to the impact of LDAS as the dominant source of differences between ONLINE and OFFLINE (as argued in Chapter 4).

In general, the LSMs' ability to support hydrological simulations can be limited by inadequate handling of numerous underlying processes, causing potential problems downstream in the hydrological applications. Even though this thesis could demonstrate this by studying the impact of two important processes, the LDAS and the snow model complexity in Chapter 4 and Chapter 5, it has to be acknowledged that there are many other areas with expected significant hydrological contributions, such as evaporation, runoff generation or soil moisture processes, which could not be part of this thesis.

Similarly, the study of the hydrological impact of the snow model complexity in Chapter 5 could not be designed to be a full sensitivity experiment for the land-surface processes in permafrost. That would have required a very large computational cost, so instead some of the most important processes, such as the vertical discretisation or the snow-soil thermal conductivity, were investigated.

It also has to be acknowledged that even with the best care taken in the experimental setup of the studies in Chapter 4, Chapter 5 and Chapter 6, the network of analysed catchments with river discharge observations is still under-representative. This is particularly true in tropical areas, which are especially poorly observed with large gaps in space and time in the reanalysis period of 1980 to 2018. Moreover, the 8 and 9 minimum number of years in the hydrological evaluations (in Chapter 4 and Chapter 5) and the 16 years in the linear trend analysis (Chapter 6) are relatively short minimum observation periods, but due to the limited availability of observations especially in more recent years, this was accepted as a compromise.

Regarding the trend analysis, it can be acknowledged that more complex trend analysis methods could have been used in Chapter 6 instead of the simple linear method, many of which could have also provided information on the significance level. However, the study in this thesis was only intended to deliver a general first picture about any non-stationarity in the reanalysis time series. Moreover, it is also relevant that large discontinuities, such as the one highlighted in relation to the introduction of the IMS satellite-based snow cover product in 2004, can make the linear trend

analysis somewhat unreliable. However, it could be argued, that the evolution of snow (and probably of other relevant variables as well) in ERA5 and ERA5-Land (see Figure 6-5) appears to be more complex than a single discontinuity in 2004, which suggests the linear trend analysis still delivers very valuable information even on the whole reanalysis period of 1981- 2018.

Chapter 10 Conclusions

Earth system models have largely improved in physical complexity and land-surface process representation in recent decades. This has helped them in reaching the level required to support the development of global hydrological flood forecasting applications, with GloFAS as one of the prime examples. The flood hazard climatologies and their flood thresholds, usually generated from reanalysis, are key elements of these systems. The quality of these climatologies is crucially important, as it has a strong influence on the value of the derived flood warnings.

The main aim of this thesis was to explore some of the crucial characteristics of the data sets used to produce these flood hazard climatologies and to identify areas of limitations, where the quality of the applied modelling or input reanalysis data sets could be improved. In addition, the thesis also explored aspects of the generation and use of the flood hazard climatologies. The presence of non-stationarity in the reanalysis time series and the suboptimal use of the reanalysis-derived flood thresholds in flood forecasting were both investigated.

It was found that even though the land data assimilation can largely improve surface weather conditions, it tends to produce systematic increments even on an annual average scale, which open the water budget and potentially cause large river discharge errors in downstream applications. In soil-moisture-dominated areas, the river discharge seems to be less impacted by the increments, but over snow-dominated areas, the assimilation deteriorates the river discharge quality by removing snow in response to an apparent snow model bias.

The new multi-layer ECMWF snow scheme was evaluated as a potential solution for the snow model bias problem. The physically more complex scheme was shown to provide an improved hydrological process representation over large parts of snow-dominated areas. However, the performance was suboptimal in permafrost regions. This was addressed by modifications in the ECLand snow and soil freezing parametrisations. It could be shown that a series of incremental changes could noticeably improve river discharge in permafrost, primarily through decreasing the soil temperature and increasing the surface runoff in the critical spring snowmelt period.

The thesis could also show that non-stationarity, measured by linear trends, is widespread in the analysed GloFAS-ERA5 hydrological reanalysis time series. The trends seem to be driven by changes in precipitation over the tropical and subtropical areas of the world, while snowmelt has a very strong influence in determining the river discharge trends in the northern latitudes.

Although large trends are likely to hinder the usability of the derived flood thresholds, non-stationary forecast biases will also be expected to reduce reliability and skill, as the same reanalysis-

based threshold will likely become less and less representative of the forecasted flood events with the increasing forecast range. To address this, a new methodology of range-dependent flood thresholds, generated from ensemble reforecasts, was developed. These ensemble-reforecast-based thresholds were shown to deliver much improved forecast reliability and skill. Overall, forecast errors could be reduced by up to 50-90%, compared with using the traditional reanalysis-based thresholds. Results also showed that using all ensemble members in the threshold generation could bring further advantages over the use of the single unperturbed control member.

The four main papers, addressing the objectives outlined in Chapter 1, and many of the additional co-authored papers (see Chapter 8) summarised in this thesis, all contributed to improving the knowledge about flood hazard climatologies. The topics covered in the scientific studies presented in this thesis are generally relevant for any hydrological forecasting systems, not just the ones that have an Earth system model with a land-surface scheme as their hydrological core, such as GloFAS before its v3.1 upgrade in May 2021.

This research highlighted hydrologically relevant limitations in the reanalysis data sets, used to generate climatologies, and suggested potential avenues for improving those areas. In the case of the multi-layer snow model's impact in permafrost, the recommended ECLand parametrisation changes were also analysed. This additional sensitivity analysis led to a change of the multi-layer snow scheme's operational candidate, to be implemented in the next IFS cycle (48r1) in 2023, which is specifically due to the highlighted hydrological improvements in permafrost.

This thesis demonstrated that diagnostic studies, such as the work presented in Chapter 4 and Chapter 5 about the hydrological impact of LDAS and the snow scheme complexity in land-surface schemes, can have great success in helping the improvement of land-surface realism in Earth system models and in contributing to improvements of the whole Earth system and not just of the hydrological variables, such as river discharge.

This thesis also demonstrated that using flood thresholds in flood forecasting systems that do not represent the extreme event behaviour of the real-time forecasts well for all lead times (for example in GloFAS) will inevitably decrease the skill and reliability of the flood warnings and jeopardise the user confidence in the system. This inconsistency between the extreme event representation of the climatology and the forecasts can be caused by sources such as large trends in the underlying reanalysis time series or the nonstationary biases in the forecasts, as shown in Chapter 6 and Chapter 7, respectively.

The future versions of reanalysis data sets, such as ERA6 of ECMWF, will undoubtedly benefit from studies like the ones presented in this thesis. These can help research teams by signposting which

areas further development activities should concentrate on, in order to improve the quality of the land-surface variables in the reanalysis data sets.

The findings of this research have implications for both the land-surface modelling and global flood forecasting communities. As a take-home message, the thesis contributed practical recommendations for system developers, in order to reduce the risk of unnecessary skill loss. Firstly, they should look out for a potentially inconsistent use of LDAS in different parts of the system (i.e. climatology and real-time forecasts), like the example of GloFAS before May 2021. Similarly, climatologies should be analysed for the occurrence of large trends, such as shown here for the state-of-the-art ERA5 and the related GloFAS-ERA5 river discharge reanalyses. In addition, potential inconsistencies should also be checked in the handling of extreme events in different parts of the systems. This should be done preferably using the methodology presented in Chapter 7, with the help of reforecasts, or when available, ensemble reforecasts, but even if no reforecasts are available, at least the largest possible collection of historical forecasts should be used.

The presence of these potential issues affecting the flood hazard climatologies and the related flood warnings, explored in this thesis, should be always investigated. Without addressing these latent issues, the reliability and skill of the flood event forecasts and thus the quality of the flood warnings could be substantially reduced. This should be avoided, as it can strongly influence the decision-making process and ultimately lead to a loss of confidence in the products of the global flood forecasting systems.

The work presented in this thesis has contributed to better understanding and refined use of flood hazard climatologies. It has also provided practical recommendations to the hydrometeorological community for improved global flood forecasting that can bring better preparedness across the world and ultimately can help saving lives and protect livelihoods.

References

- Abramowitz, G., R. Leuning, M. Clark, and A. Pitman, 2008: Evaluating the Performance of Land Surface Models, *J. Clim.*, 21, 5468–5481, <https://doi.org/10.1175/2008JCLI2378.1>
- Adam, J. C., A. F. Hamlet, and D. P. Lettenmaier, 2009: Implications of global climate change for snowmelt hydrology in the twenty-first century, *Hydrological Processes*, 23(7), 962–972. <https://doi.org/10.1002/hyp.7201>
- Albergel, C., P. de Rosnay, G. Balsamo, L. Isaksen, and J. Muñoz Sabater, 2012: Soil moisture analyses at ECMWF: Evaluation using global ground-based in situ observations, *J. Hydrometeor.*, 13, 1442–1460, <https://doi.org/10.1175/JHM-D-11-0107.1>
- Albergel C., and Coauthors, 2013: Skill and Global Trend Analysis of Soil Moisture from Reanalyses and Microwave Remote Sensing, *J. Hydrometeor.*, 14, 1259–1277, <https://doi.org/10.1175/JHM-D-12-0161.1>
- 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, *Hydrology and Earth System Sciences*, 17, 1161–1175, <https://doi.org/10.5194/hess-17-1161-2013>
- Alfieri L., M. Berenguer, V. Knechtel, K. Liechti, D. Sempere-Torres, and M. Zappa, 2015: Flash Flood Forecasting Based on Rainfall Thresholds, in *Handbook of Hydrometeorological Ensemble Forecasting*, edited by: Q. Duan, F. Pappenberger, J. Thielen, A. Wood, H. Cloke, J. Schaake, 1–38, Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-40457-3_49-1
- Alfieri L., S. Cohen, J. Galantowicz, G. Schumann, M. Trigg, E. Zsoter, C. Prudhomme, A. Kruczkiwicz, E. Coughlan de Perez, Z. Flamig, R. Rudari, H. Wu, R. Adler, R. Brakenridge, A. Kettner, A. Weerts, P. Matgen, S. Islam, T. de Groeve, P. Salamon, 2018: A global network for operational flood risk reduction, *Environmental Science & Policy*, 84, 149–158, <https://doi.org/10.1016/j.envsci.2018.03.01.014>
- Alfieri, L., E. Zsoter, S. Harrigan, F. Hirpa, C. Lavaysse, C. Prudhomme, and P. Salamon, 2019: Range-dependent thresholds for global flood early warning, *Journal of Hydrology*, <https://doi.org/10.1016/j.hydroa.2019.100034>.
- Alfieri, L., V. Lorini, F. A. Hirpa, S. Harrigan, E. Zsoter, C. Prudhomme, and P. Salamon, 2020: A global streamflow reanalysis for 1980–2018, *J. Hydrol.*, X 6, 100049. <https://doi.org/10.1016/j.hydroa.2019.100049>.

- Anabalón A., and A. Sharma, 2017: On the divergence of potential and actual evapotranspiration trends: An assessment across alternate global datasets, *Earth's Future*, 5, 905–917, <https://doi.org/10.1002/2016EF000499>
- Anderson, E. A., 1976: A Point Energy and Mass Balance Model of a Snow Cover, Stanford University: Stanford, CA, USA, 150p
- Andreadis, K. M., and D. P. Lettenmaier, 2006a: Assimilating remotely sensed snow observations into a macroscale hydrology model. *Adv. Water Resour.*, 29, 872–886, <https://doi.org/10.1016/j.advwatres.2005.08.004>.
- Andreadis, K.M., and D. P. Lettenmaier, 2006b: Trends in 20th century drought over the continental United States. *Geophys. Res. Lett.*, 33, L10403, <https://doi.org/10.1029/2006GL025711>.
- Andresen, C. G., D. M. Lawrence, C. J. Wilson, A. D. McGuire, C. Koven, K. Schaefer, E. Jafarov, S. Peng, X. Chen, I. Gouttevin, I. Burke, S. Chadburn, D. Ji, G. Chen, D. Hayes, and W. Zhang, 2020: Soil moisture and hydrology projections of the permafrost region – A model intercomparison, *Cryosphere*, 14, 445–459, <https://doi.org/10.5194/tc-14-445-2020>
- Archfield, S. A., M. Clark, B. Arheimer, L. E. Hay, H. McMillan, J. E. Kiang, J. Seibert, K. Hakala, A. Bock, S. T. Wagener, W. H. Farmer, V. Andréassian, S. Attinger, A. Viglione, R. Knight, S. Markstrom, and T. Over, 2015: Accelerating advances in continental domain hydrologic modeling, *Water Resour. Res.*, 51(12), 10078–10091, <https://doi.org/10.1002/2015WR017498>
- Archfield, S. A., R. M. Hirsch, A. Viglione, and G. Blöschl, 2016: Fragmented patterns of flood change across the United States, *Geophys. Res. Lett.*, 43, 10–232, <https://doi.org/10.1029/2015GL065797>
- Arduini, G., G. Balsamo, E. Dutra, J. Day, I. Sandu, S. Boussetta, and T. Haiden, 2019: Impact of a multi-layer snow scheme on near-surface weather forecasts, *J. Adv. Model. Earth Syst.*, 11, 4687–4710, <https://doi.org/10.1029/2019MS001725>
- Arheimer, B., R. Pimentel, K. Isberg, L. Crochemore, J. C. M. Andersson, A. Hasan, and L. Pineda, 2020: Global catchment modelling using World-Wide HYPE (WWH), open data, and stepwise parameter estimation, *Hydrol. Earth Syst. Sci.*, 24, 535–559, <https://doi.org/10.5194/hess-24-535-2020>
- Armstrong, R. L., and E. Brun, 2008: *Snow and Climate: Physical Processes, Surface Energy Exchange and Modeling*. Cambridge University Press, 222 pp

- Arnone, E., M. Cucchi, S. D. Gesso, M. Petitta, and S. Calmanti, 2020: Droughts Prediction: a Methodology Based on Climate Seasonal Forecasts, *Water Resour. Manage.*, 34, 4313–4328, <https://doi.org/10.1007/s11269-020-02623-3>
- Arsenault, K. R., P. R. Houser, G. J. M. De Lannoy, and P. A. Dirmeyer, 2013: Impacts of snow cover fraction data assimilation on modeled energy and moisture budgets, *J. Geophys. Res. Atmos.*, 118, 7489–7504, <https://doi.org/10.1002/jgrd.50542>
- Agustí-Panareda, A., G. Balsamo, and A. Beljaars, 2010: Impact of improved soil moisture on the ECMWF precipitation forecast in West Africa, *Geophys. Res. Lett.*, 37, L20808, <https://doi.org/10.1029/2010GL044748>
- 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, <http://dx.doi.org/10.1175/2008JHM1068.1>
- Balsamo, G., S. Boussetta, P. Lopez, and L. Ferranti, 2010: Evaluation of ERA-Interim and ERA-Interim-GPCP-rescaled precipitation over the U.S.A., ERA Rep. Series 5, ECMWF, 10 pp., <https://www.ecmwf.int/node/7926>
- Balsamo, G., F. Pappenberger, E. Dutra, P. Viterbo, 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. Proc.*, 25, 1046–1054, <https://doi.org/10.1002/hyp.7808>
- Balsamo, G., and Coauthors, 2015: ERA-Interim/Land: A global land surface reanalysis data set, *Hydrol. Earth Syst. Sci.*, 19, 389–407, <https://doi.org/10.5194/hess-19-389-2015>
- Baugh, C.; P. de Rosnay, H. Lawrence, T. Jurlina, M. Drusch, E. Zsoter, and C. Prudhomme, 2020: The Impact of SMOS Soil Moisture Data Assimilation within the Operational Global Flood Awareness System (GloFAS), *Remote Sens.*, 12, 1490. <https://doi.org/10.3390/rs12091490>
- Beck H. E., E. F. Wood, M. Pan, and C. K. Fisher, 2019: MSWEP V2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment. *Bull. Amer. Meteor. Soc.*, 100, 473–500, <https://doi.org/10.1175/BAMS-D-17-0138.1>
- Bélair, S., L.-P. Crevier, J. Mailhot, B. Bilodeau, and Y. Delage, 2003a: Operational implementation of the ISBA land surface scheme in the Canadian Regional Weather Forecast Model. Part I: Warm season results. *J. Hydrometeor.*, 4, 352–370, [https://doi.org/10.1175/1525-7541\(2003\)4<352:OIOTIL>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)4<352:OIOTIL>2.0.CO;2).

Bélaïr, S.; R. Brown, J. Mailhot, B. Bilodeau, and L. P. Crevier, 2003b: Operational implementation of the ISBA land surface scheme in the Canadian regional weather forecast model. Part II: Cold season results. *J. Hydrometeorol.*, 4, 371–386. [https://doi.org/10.1175/1525-7541\(2003\)4<371:OIOTIL>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)4<371:OIOTIL>2.0.CO;2).

Beljaars, A. C. M., P. Viterbo, M. Miller, and A. K. Betts, 1996: The anomalous rainfall over the United States during July 1993: Sensitivity to land surface parameterization and soil anomalies. *Mon. Wea. Rev.*, 124, 362–383, [https://doi.org/10.1175/1520-0493\(1996\)124<0362:TAROTU>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<0362:TAROTU>2.0.CO;2).

Best, M. J., and Coauthors, 2011: The Joint UK Land Environment Simulator (JULES), Model description—Part 1: Energy and water fluxes. *Geosci. Model Dev.*, 4, 677–699, <https://doi.org/10.5194/gmd-4-677-2011>.

Betts, A. K., R. Desjardins, D. Worth, S Wang, and J. Li, 2014: Coupling of winter climate transitions to snow and clouds over the Prairies. *Journal of Geophysical Research: Atmospheres*, 119, 1118–1139, <https://doi.org/10.1002/2013JD021168>.

Biancamaria, S., D. P. Lettenmaier, and T. M. Pavelsky, 2016: The SWOT Mission and Its Capabilities for Land Hydrology. *Surv. Geophys.*, 37, 307–337, <https://doi.org/10.1007/s10712-015-9346-y>.

Bischiniotis K, B. van den Hurk, E. Zsoter, E. Coughlan de Perez, M. Grillakis, and J. Aerts, 2019: Evaluation of a global ensemble flood prediction system in Peru. *Hydrological Sciences Journal*, 64:10, 1171–1189, <https://doi.org/10.1080/02626667.2019.1617868>.

Bischiniotis, K., H. de Moel, M. van den Homberg, A. Couasnon, J. Aerts, G. Guimarães Nobre, E. Zsoter, and B. van den Hurk, 2020: A framework for comparing permanent and forecast-based flood risk-reduction strategies. *Sci. Total Environ*, 720, <https://doi.org/10.1016/j.scitotenv.2020.137572>.

Blyth, E., D. B. Clark, R. Ellis, C. Huntingford, S. Los, M. Pryor, M. Best, and S. Sitch, 2011: A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale. *Geosci. Model Dev.*, 4, 255–269, <https://doi.org/10.5194/gmd-4-255-2011>.

Blyth, E. M., and Coauthors, 2021: Advances in Land Surface Modelling. *Curr. Clim. Change Rep.*, 7, 45–71, <https://doi.org/10.1007/s40641-021-00171-5>.

Bouilloud, L., and Coauthors, 2010: Coupling the ISBA Land Surface Model and the TOPMODEL Hydrological Model for Mediterranean Flash-Flood Forecasting: Description, Calibration, and Validation. *J. Hydrometeorol.*, 11, 315–333, <https://doi.org/10.1175/2009JHM1163.1>.

Bonan, G., 2008: Ecological Climatology: Concepts and Applications (2nd ed.). *Cambridge University Press*, <https://doi.org/10.1017/CBO9780511805530>.

- Boone, A., and P. Etchevers, 2001: An intercomparison of three snow schemes of varying complexity coupled to the same land surface model: Local-scale evaluation at an Alpine site. *J. Hydrometeorol.*, 2, 374–394, [https://doi.org/10.1175/1525-7541\(2001\)002<0374:AIOTSS>2.0.CO;2](https://doi.org/10.1175/1525-7541(2001)002<0374:AIOTSS>2.0.CO;2).
- Boussetta, S., and Coauthors, 2013: Natural land carbon dioxide exchanges in the ECMWF Integrated Forecasting System: Implementation and offline validation. *J. Geophys. Res.*, 118, 5923–5946. <https://doi.org/10.1002/jgrd.50488>.
- Boussetta, S., G. Balsamo, G. Arduini, E. Dutra, J. McNorton, M. Choulga, A. Agustí-Panareda, A. Beljaars, N. Wedi, J. Muñoz-Sabater, P. de Rosnay, I. Sandu, I. Hadade, G. Carver, C. Mazzetti, C. Prudhomme, D. Yamazaki, and E. Zsoter, 2021: ECLand: The ECMWF Land Surface Modelling System. *Atmosphere*, 12, 723, <https://doi.org/10.3390/atmos12060723>.
- Bouttier, F., and P. Courtier, 1999: Data Assimilation Concepts and Methods. *NWP lecture notes*, pp 59, ECMWF, available online at <https://www.ecmwf.int/node/16928>.
- Brasnett, B., 1999: A global analysis of snow depth for numerical weather prediction. *J. Appl. Meteor.*, 38, 726–740, [https://doi.org/10.1175/1520-0450\(1999\)038<0726:AGAOSD>2.0.CO;2](https://doi.org/10.1175/1520-0450(1999)038<0726:AGAOSD>2.0.CO;2).
- Burek, P., J. M. van der Knijff, and A. P. J. D. de Roo, 2013: LISFLOOD – Distributed Water Balance and Flood Simulation Model – Revised User Manual. *Publications Office of the European Union*, <https://doi.org/10.2788/24719>.
- Burke, E. J., R. Dankers, C. D. Jones, and A. J. Wiltshire, 2013: A retrospective analysis of pan Arctic permafrost using the JULES land surface model. *Climate Dynamics*, 41(3-4), 1025–1038, <https://doi.org/10.1007/s00382-012-1648-x>.
- Cao, B., S. Gruber, D. Zheng, and X. Li, 2020: The ERA5-Land soil temperature bias in permafrost regions. *Cryosphere*, 14, 2581–2595, <https://doi.org/10.5194/tc-14-2581-2020>.
- Cao, B., G. Arduini, and E. Zsoter, 2022: Brief communication: Improving ERA5-Land soil temperature in permafrost regions using an optimized multi-layer snow scheme. *Cryosphere*, 16, 2701–2708, <https://doi.org/10.5194/tc-16-2701-2022>.
- Carrera, M. L., S. Bélair, and B. Bilodeau, 2015: The Canadian land data assimilation system (CaLDAS): Description and synthetic evaluation study. *J. Hydrometeorol.*, 16, 1293–1314, <https://doi.org/10.1175/JHM-D-14-0089.1>.
- Chow, V.T., D. Maidment, and L. W. Mays, 1988: Applied Hydrology. *McGraw-Hill*, New York.
- Cloke, H. L., and F. Pappenberger, 2009: Ensemble flood forecasting: a review. *Journal of Hydrology*, 375, 613–626, <https://doi.org/10.1016/j.jhydrol.2009.06.005>.

Clark, M. P., A. Slater, A. P. Barrett, L. E. Hay, G. J. McCabe, B. Rajagopalan, and G. H. Leavesley, 2006: Assimilation of snow covered area information into hydrologic and land-surface models. *Advances in Water Resources*, 29, 1209–1221, <https://doi.org/10.1016/j.advwatres.2005.10.001>.

Clark, M. P., and Coauthors, 2015: Improving the representation of hydrologic processes in earth system models. *Water Resources Research*, 51, 5929–5956, <https://doi.org/10.1002/2015wr017096>.

Clark, M. P., R. Zolfaghari, K. R. Green, S. Trim, W. J. Knoben, A. Bennett, and R. J. Spiteri, 2021: The numerical implementation of land models: Problem formulation and laugh tests. *J. Hydrometeorol.*, 22, 1627–1648, <https://doi.org/10.1175/JHM-D-20-0175.1>.

Connolly R., M. Connolly, W. Soon, D. R. Legates, R. G. Cionco, and H. V. M. Velasco Herrera, 2019: Northern Hemisphere Snow-Cover Trends (1967–2018): A Comparison between Climate Models and Observations. *Geosciences*, 9, 135, <https://doi.org/10.3390/geosciences9030135>.

Coughlan de Perez, E. C., 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>.

Coughlan de Perez, E. C., M. van Aalst, K. Bischiniotis, S. Mason, H. Nissan, F. Pappenberger, E. Stephens, E. Zsoter, and B. van den Hurk, 2018: Global predictability of temperature extremes. *Environmental Research Letters*, 12(5), 054017, <https://doi.org/10.1088/1748-9326/aab94a>.

Cucchi M., G. P. Weedon, A. Amici, N. Bellouin, S. Lange, H. M. Schmied, H. Hersbach, and C. Buontempo, 2020: WFDE5: bias adjusted ERA5 reanalysis data for impact studies. *Earth Syst. Sci. Data*, 12, 2097–2120, <https://doi.org/10.5194/essd-12-2097-2020>.

Dai, A., T. Qian, K. E. Trenberth, and J. D. Milliman, 2009: Changes in continental freshwater discharge from 1948 to 2004. *Journal of Climate*, 22(10), 2773–2792, <https://doi.org/10.1175/2008JCLI2592.1>.

Day, J. J., G. Arduini, I. Sandu, M. Linus, A. Beljaars, G. Balsamo, M. Rodwell, and D. Richardson, 2020: Measuring the impact of a new snow model using surface energy budget process relationships. *J. of Adv. in Model. Earth Syst.*, 12, e2020MS002144, <https://doi.org/10.1029/2020MS002144>.

de Boisseson, E., H. Zuo, E. Zsoter, S. Harrigan, F. Wetterhall, P. de Rosnay, and C. Prudhomme, 2021: Evaluating a dynamically modelled river discharge as input for ocean systems through the monitoring of the ocean state from reanalysis. *EGU General Assembly*, online, 19–30 Apr 2021, EGU21-14450, <https://doi.org/10.5194/egusphere-egu21-14450>.

- Decharme, B., E. Brun, A. Boone, C. Delire, P. Le Moigne, and S. Morin, 2016: Impacts of snow and organic soils parameterization on northern Eurasian soil temperature profiles simulated by the ISBA land surface model. *Cryosphere*, 10, 853–877, <https://doi.org/10.5194/tc-10-853-2016>.
- Decharme, B., and Coauthors, 2019: Recent changes in the ISBA-CTRIP land surface system for use in the CNRM-CM6 climate model and in global off-line hydrological applications, *J. Adv. Model. Earth Syst.*, 11, 1207–1252, <https://doi.org/10.1029/2018MS001545>.
- Dee, D., and A. Da Silva, 1998: Data Assimilation in the presence of forecast bias. *Quarterly Journal of the Royal Meteorological Society*, 124(545), 269–295, <https://doi.org/10.1016/B978-0-12-088759-0.00022-5>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.*, 137, 553–597, <https://doi.org/10.1002/qj.828>.
- De Lannoy, G. J. M., R. H. Reichle, K. R. Arsenault, P. R. Houser, S. Kumar, N. E. C. Verhoest, and V. R. N. Pauwels, 2012: Multiscale assimilation of Advanced Microwave Scanning Radiometer-EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado. *Water Resources Research*, 48, W01522, <https://doi.org/10.1029/2011WR010588>.
- Demeritt, D., S. Nobert, H. Cloke, and F. Pappenberger, 2010: Challenges in communicating and using ensembles in operational flood forecasting. *Meteorological Applications*, 17, 209–222, <https://doi.org/10.1002/met.194>.
- Depetris, P. J., 2021: The importance of monitoring river water discharge. *Front. water*, <https://doi.org/10.3389/frwa.2021.745912>.
- Do, H. X., S. Westra, and M. Leonard, 2017: A global-scale investigation of trends in annual maximum streamflow. *J. Hydrol.*, 552, 28–43, <https://doi.org/10.1016/j.jhydrol.2017.06.015>.
- Donat, M. G., A. L. Lowry, L. V. Alexander, P. A. O’Gorman, and N. Maher, 2016: More extreme precipitation in the world’s dry and wet regions. *Nat. Clim. Change*, 6, 508–513, <https://doi.org/10.1038/nclimate2941>.
- Donnelly, C, J. M. Andersson, and B. Arheimer, 2016: Using flow signatures and catchment similarities to evaluate a multi-basin model (E-HYPE) across Europe. *Hydr. Sciences Journal*, 61(2), 255–273, <https://doi.org/10.1080/02626667.2015.1027710>.
- Dorigo, W., R. de Jeu, D. Chung, R. Parinussa, Y. Liu, W. Wagner, and D. Fernández-Prieto, 2012: Evaluating global trends (1988–2010) in harmonized multi-satellite surface soil moisture. *Geophys. Res. Lett.*, 39 (18), <https://doi.org/10.1029/2012GL052988>.

- Dottori, F., W. Szewczyk, J.-C. Ciscar, F. Zhao, L. Alfieri, Y. Hirabayashi, and A. Bianchi, 2018: Increased Human and Economic Losses from River Flooding with Anthropogenic Warming. *Nat. Clim. Chang.*, 8, 781–786, <https://doi.org/10.1038/s41558-018-0257-z>.
- Drusch, M., D. Vasiljevic, and P. Viterbo, 2004: ECMWF's Global Snow Analysis: Assessment and Revision Based on Satellite Observations. *J. Appl. Meteor.*, 43, 1282–1294, [https://doi.org/10.1175/1520-0450\(2004\)043<1282:EGSAAA>2.0.CO;2](https://doi.org/10.1175/1520-0450(2004)043<1282:EGSAAA>2.0.CO;2).
- Drusch, M., and P. Viterbo, 2007: Assimilation of screen-level variables in ECMWF's Integrated Forecast System: A study on the impact on the forecast quality and analyzed soil moisture. *Mon. Wea. Rev.*, 135, 300–314, <https://doi.org/10.1175/MWR3309.1>.
- Dutra, E., G. Balsamo, P. Viterbo, P. Miranda, A. Beljaars, C. Schär, and K. Elder, 2010: An improved snow scheme for the ECMWF land surface model: Description and offline validation. *J. Hydrometeorol.*, 11, 899–916, <https://doi.org/10.1175/2010JHM1249.1>.
- Dutra, E., C. Schär, P. Viterbo, and P. M. A. Miranda, 2011: Land-atmosphere coupling associated with snow cover. *Geophys. Res. Lett.*, 38, L15707, <https://doi.org/10.1029/2011GL048435>.
- Dutra E., P. Viterbo, P. M. A. Miranda, and G. Balsamo, 2012: Complexity of Snow Schemes in a Climate Model and Its Impact on Surface Energy and Hydrology. *J. Hydrometeorol.*, 13, 521–538, <https://doi.org/10.1175/JHM-D-11-072.1>.
- Emerton, R., and Coauthors, 2016: Continental and global scale flood forecasting systems. *WIREs Water*, 3, 391–418, <https://doi.org/10.1002/wat2.1137>.
- Emerton, R., H. Cloke, E. Stephens, E. Zsoter, S. J. Woolnough, and F. Pappenberger, 2017: Complex picture for likelihood of ENSO-driven flood hazard. *Nat Commun.*, 8, 14796, <https://doi.org/10.1038/ncomms14796>.
- Emerton, R., E. Zsoter, L. Arnal, H. L. Cloke, D. Muraro, C. Prudhomme, E. Stephens, P. Salamon, and F. Pappenberger, 2018: Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0. *Geosci. Model Dev.*, 11, 3327–3346, <https://doi.org/10.5194/gmd-11-3327-2018>.
- Faulkner D., T. Sampson, S. Warren and F. Byrne, 2019: Can we still predict the future from the past? Non-stationary flood frequency analysis in Ireland. *Irish National Hydrology Conference 2019*.
- Fekete, B. M., U. Looser, A. Pietroniro, and R. D. Robarts, 2012: Rationale for monitoring discharge on the ground. *J. Hydrometeorol.*, 13, 1977–1986, <https://doi.org/10.1175/jhm-d-11-0126.1>.

- Feng, H., and M. Zhang, 2016: Global land moisture trends: drier in dry and wetter in wet over land. *Sci. Rep.*, 5, 18018, <https://doi.org/10.1038/srep18018>.
- Feng, D., and Coauthors, 2021: Recent changes to Arctic river discharge. *Nat. Commun.*, 12, 6917, <https://doi.org/10.1038/s41467-021-27228-1>.
- Ferguson, C. R., and G. Villarini, 2012: Detecting inhomogeneities in the twentieth century reanalysis over the central United States. *Journal of Geophysical Research: Atmospheres*, 117, <https://doi.org/10.1029/2011JD016988>.
- Ficchi, A., H. Cloke, C. Neves, S. Woolnough, E. Coughlan de Perez, E. Zsoter, I. Pinto, A. Meque, and E. Stephens, 2021: Beyond El Nino: unsung climate modes drive African floods. *Weather and Climate Extremes*, 33, 100345. ISSN 2212-0947, <https://doi.org/10.1016/j.wace.2021.100345>.
- Fisher, R. A., C. D. Koven, 2020: Perspectives on the Future of Land Surface Models and the Challenges of Representing Complex Terrestrial Systems. *J. Adv. Model. Earth Syst.*, 12, e2018MS001453, <https://doi.org/10.1029/2018MS001453>.
- Flato, G. M., 2011: Earth system models: An overview. *WIREs Clim. Change*, 2, 783–800, <https://doi.org/10.1002/wcc.148>.
- Folton, N., E. Martin, P. Arnaud, P. L'Hermite, and M. Tolsa, 2019: A 50-year analysis of hydrological trends and processes in a Mediterranean catchment. *Hydrol. Earth Syst. Sci.*, 23, 2699–2714, <https://doi.org/10.5194/hess-23-2699-2019>.
- Gómez, B., C. L. Charlton-Pérez, H. Lewis, and B. Candy, 2020: The Met Office Operational Soil Moisture Analysis System. *Remote Sensing*, 12(22), 3691, <https://doi.org/10.3390/rs12223691>.
- Gouttevin, I., G. Krinner, P. Ciais, J. Polcher, and C. Legout, 2012: Multi-scale validation of a new soil freezing scheme for a land-surface model with physically-based hydrology. *Cryosphere*, 6, 407–430, <https://doi.org/10.5194/tc-6-407-2012>.
- Griessinger, N., J. Seibert, J. Magnusson, and T. Jonas, 2016: Assessing the benefit of snow data assimilation for runoff modeling in Alpine catchments. *Hydrol. Earth Syst. Sci.*, 20, 3895–3905, <https://doi.org/10.5194/hess-20-3895-2016>.
- Gumbel, E. J., 1941: The return period of flood flows. *Ann. Math. Stat.*, 12(2), 163–190, <https://doi.org/10.1214/aoms/1177731747>.
- Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez, 2009: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.*, 377, 80–91, <https://doi.org/10.1016/j.jhydrol.2009.08.003>.

Haddeland, I., and Coauthors, 2011: Multimodel estimate of the global terrestrial water balance: Setup and first results. *J. Hydrometeor.*, 12, 869–884, <https://doi.org/10.1175/2011JHM1324.1>.

Hancock, S., R. Baxter, J. Evans, and B. Huntley, 2013: Evaluating global snow water equivalent products for testing land surface models. *Remote Sens. Environ.*, 128, 107–117, <https://doi.org/10.1016/j.rse.2012.10.004>.

Hansen, J., S. Makiko, R. Ruedy, K. Lo, D. W. Lea, and M. Medina-Elizade, 2006: Global temperature change. *PNAS*, 6, 14288–14293, <https://doi.org/10.1073/pnas.0606291103>.

Harrigan, S., H. Cloke, F. and Pappenberger, 2020a: Innovating global hydrological prediction through an Earth system approach. *WMO Bulletin*, 69 (1), ISSN 0042-9767.

Harrigan, S., E. Zsoter, L. Alfieri, C. Prudhomme, P. Salamon, F. Wetterhall, H. Cloke, and F. Pappenberger, 2020b: GloFAS-ERA5 operational global river discharge reanalysis 1979–present. *Earth Syst. Sci. Data*, <https://doi.org/10.5194/essd-12-2043-2020>.

Harrigan, S., E. Zsoter, H. Cloke, P. Salamon, and C. Prudhomme, 2023: 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-27-1-2023>.

Hawkins, E., and Coauthors, 2017: Estimating changes in global temperature since the preindustrial period. *Bull. Am. Meteorol. Soc.*, 98, 1841–1856, <https://doi.org/10.1175/BAMS-D-16-0007.1>.

Helfrich, S. R., D. McNamara, B. Ramsay, T. Baldwin, and T. Kasheta, 2007: Enhancements to, and forthcoming developments in the interactive multisensor snow and ice mapping system, (IMS). *Hydrol. Processes*, 21, 1576–1586, <https://doi.org/10.1002/hyp.6720>.

Helmert, J., and Coauthors, 2018: Review of snow data assimilation methods for hydrological, land surface, meteorological and climate models: Results from a COST HarmoSnow survey. *Geosciences*, 8, 489, <https://doi.org/10.3390/geosciences8120489>.

Hersbach, H., and Coauthors, 2015: ERA-20CM: a twentieth-century atmospheric model ensemble. *Q. J. R. Meteorol. Soc.*, 141, 2350–2375, <https://doi.org/10.1002/qj.2528>.

Hersbach, H., and D. Dee, 2016: ERA5 reanalysis is in production. *ECMWF Newsletter*, No. 147, ECMWF, Reading, UK, <https://www.ecmwf.int/sites/default/files/elibrary/2016/16299-newsletter-no147-spring-2016.pdf>.

Hersbach, H., P. de Rosnay, B. Bell, D. Schepers, A. J. Simmons, C. Soci, S. Abdalla, M. A. Balmaseda, G. Balsamo, P. Bechtold, P. Berrisford, J. Bidlot, E. de Boisseson, M. Bonavita, P. Browne, R. Buizza, P. Dahlgren, D. P. Dee, R. Dragani, M. Diamantaki, J. Flemming, R. Forbes, A. J. Geer, T. Haiden, E.

- V. Holm, L. Haimberger, R. Hogan, A. Horanyi, M. Janiskova, P. Laloyaux, P. Lopez, J. Munoz Sabater, C. Peubey, R. Radu, D. Richardson, J.-N. Thepaut, F. Vitart, X. Yang, E. Zsoter, and H. Zuo, 2018: Operational global reanalysis: progress, future directions and synergies with NWP. *ERA Report Series*, No. 27, ECMWF, Reading, UK, <https://doi.org/10.21957/tkic6g3wm>.
- Hersbach H., and Coauthors, 2020: The ERA5 Global Reanalysis. *Q. J. R. Meteorol. Soc.*, 146, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Hirpa, F. A., P. Salamon, L. Alfieri, J. T. Pozo, E. Zsoter, and F. Pappenberger, 2016: The effect of reference climatology on global flood forecasting. *J. Hydrometeorol.*, 17, 1131–1145, <https://doi.org/10.1175/JHM-D-15-0044.1>.
- Hirpa, F. A., F. Pappenberger, L. Arnal, C. A. Baugh, H. L. Cloke, E. Dutra, R. E. Emerton, B. Bevilla-Romero, P. Salamon, P. J. Smith, E. Stephens, F. Wetterhall, E. Zsoter, and J. Thielen-del Pozo, 2018a: Global Flood Forecasting for Averting Disasters Worldwide, chapter 12 in “Global Flood Hazard: Applications in modeling, mapping and forecasting”. In: Guy Schumann, Paul Bates, Giuseppe Aronica, Heiko Apel (eds.), *Wiley*, <https://doi.org/10.1002/9781119217886.ch12>.
- Hirpa, F. A., P. Salamon, H. E. Beck, V. Lorini, L. Alfieri, E. Zsoter, S. J. Dadson, 2018b: Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data. *J. Hydrol.*, 566, 595–606, doi:10.1016/j.jhydrol.2018.09.052.
- Hodges, K. I., R. W. Lee, and L. Bengtsson, 2011: A comparison of extratropical cyclones in recent reanalyses ERA-Interim, NASA MERRA, NCEP CFSR, and JRA-25. *Journal of Climate*, 24(18), 4888–4906. <https://doi.org/10.1175/2011JCLI4097.1>.
- Horton, P.; Schaefli, B.; Kauzlaric, M. Why do we have so many different hydrological models? A review based on the case of Switzerland. *WIREs Water* 2021, 9, <https://doi.org/10.1002/wat2.1574>.
- Hosking, J. R. M., 1990: L-moments: analysis and estimation of distributions using linear combinations of order statistics. *J. R. Stat. Soc. Ser. B Methodol.*, 52, 105–124, <https://doi.org/10.1111/j.2517-6161.1990.tb01775.x>.
- Hossain, F., A. H. Siddique-E-Akbor, S. Biancamaria, H. Lee, and C. K. Shum, 2014: Proof of concept of an altimeter-based river forecasting system for transboundary flow inside Bangladesh. *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, 7, 587–601, <https://doi.org/10.1109/JSTARS.2013.2283402>.
- Hsu, W. R., and A. H. Murphy, 1986: The attributes diagram: a geometrical framework for assessing the quality of probability forecast. *International Journal of Forecasting*, 285–293, [https://doi.org/10.1016/0169-2070\(86\)90048-8](https://doi.org/10.1016/0169-2070(86)90048-8).

Huffman, G. J., R. F. Adler, D. T. Bolvin, and G. Gu, 2009: Improving the global precipitation record: GPCP version 2.1. *Geophys. Res. Lett.*, 36, <https://doi.org/10.1029/2009GL040000>.

IPCC AR4 report, 2007, Intergovernmental Panel on Climate Change, https://www.ipcc.ch/site/assets/uploads/2018/05/ar4_wg1_full_report-1.pdf.

IPCC AR5 report WG1 contribution, 2013, Intergovernmental Panel on Climate Change, https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_all_final.pdf.

IPCC AR5 report, 2014, Intergovernmental Panel on Climate Change, https://www.ipcc.ch/site/assets/uploads/2018/05/SYR_AR5_FINAL_full_wcover.pdf.

Javadian, M., A. Behrangi, W. K. Smith, and J. B. Fisher, 2020: Global Trends in Evapotranspiration Dominated by Increases across Large Cropland Regions. *Remote Sens.*, 12(7), <https://doi.org/10.3390/rs12071221>.

Jiang S., L. Kang, 2019: Flood frequency analysis for annual maximum streamflow using a non-stationary GEV model. *E3S Web of Conferences*, 79:203022, <https://doi.org/10.1051/e3sconf/20197903022>.

Johnson, S., and Coauthors, 2018: SEAS5: the new ECMWF seasonal forecast system. *Geoscientific Model Development*, 1–44, <https://doi.org/10.5194/gmd-2018-228>.

Jung, M., M. Reichstein, and P. Ciais, 2020: Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, 467, 951–954, <https://doi.org/10.1038/nature09396>.

Kauffeldt, A., S. Halldin, F. Pappenberger, F. Wetterhall, C.-Y. Xu, and H. L. Cloke, 2015: Imbalanced land surface water budgets in a numerical weather prediction system. *Geophys. Res. Lett.*, 42, 4411–4417, <https://doi.org/10.1002/2015GL064230>.

Keller, J. D., and S. Wahl, 2020: Representation of climate in reanalyses—an intercomparison for Europe and North America. *Journal of Climate*, 34, 1667–1684, <https://doi.org/10.1175/jcli-d-20-0609.1>.

Kjeldsen, T. R., R. Lamb, and S. D. Blazkova, 2014: Uncertainty in flood frequency analysis. In K. Beven, & J. Hall (Eds.), *Applied Uncertainty Analysis for Flood Risk Management*, 153–197, Imperial College Press.

Kjeldsen, T. R., R. Lamb, and S. D. Blazkova, 2014: Uncertainty in flood frequency analysis. In: Beven, K., Hall, J. (Eds.). *Applied Uncertainty Analysis for Flood Frequency Analysis*, Imperial Collage Press, London.

- Kling, H., M. Fuchs, and M. Paulin, 2012: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.*, 424–425, 264–277, <https://doi.org/10.1016/j.jhydrol.2012.01.011>.
- Knoben, W. J. M., J. E. Freer, and R. A. Woods, 2019: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrol. Earth Syst. Sci.*, 23, 4323–4331, <https://doi.org/10.5194/hess-23-4323-2019>.
- Knowles N., 2015: Trends in Snow Cover and Related Quantities at Weather Stations in the Conterminous United States. *J. Climate*, 28, 7518–7528, <https://doi.org/10.1175/JCLI-D-15-0051.1>.
- Koczo, K. M., A. E. Jeton, B. J. McGurk, and M. D. Dettinger, 2005: Precipitation-runoff processes in the Feather River Basin, northeastern California, with prospects for streamflow predictability, water years 1971–97. U.S. *Geological Survey Scientific Investigations*, Rep 2004-5202, 82 pp., <https://doi.org/10.3133/sir20045202>.
- Koven, C. D., W. J. Riley, and A. Stern, 2013: Analysis of permafrost thermal dynamics and response to climate change in the CMIP5 earth system models. *J. Clim.*, 26, 1877–1900, <https://doi.org/10.1175/JCLI-D-12-00228.1>.
- Koren, V., M. Smith, and Z. Cui, 2014: Physically-based modifications to the Sacramento Soil Moisture Accounting model. Part A: Modeling the effects of frozen ground on the runoff generation process. *J. Hydrol.*, 519, 3475–3491, <https://doi.org/10.1016/j.jhydrol.2017.05.042>.
- Krogh, S. A., J. W. Pomeroy, and P. Marsh, 2017: Diagnosis of the hydrology of a small Arctic basin at the tundra-taiga transition using a physically based hydrological model. *J. Hydrol.*, 550, 685–703, <https://doi.org/10.1016/j.jhydrol.2017.05.042>.
- Kunkel, K. E., D. A. Robinson, and S. Champion, 2016: Trends and Extremes in Northern Hemisphere Snow Characteristics. *Curr. Clim. Change Rep.*, 2, 65–73, <https://doi.org/10.1007/s40641-016-0036-8>.
- Larger, C., and Coauthors, 2020: Toward snow cover estimation in mountainous areas using modern data assimilation methods: a review. *Front. Earth Sci.*, 8, <https://doi.org/10.3389/feart.2020.00325>.
- Lavers, D. A., F. Pappenberger, D. S. Richardson, and E. Zsoter, 2016: ECMWF extreme forecast index for water vapor transport: A forecast tool for atmospheric rivers and extreme precipitation. *Geophysical Research Letters*, 43, 11, 852–858. <https://doi.org/10.1002/2016GL071320>.

Lavers, D. A., E. Zsoter, D. S. Richardson, and F. Pappenberger, 2017: An assessment of the ECMWF extreme forecast index for water vapor transport during boreal winter. *Weather and Forecasting*, 32(4), 1667–1674, <https://doi.org/10.1175/WAF-D-17-0073.1>.

Lavers, D., and D. Richardson, A. Ramos, E. Zsótér, F. Pappenberger, and R. Trigo, 2018: Earlier awareness of extreme winter precipitation across the western Iberian Peninsula. *Meteorological Applications*, 25, <https://doi.org/10.1002/met.1727>.

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.*, <https://doi.org/10.1088/1748-9326/ab52b2>.

Lavers, D. A., A. Simmons, F. Vamborg, and M. J. Rodwell, 2022: An evaluation of ERA5 precipitation for climate monitoring. *Q. J. R. Meteorol. Soc.*, 148, 3152–3165, <https://doi.org/10.1002/qj.4351>.

Legates, D. R., and G. J. McCabe Jr., 1999: Evaluating the use of “goodness-of-fit” Measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.*, 35, 233–241, <https://doi.org/10.1029/1998WR900018>.

Lehner, B., and Coauthors, 2011: High-resolution mapping of the world’s reservoirs and dams for sustainable river-flow management. *Front. Ecol. Environ.*, 9, 494–502, <https://doi.org/10.1890/100125>.

Lehner, B., and P. Döll, 2004: Development and validation of a global database of lakes, reservoirs and wetlands. *J. Hydrol.*, 296, 1–22, <https://doi.org/10.1016/j.jhydrol.2004.03.028>.

Leutbecher, M., and Coauthors, 2017: Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Q. J. R. Meteorol. Soc.*, <https://doi.org/10.1002/qj.3094>.

Lin, P., and Coauthors, 2019: Global Reconstruction of Naturalized River Flows at 2.94 Million Reaches. *Water Resour. Res.*, 55, 6499–6516, <https://doi.org/10.1029/2019WR025287>.

Lievens, H., and Coauthors, 2015: SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia. *Remote Sensing of Environment*, 168, 146–162, <https://doi.org/10.1016/j.rse.2015.06.025>.

López-Moreno, J. I., and J. M. García-Ruiz, 2004: Influence of snow accumulation and snowmelt on streamflow in the Central Spanish Pyrenees. *Hydrol. Sci. J.*, 49, 787–802, <https://doi.org/10.1623/hysj.49.5.787.55135>.

- Magnusson, J., N. Wever, R. Essery, N. Helbig, A. Winstral, and T. Jonas, 2015: Evaluating snow models with varying process representations for hydrological applications. *Water Resour. Res.*, 51, 2707–2723, <https://doi.org/10.1002/2014WR016498>.
- Mahfouf J.-F., P. Viterbo, H. Douville, A. C. M. Beljaars, and S. Saarinen, 2000: A revised land-surface analysis scheme in the Integrated Forecasting System. *ECMWF Newsletter*, No. 88, ECMWF, Reading, United Kingdom, 8–13, <https://www.ecmwf.int/node/14636>.
- Mahto, S. S., and V. Mishra, 2019: Does ERA-5 Outperform Other Reanalysis Products for Hydrologic Applications in India? *Journal of Geophysical Research: Atmospheres*, 124 (16), 9423–9441, <https://doi.org/10.1029/2019JD031155>.
- Melton, J. R., D. L. Versegny, R. Sospedra-Alfonso, and S. Gruber, 2019: Improving permafrost physics in the coupled Canadian Land Surface Scheme (v.3.6.2) and Canadian Terrestrial Ecosystem Model (v.2.1) (CLASS-CTEM). *Geosci. Model Dev.*, 12, 4443–4467, <https://doi.org/10.5194/gmd-12-4443-2019>.
- Mengelkamp, H.-T., K. Warrach, C. Ruhe, and E. Raschke, 2001: Simulation of runoff and streamflow on local and regional scales. *Meteor. Atmos. Phys.*, 76, 76–107, <https://doi.org/10.1007/s007030170042>.
- Molotch, N. P., and R. C. Bales, 2005: Scaling snow observations from the point to the grid element: Implications for observation network design. *Water Resour. Res.*, 41, W11421, <https://doi.org/10.1029/2005WR004229>.
- Muñoz-Sabater, J., E. Dutra, A. Agustí-Panareda, C. Albergel, G. Arduini, G. Balsamo, S. Boussetta, M. Choulga, S. Harrigan, H. Hersbach, B. Martens, D. G. Miralles, M. Piles, N. J. Rodríguez-Fernández, E. Zsoter, C. Buontempo, and J.-N. Thépaut, 2021: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data*, 13, 4349–4383, <https://doi.org/10.5194/essd-13-4349-2021>.
- Murphy, A. H., 1973: A new vector partition of the probability score. *Journal of Appl. Meteorology*, 12 (4), 595–600, [https://doi.org/10.1175/1520-0450\(1973\)012<0595:ANVPOT>2.0.CO;2](https://doi.org/10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2).
- Nash, J. E., and J. V. Sutcliffe, 1970: River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.*, 10, 282–290, [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6).
- Nguyen, P., A. Thorstensen, S. Sorooshian, K. Hsu, A. Aghakouchak, H. Ashouri, H. Tran, and D. Braithwaite, 2018: Global precipitation trends across spatial scales using satellite observations. *Bull. Amer. Meteor. Soc.*, 99, 689–697, <https://doi.org/10.1175/BAMS-D-17-0065.1>.

Niu, G. Y., Z. L. Yang, 2006: Effects of frozen soil on snowmelt runoff and soil water storage at a continental scale. *J. Hydrometeorol*, 7, 937–952, <https://doi.org/10.1175/JHM538.1>.

Niu, G. Y., Z. L. Yang, 2007: An observation-based formulation of snow cover fraction and its evaluation over large North American river basins. *J. Geophys. Res. Atmos.*, 112, <https://doi.org/10.1029/2007JD008674>.

McNorton, J., and Coauthors, 2021: An Urban Scheme for the ECMWF Integrated Forecasting System: Single-Column and Global Offline Application. *J. Adv. Model. Earth Syst.*, 4, <https://doi.org/10.1029/2020MS002375>.

d'Orgeval, T., J. Polcher, and P. de Rosnay, 2008: Sensitivity of the West African hydrological cycle in ORCHIDEE to infiltration processes. *Hydrol. Earth Syst. Sci.*, 12, 1387–1401, <https://doi.org/10.5194/hess-12-1387-2008>.

Overgaard, J., D. Rosbjerg, and M. B. Butts, 2006: Land-surface modelling in hydrological perspective—A review. *Biogeosciences*, 3, 229–241, <https://doi.org/10.5194/bg-3-229-2006>.

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

Pan, M., and E. Wood, 2006: Data assimilation for estimating the terrestrial water budget using a constrained ensemble Kalman filter. *J. Hydrometeor.*, 7, 534–547, <https://doi.org/10.1175/JHM495.1>.

Pan, N., S. Wand, Y. Liu, W. Zhao, and B. Fu, 2019: Global Surface Soil Moisture Dynamics in 1979–2016 Observed from ESA CCI SM Dataset. *Water*, 11(5), 883; <https://doi.org/10.3390/w11050883>.

Papalexiou, S. M., and D. Koutsoyiannis, 2013: Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resour. Res.*, 49, 187–201, doi:10.1029/2012WR012557.

Pappenberger, F., E. Dutra, F. Wetterhall, and H. L. Cloke, 2012: Deriving global flood hazard maps of fluvial floods through a physical model cascade. *Hydrol. Earth Syst. Sci.*, 16, 4143–4156, <https://doi.org/10.5194/hess-16-4143-2012>.

Parmesan C., and G. Yohe, 2003: Globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421, 37–42, <https://doi.org/10.1038/nature01286>.

Passerotti, G., G. Massazza, A. Pezzoli, V. Bigi, E. Zsótér, and M. Rosso, 2020: Hydrological Model Application in the Sirba River: Early Warning System and GloFAS Improvements. *Water*, 12, 620, <https://doi.org/10.3390/w12030620>.

- Pathiraja, S., D. Anghileri, P. Burlando, A. Sharma, L. Marshall, and H. Moradkhani, 2018: Insights on the impact of systematic model errors on data assimilation performance in changing catchments. *Advances in Water Resources*, 113, 202-222, ISSN 0309-1708, <https://doi.org/10.1016/j.advwatres.2017.12.006>.
- Pavelsky, T. M., M. T. Durand, K. M. Andreadis, R. E. Beighley, R. C. D. Paiva, G. H. Allen, and Z. F. Miller, 2014: Assessing the potential global extent of SWOT river discharge observations. *Journal of Hydrology*, 519 (PB), 1516-1525, <https://doi.org/10.1016/j.jhydrol.2014.08.044>.
- Pearson, K., 1896: Mathematical contributions to the theory of evolution: III. Regression, heredity and panmixia. *Philos. Trans. Roy. Soc., London*, 187A, 253–318, <https://doi.org/10.1098/rsta.1896.0007>.
- Pitman, A. J., 2003: The evolution of, and revolution in, land surface schemes designed for climate models. *International Journal of Climatology*, 23(5), 479–510. <https://doi.org/10.1002/joc.893>.
- Prentice, I. C.; X. Liang, B. E. Medlyn, and Y.-P. Wang, 2015: Reliable, robust and realistic: The three R's of next-generation land-surface modelling. *Atmos. Chem. Phys.*, 15, 5987–6005, <https://doi.org/10.5194/acp-15-5987-2015>.
- Pullen, S., C. Jones, and G. Rooney, 2011: Using satellite-derived snow cover data to implement a snow analysis in the met office NWP model. *J. Appl. Meteor.*, 50, 958–973, <https://doi.org/10.1175/2010JAMC2527.1>.
- Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, 126, 1143–1170, <https://doi.org/10.1002/qj.49712656415>.
- Rabier, F., 2005: Overview of global data assimilation developments in numerical weather prediction centres. *Quart. J. Roy. Meteor. Soc.*, 131, 3215–3233, <https://doi.org/10.1256/qj.05.129>.
- Riihelä, A., R. M. Bright, and K. Anttila, 2021: Recent strengthening of snow and ice albedo feedback driven by Antarctic sea-ice loss. *Nat. Geosci.*, 14, 832–836, <https://doi.org/10.1038/s41561-021-00841-x>.
- Rodda J. C., S. A. Pieyns, N. S. Sehmi, and G. Matthews, 1993: Towards a world hydrological cycle observing system. *Hydrol. Sci. J.* 38, 373–378., <https://doi.org/10.1080/026266693099492687>.
- Rodell, M., Coauthors, 2004: The Global Land Data Assimilation System, *Bull. Amer. Meteor. Soc.*, 85 (3), 381–394, <https://doi.org/10.1175/BAMS-85-3-381>.

Romanovsky, V. E., M. Burgess, S. Smith, K. Yoshikawa, and J. Brown, 2002: Permafrost temperature records: Indicators of climate change. *Eos. Trans. Am. Geophys. Union*, 83, 589–594, <https://doi.org/10.1029/2002EO000402>.

Rosmann, T., E. Domínguez, and J. Chavarro, 2016: Comparing trends in hydrometeorological average and extreme data sets around the world at different time scales. *J. Hydrol. Reg. Stud.*, 5, 200–212, <https://doi.org/10.1016/j.ejrh.2015.12.061>.

de Rosnay, P., M. Drusch, D. Vasiljevic, G. Balsamo, C. Albergel, and L. Isaksen, 2013: A simplified extended Kalman filter for the global operational soil moisture analysis at ECMWF. *Quart. J. Roy. Meteor. Soc.*, 139, 1199–1213, <https://doi.org/10.1002/qj.2023>.

de Rosnay P., G. Balsamo, and C. Albergel, 2014: Initialisation of Land Surface Variables for Numerical Weather Prediction. *Surv. Geophys.*, 35, 607–621, <https://doi.org/10.1007/s10712-012-9207-x>.

de Rosnay P., L. Isaksen, and M., Dahoui, 2015: Snow data assimilation at ECMWF. *ECMWF newsletter*, 143, <https://www.ecmwf.int/sites/default/files/elibrary/2015/17328-snow-data-assimilation-ecmwf.pdf>.

Saha, S. K., K. Sujith, S. Pokhrel, H. S. Chaudhari, and A. Hazra, 2017: Effects of multilayer snow scheme on the simulation of snow: Offline Noah and coupled with NCEP CFS v2. *J. Adv. Mode. Earth Syst.*, 9, 271–290, <https://doi.org/10.1002/2016MS000845>.

Santoro, M., V. Andres, S. Jirka, T. Koike, U. Looser, S. Nativi, F. Pappenberger, M. Schlummer, A. Strauch, M. Utech, and E. Zsoter, 2018: Interoperability challenges in river discharge modelling: a cross domain application scenario. *Comput. Geosci.*, 115, 66–74, <https://doi.org/10.1016/j.cageo.2018.03.008>.

Schneider, T., S. Lan, A. Stuart, and J. Teixeira, 2017: Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophysical Research Letters*, 44(24), 12396–12417, <https://doi.org/10.1002/2017gl076101>.

Scheppe, R., S. Thober, M. Kelbling, R. Kumar, S. Attinger, and L. Samaniego, 2022: MPR 1.0: A stand-alone Multiscale Parameter Regionalization Tool for Improved Parameter Estimation of Land Surface Models. *Geosci. Model Dev.*, 15, 859–882, <https://doi.org/10.5194/gmd-15-859-2022>.

Slater, A. G., Coauthors, 2001: The representation of snow in land surface schemes: Results from PILPS 2(d). *J. Hydrometeorol.*, 2, 7–25, [https://doi.org/10.1175/1525-7541\(2001\)002<0007:TROSIL>2.0.CO;2](https://doi.org/10.1175/1525-7541(2001)002<0007:TROSIL>2.0.CO;2).

- Slater, A. G., and M. P. Clark, 2006: Snow data assimilation via an ensemble Kalman filter. *J. Hydrometeor.*, 7, 478–493, <https://doi.org/10.1175/JHM505.1>.
- Slater, L. J., and Coauthors, 2021: Nonstationary weather and water extremes: a review of methods for their detection, attribution, and management. *Hydrol. Earth Syst. Sci.*, 25, 3897–3935, <https://doi.org/10.5194/hess-25-3897-2021>.
- Sleigh, M., W. Deconinck, M. Lange, O. Marsden, and B. Reuter, 2022: An open-source Integrated Forecasting System. *ECMWF Newsletter*, No. 171, ECMWF, Reading, United Kingdom, 13–14.
- Spinoni, J., G. Naumann, and J. V. Vogt, 2017: Pan-European seasonal trends and recent changes of drought frequency and severity. *Global Planet. Change*, 148, 113–130, <https://doi.org/10.1016/j.gloplacha.2016.11.013>.
- Stahl K., L. M. Tallaksen, J. Hannaford, and H. A. J. van Lanen, 2012: Filling the white space on maps of European runoff trends: estimates from a multi-model ensemble. *Hydrology Earth System Sciences*, 16 (7), 2035–2047, <https://doi.org/10.5194/hess-16-2035-2012>.
- Stephens, E., J. J. Day, F. Pappenberger, and H. Cloke, 2015: Precipitation and floodiness. *Geophys. Res. Lett.*, 42, 316–323, doi:10.1002/2015GL066779.
- Su L., C. Miao, D. Kong, Q. Duan, X. Lei, Q. Hou, and H. Li, 2018: Long-term trends in global river flow and the causal relationships between river flow and ocean signals. *Journal of Hydrology*, <https://doi.org/10.1016/j.jhydrol.2018.06.058>.
- Sun Q., C. Miao, Q. Duan, H. Ashouri, S. Sorooshian, and K. L. Hsu, 2018: A review of global precipitation datasets: Data sources, estimation, and intercomparisons. *Rev. Geophys.*, 56, 79–107, <https://doi.org/10.1002/2017RG000574>.
- Sun, Q., X. Zhang, F. Zwiers, S. Westra, and L. V. Alexander, 2020: A global, continental and regional analysis of changes in extreme precipitation. *J. Climate*, 34, 1–52, <https://doi.org/10.1175/JCLI-D-19-0892.1>.
- Thielen, J., J. Bartholmes, M.-H. Ramos, and A. de Roo, 2009: The European Flood Alert System – Part 1: Concept and development. *Hydrol. Earth Syst. Sci.*, 13, 125–140, <https://doi.org/10.5194/hess-13-125-2009>.
- Thorne, P.W., and R. S. Vose, 2010: Reanalyses suitable for characterizing long-term trends. *Bulletin of the American Meteorological Society*, 91(3), 353–361. <https://doi.org/10.1175/2009BAMS2858.1>.

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*, 22 (7). 1801–1817, ISSN 1525-7541, <https://doi.org/10.1175/JHM-D-20-0250.1>.

Towner, J., H. L. Cloke, E. Zsoter, Z. Flamig, J. M. Hoch, J. Bazo, E. Coughlan de Perez, and E. M. Stephens, 2019: Assessing the performance of global hydrological models for capturing peak river flows in the Amazon basin. *Hydrol. Earth Syst. Sci.*, 23, 3057–3080, <https://doi.org/10.5194/hess-23-3057-2019>.

Tsonevsky, I., C. A. Doswell, and H. E. Brooks, 2018: Early Warnings of Severe Convection Using the ECMWF Extreme Forecast Index. *Wea. Forecasting*, 33, 857–871, <https://doi.org/10.1175/WAF-D-18-0030.1>.

UNDRR, Human Cost of Disasters: an Overview of the Last 20 Years, 2019, <https://dds.cepal.org/redesoc/publication?id=5361>.

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. Inf. Sci.*, 24, 189–212, <https://doi.org/10.1080/13658810802549154>.

Verkade, J. S., J. D. Brown, P. Reggiani, and A. Weerts, 2013: Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. *J. Hydrol.*, 501, 73–91, <https://doi.org/10.1016/j.jhydrol.2013.07.039>.

Le Vine, N., A. Butler, N. McIntyre, and C. Jackson, 2016: Diagnosing hydrological limitations of a land surface model: Application of JULES to a deep-groundwater chalk basin. *Hydrol. Earth Syst. Sci.*, 20, 143–159, <https://doi.org/10.5194/hess-20-143-2016>.

Vionnet, V., E. Brun, S. Morin, A. Boone, S. Faroux, P. Le Moigne, E. Martin, and J.-M. Willemet, 2012: The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2. *Geosci. Model Dev.*, 5, 773–791, <https://doi.org/10.5194/gmd-5-773-2012>.

Vitart, F., 2014: Evolution of ECMWF sub-seasonal forecast scores. *Quart. J. Roy. Meteor. Soc.*, 140, 1889–1899, <https://doi.org/10.1002/qj.2256>.

Viterbo, P., A. Beljaars, J. F. Mahfouf, and J. Teixeira, 1999: The representation of soil moisture freezing and its impact on the stable boundary layer. *Q. J. R. Meteorol. Soc.*, 125, 2401–2426.

Walters, D., Coauthors, 2019: The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations. *Geosci. Model Dev.*, 12, 1909–1963, <https://doi.org/10.5194/gmd-12-1909-2019>.

- Wang, T., and Coauthors, 2013: Evaluation of an improved intermediate complexity snow scheme in the ORCHIDEE land surface model. *J. Geophys. Res. Atmos.*, 118, 6064–6079, <https://doi.org/10.1002/jgrd.50395>.
- Wang, L. L., D. H. Chen, and H. J. Bao, 2016: The improved Noah land surface model based on storage capacity curve and Muskingum method and application in GRAPES model. *Atmos. Sci. Lett.*, 17, 190–198, <https://doi.org/10.1002/asl.642>.
- Wedi N. P., and Coauthors, 2015: The modelling infrastructure of the Integrated Forecasting System: Recent advances and future challenges. *ECMWF Technical Memoranda*, 760, pp50.
- Weeink, W. H., 2010: Evaluation of streamflow and ensemble thresholds (Master's Thesis). *University of Twente*, <https://webgr.irstea.fr/wp-content/uploads/2012/07/2010-WEEINK-MASTER.pdf>.
- Wentao, L., D. Qingyun, M. Chiyuan, Y. Aizhong, G. Wei, and D. Zhenhua, 2017: A review on statistical postprocessing methods for hydrometeorological ensemble forecasting. *WIREs Water*, 4, e1246, <https://doi.org/10.1002/wat2.1246>.
- Westra, S., L. V. Alexander, and F. W. Zwiers, 2013: Global increasing trends in annual maximum daily precipitation. *J. Climate*, 26, 3904–3918, <https://doi.org/10.1175/JCLI-D-12-00502.1>.
- Wiltshire, A. J., and Coauthors, 2020: JULES-GL7: the Global Land configuration of the Joint UK Land Environment Simulator version 7.0 and 7.2. *Geosci. Model Dev.*, 13, 483–505, <https://doi.org/10.5194/gmd-13-483-2020>.
- Winkelbauer, S., M. Mayer, V. Seitner, E. Zsoter, H. Zuo, and L. Haimberger, 2022: Diagnostic evaluation of river discharge into the Arctic Ocean and its impact on oceanic volume transports. *Hydrol. Earth Syst. Sci.*, 26, 279–304, <https://doi.org/10.5194/hess-26-279-2022>.
- WMO Guidelines on the calculation of Climate Normals. 2017, *WMO*, No. 1203.
- Wu, H., R. F. Adler, Y. Tian, G. J. Huffman, H. Li, and J. Wang, 2014: Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model. *Water Resour. Res.*, 50, 2693–2717, <https://doi.org/10.1002/2013WR014710>.
- Wu, W., R. Emerton, Q. Duan, A. W. Wood, F. Wetterhall, and D. E. Robertson, 2020: Ensemble flood forecasting: Current status and future opportunities. *WIREs Water*, 3, <https://doi.org/10.1002/wat2.1432>.

Yamazaki, D., S. Kanae, H. Kim, and T. Oki, 2011: A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resour. Res.*, 47, W04501, <https://doi.org/10.1029/2010WR009726>.

Yang, Z. L., R. E. Dickinson, A. Robock, and K. Y. Vinnikov, 1997: Validation of the snow submodel of the biosphere–atmosphere transfer scheme with Russian snow cover and meteorological observational data. *J. Clim.*, 10, 353–373, [https://doi.org/10.1175/1520-0442\(1997\)010<0353:VOTSSO>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<0353:VOTSSO>2.0.CO;2).

Yokohata, T., K. Saito, K. Takata, T. Nitta, Y. Satoh, T. Hajima, T. Sueyoshi, and G. Iwahana, 2020: Model improvement and future projection of permafrost processes in a global land surface model. *Prog. Earth Planet. Sci.*, 69, <https://doi.org/10.1186/s40645-020-00380-w>

Yuan, X., and E. F. Wood, 2012: Downscaling precipitation or bias-correcting streamflow? Some implications for coupled general circulation model (CGCM)-based ensemble seasonal hydrologic forecast. *Water Resour. Res.*, 48, <https://doi.org/10.1029/2012WR012256>.

Zaitchik, B. F., and M. Rodell, 2009: Forward-looking assimilation of MODIS-derived snow-covered area into a land surface model. *J. Hydrometeor.*, 10, 130–148, <https://doi.org/10.1175/2008JHM1042.1>.

Zajac, Z., B. Revilla-Romero, P. Salamon, P. Burek, F. A. Hirpa, and H. Beck, 2017: The impact of lake and reservoir parameterization on global streamflow simulation. *J. Hydrol.*, 548, 552–568, <https://doi.org/10.1016/j.jhydrol.2017.03.022>.

T. Zhang, R. G. Barry, K. Knowles, J. A. Heginbottom, and J. Brown, 2008: Statistics and characteristics of permafrost and ground-ice distribution in the Northern Hemisphere. *Polar Geography*, 31:1-2, 47-68, <https://doi.org/10.1080/10889370802175895>.

Zhang Y., J. Peña-Arancibia, and T. McVicar, 2016: Multi-decadal trends in global terrestrial evapotranspiration and its components. *Sci. Rep.*, 6, 19124, <https://doi.org/10.1038/srep19124>.

Zhao, F., and Coauthors, 2017: The critical role of the routing scheme in simulating peak river discharge in global hydrological models. *Environ. Res. Lett.*, 12, 075003, <https://doi.org/10.1088/1748-9326/aa7250>.

Zsoter, E., F. Pappenberger, and D. Richardson, 2014: Sensitivity of model climate to sampling configurations and the impact on the Extreme Forecast Index. *Meteorol. Appl.*, 22, 236–247, <https://doi.org/10.1002/met.1447>.

-
- Zsoter, E., F. Pappenberger, P. Smith, R. E. Emerton, E. Dutra, F. Wetterhall, D. Richardson, K. Bogner, and G. Balsamo, 2016: Building a Multimodel Flood Prediction System with the TIGGE Archive. *J. Hydrometeorol.*, 17:11, 2923-2940, <https://doi.org/10.1175/JHM-D-15-0130.1>.
- Zsoter E., H. Cloke, E. Stephens, P. de Rosnay, J. Muñoz-Sabater, C. Prudhomme, and F. Pappenberger, 2019: How well do operational Numerical Weather Prediction setups represent hydrology? *J. Hydrometeorol.*, 14, 1533–1552, <https://doi.org/10.1175/JHM-D-18-0086.1>.
- Zsoter, E., C. Prudhomme, E. Stephens, F. Pappenberger, and H. Cloke, 2020a: Using ensemble reforecasts to generate flood thresholds for improved global flood forecasting. *J. Flood Risk Manag.*, <https://doi.org/10.1111/jfr3.12658>.
- Zsoter, E., H. L. Cloke, C. Prudhomme, S. Harrigan, P. de Rosnay, J. Munoz-Sabater, and E. Stephens, 2020b: Trends in the GloFAS-ERA5 river discharge reanalysis. *ECMWF Technical Memoranda*, 871, <https://doi.org/10.21957/p9jrh0xp>.
- Zsoter, E., G. Arduini, C. Prudhomme, E Stephens, and H. Cloke, 2022: Hydrological Impact of the New ECMWF Multi-Layer Snow Scheme. *Atmosphere*, 13, 727. <https://doi.org/10.3390/atmos13050727>.

Appendix

This appendix contains the typeset versions of each of the published chapters presented in this thesis, alongside the more important further publications during this PhD. All author contribution statements (provided in the respective chapters or in the Appendix) have been approved by Professor Hannah Cloke, supervisor.

Hannah L. Cloke

A1: How well do operational numerical weather prediction setups represent hydrology?

This paper presents the published version of chapter 4 of this thesis, with the following reference:

Zsoter, E., H Cloke, E. Stephens, P. de Rosnay, J. Muñoz-Sabater, C. Prudhomme and F. Pappenberger, 2019: How well do operational Numerical Weather Prediction setups represent hydrology?, *J. Hydrometeorol.*, 14, doi:10.1175/JHM-D-18-0086.1

How Well Do Operational Numerical Weather Prediction Configurations Represent Hydrology?

ERVIN ZSOTER

European Centre for Medium-Range Weather Forecasts, and Department of Geography and Environmental Science, University of Reading, Reading, United Kingdom

HANNAH CLOKE

Department of Geography and Environmental Science, and Department of Meteorology, University of Reading, Reading, United Kingdom, and Department of Earth Sciences, Uppsala University, and Centre of Natural Hazards and Disaster Science, Uppsala, Sweden

ELISABETH STEPHENS

Department of Geography and Environmental Science, University of Reading, Reading, United Kingdom

PATRICIA DE ROSNAY, JOAQUIN MUÑOZ-SABATER, CHRISTEL PRUDHOMME, AND FLORIAN PAPPENBERGER

European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

(Manuscript received 27 April 2018, in final form 16 April 2019)

ABSTRACT

Land surface models (LSMs) have traditionally been designed to focus on providing lower-boundary conditions to the atmosphere with less focus on hydrological processes. State-of-the-art application of LSMs includes a land data assimilation system (LDAS), which incorporates available land surface observations to provide an improved realism of surface conditions. While improved representations of the surface variables (such as soil moisture and snow depth) make LDAS an essential component of any numerical weather prediction (NWP) system, the related increments remove or add water, potentially having a negative impact on the simulated hydrological cycle by opening the water budget. This paper focuses on evaluating how well global NWP configurations are able to support hydrological applications, in addition to the traditional weather forecasting. River discharge simulations from two climatological reanalyses are compared: one “online” set, which includes land–atmosphere coupling and LDAS with an open water budget, and an “offline” set with a closed water budget and no LDAS. It was found that while the online version of the model largely improves temperature and snow depth conditions, it causes poorer representation of peak river flow, particularly in snowmelt-dominated areas in the high latitudes. Without addressing such issues there will never be confidence in using LSMs for hydrological forecasting applications across the globe. This type of analysis should be used to diagnose where improvements need to be made; considering the whole Earth system in the data assimilation and coupling developments is critical for moving toward the goal of holistic Earth system approaches.

1. Introduction

Land surface models (LSMs) have traditionally been designed to focus on providing lower-boundary conditions to the atmosphere by describing the vertical fluxes

 Denotes content that is immediately available upon publication as open access.

Corresponding author: E. Zsoter, ervin.zsoter@ecmwf.int

DOI: 10.1175/JHM-D-18-0086.1

© 2019 American Meteorological Society



This article is licensed under a Creative Commons Attribution 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

of energy and water between the land surface and the atmosphere, with less focus on predicting runoff (Mengelkamp et al. 2001). LSMs therefore maximize the quality of the atmospheric forecast, but do not necessarily bring the same benefits in the representation of the hydrological cycle (Kauffeldt et al. 2015).

There is a wide literature on assessing the hydrological capabilities of LSMs and describing various improvements in the modeling of the hydrological cycle (e.g., Balsamo et al. 2009; Wang et al. 2016; Blyth et al. 2011; Wu et al. 2014). However, there are significant limitations in the representation of hydrological fluxes and storages in LSMs, largely due to the large-scale focus of LSM applications, which has led to the neglect of some important processes for runoff generation (Overgaard et al. 2006; Le Vine et al. 2016), including inadequate snowmelt processes (Dutra et al. 2012; Zaitchik and Rodell 2009).

Data assimilation is an essential part of any numerical weather prediction (NWP) system (Rabier 2005). It is designed to provide initial conditions for the Earth system by updating the model in all of the components: atmosphere, land, ocean, and sea ice. State-of-the-art NWP configurations, such as used at the European Centre for Medium-Range Weather Forecasts (ECMWF), include both an LSM and a land data assimilation system (LDAS). The objective of the data assimilation in this context is to combine the land surface model state with the available land surface observations to initialize the land surface model prognostic variables of the forecasting system (Bélair et al. 2003). The current ECMWF LDAS analyses soil moisture, soil temperature, snow mass, density, and temperature (de Rosnay et al. 2014). Land data assimilation was shown to contribute significantly to more skillful atmospheric forecasts, with the soil moisture data assimilation also proven essential in countering a positive precipitation/evapotranspiration feedback which can cause large positive precipitation biases (e.g., de Rosnay et al. 2013; Drusch and Viterbo 2007; Beljaars et al. 1996).

While the improved surface conditions make LDAS an essential component of the ECMWF NWP system, by design the related increments remove or add water which can potentially have a negative impact on the representation of the hydrological cycle by opening the water budget (Zaitchik and Rodell 2009; Arsenault et al. 2013; Andreadis and Lettenmaier 2006; De Lannoy et al. 2012; Pan and Wood 2006). On the contrary, in a system without LDAS and coupling, the errors resulting from atmospheric forcing insufficiencies and imperfect land surface process

representations are not corrected by the assimilation of land surface observations.

As an ideal configuration, an Earth system model should always maintain a closed water budget, where the amount of water in the system remains the same. By opening the water budget, river discharge biases could emerge in situations where the LSM has an energy balance bias that is not corrected by the assimilation but only by accurate precipitation and snow accumulation forcing. For example, if the snow in the LSM is melting too slowly, this forces the LDAS to remove water (through snow) artificially to correct for the excessive amount of snow on the surface. If the water that is removed with the snow (and thus could not melt) is not retained within the Earth system that could lead to soil water deficit downstream, potentially causing an incorrect rate of river discharge. In such cases, LDAS could lead to replace incorrect snowmelt timing issue with incorrect snowmelt runoff amount.

Thus, an open water budget could cause problems for associated hydrological forecasting applications, which uses runoff calculated from LSMs with LDAS, such as the Global Flood Awareness System (GloFAS; Alfieri et al. 2013). As global hydrological modeling is increasingly possible with the improved realism that the state-of-the-art LSMs can nowadays offer (Overgaard et al. 2006), it is important to investigate how an LSM with LDAS can support the combined task of traditional weather forecasting and hydrology at the same time. This investigation was undertaken with this dual focus in mind, by analyzing the hydrological cycle and the open water budget issues that can help the Earth system model developments with highlighting areas where the coupled system with LDAS does not yet work effectively for flood simulations.

To understand how well an NWP configuration with LSM and LDAS represents hydrology, and in particular to interpret the influence of the LDAS on hydrological simulations from LSMs, in this paper river discharge simulations from two climatological reanalyses of GloFAS are compared: one operational set, which includes land-atmosphere coupling and LDAS with an open water budget, and also an "offline" set with a closed water budget and no LDAS. From these two datasets, a range of hydrological and atmospheric variables will be analyzed globally.

2. System description, datasets, and methods

Two hydrological experiments, ONLINE (run in operational mode with active land-atmosphere coupling and LDAS) and OFFLINE (run in offline mode without coupling and LDAS) provide time series of various surface variables (e.g., 2-m temperature, snow depth, and runoff), and also discharge after routing the

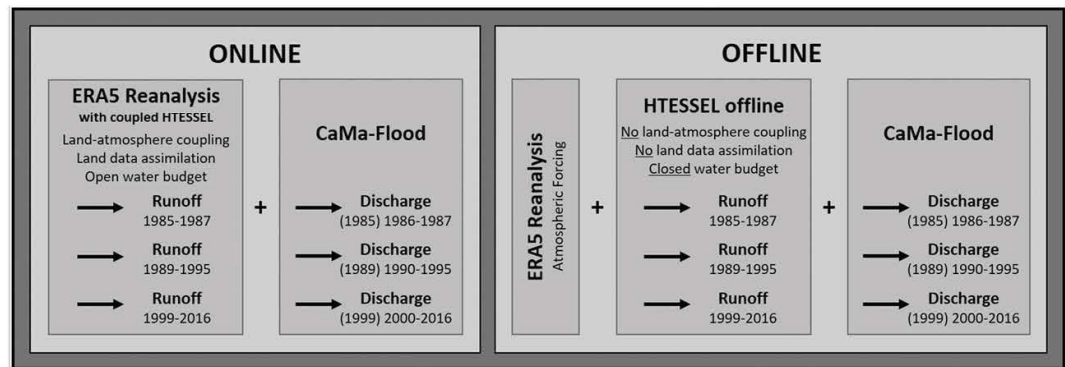


FIG. 1. Schematic of the ONLINE and OFFLINE experiments that were carried out to produce the ERA5-D25 dataset. The years in parentheses for the discharge indicate the first spinup year in each period that was excluded from the analysis.

runoff. Figure 1 highlights the schematic of ONLINE and OFFLINE with the main characteristics, components and data periods. In this section the two experiments with the model and data aspects, and the data analysis methods will be described in detail.

a. Land surface model HTESEL

The hydrological component of the analyzed datasets is based on the Hydrology Tiled ECMWF Scheme of Surface Exchanges over Land (HTESEL) land surface model (Balsamo et al. 2009, 2011). HTESEL is part of the ECMWF NWP system and used in coupled land-atmosphere mode on time ranges from short-range to seasonal forecasts. It includes a snow parameterization based on a single-layer snowpack model (Dutra et al. 2010). The soil vertical diffusion solves the Richards equation using a four-layer vertical discretization with layer depths at 7, 28, 100, and 289 cm (Balsamo et al. 2009). HTESEL provides boundary conditions for the atmosphere (heat, moisture, and momentum) by simulating water and energy budgets on the surface and through the soil, snowpack, and vegetation interception. HTESEL generates surface (fast) and subsurface (slow) runoff components at each grid point (Balsamo et al. 2009). Surface runoff depends on the standard deviation of the orography, soil texture, and soil moisture, while subsurface runoff is determined by the soil water percolation.

b. Land data assimilation

The ECMWF LDAS is part of the ECMWF Integrated Forecasting System (IFS). It is coupled to the atmospheric four-dimensional variational data assimilation (4D-Var) scheme (Rabier et al. 2000), both using a 12-h assimilation window. The upper-air and land surface analyses are running separately and are used to initialize a coupled land-atmosphere short-term forecast, which provides the

background for the next data assimilation window. The land data assimilation relies on advanced methods to optimally combine in situ and satellite observations with model background information. A schematic diagram of the ECMWF LDAS is provided in Fig. 2.

Initial implementations of the ECMWF LDAS relied on simple assimilation methods for snow and soil moisture analyses (Drusch et al. 2004; Mahfouf et al. 2000), with air temperature and humidity measurements being the main input for the soil moisture analysis (Mahfouf et al. 2000; Drusch and Viterbo 2007). The system has evolved in the past decade to use a more physically based approach and to combine satellite and in situ data in the soil analysis (de Rosnay et al. 2014; de Rosnay et al. 2013; Albergel et al. 2012).

In the current LDAS, a simplified extended Kalman filter (SEKF) is used to analyze soil moisture. The approach combines analyzed 2-m air temperature and humidity with satellite measurements from the Advanced Scatterometer (ASCAT) sensor on board of MetOp, as described in de Rosnay et al. (2013) and Albergel et al. (2012). For snow, a two-dimensional optimal interpolation (OI) is used to analyze snow mass and snow density following the method described in Brasnett (1999). In situ snow depth observations, available on the SYNOP network are used along with the 4-km resolution snow cover product from the NOAA National Environmental Satellite, Data, and Information Service (NOAA/NESDIS) Interactive Multisensor Snow and Ice Mapping System (IMS) product (Helfrich et al. 2007).

Even though it provides significant improvements to the atmospheric forecasts and independent in situ snow depth measurements (de Rosnay et al. 2015), the current ECMWF snow data assimilation follows a relatively basic method. Operational NWP configurations generally rely on simple approaches, compared to research environment,

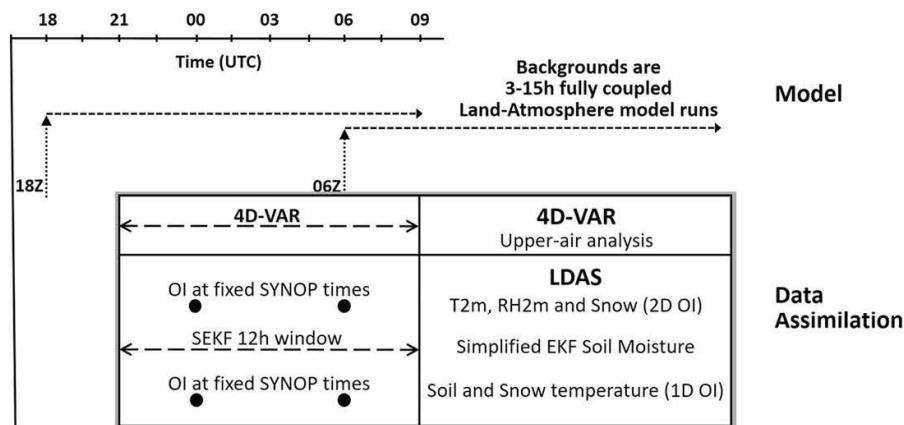


FIG. 2. Schematic diagram of the land data assimilation system at ECMWF.

that are based on more sophisticated snow assimilation methods using in situ and remotely sensed observations (e.g., Helmert et al. 2018; De Lannoy et al. 2012; Pan and Wood 2006; Slater and Clark 2006).

The ECMWF LDAS and its performance is presented and discussed in de Rosnay et al. (2014) and de Rosnay et al. (2015). A full description of the technical implementation is provided in the IFS documentation (<https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model/ifs-documentation>). The system used for this study is that used for the production of ERA5 (section 2f), with IFS cycle 41r2 at a resolution of ~31 km.

c. CaMa-Flood river routing

The Catchment-based Macroscale Floodplain model (CaMa-Flood; Yamazaki et al. 2011) was applied in this study to simulate the hydrodynamics and produce river discharge from the HTESSEL runoff outputs. CaMa-Flood is a distributed global river-routing model which uses a river network map and routes runoff to oceans or inland seas. The CaMa-Flood model was chosen for the routing component as it had already been used in several similar climatological research experiments such as Emerton et al. (2017).

d. GloFAS

GloFAS is one of the few global scale flood forecasting systems that currently exist (Emerton et al. 2016). It is part of the Copernicus Emergency Management Service (CEMS), developed by the Joint Research Centre of the European Commission (JRC) and ECMWF. The HTESSEL runoff output is coupled to the Lisflood hydrological model over a global river network to produce

river discharge with a forecast horizon of 30 days across a global river network at 0.1° resolution (van der Knijff et al. 2010; Alfieri et al. 2013). As part of the GloFAS configuration, the real-time river discharge forecasts are compared with climatological simulations (called reanalysis) to detect the likelihood of high-flow situations. These real-time and climatological datasets also present a unique opportunity for experimental analysis (Emerton et al. 2017; Stephens et al. 2015).

e. Offline land surface modeling

The current GloFAS operational setup uses a climatology based on the ERA-Interim/Land reanalysis of ECMWF (Balsamo et al. 2015). ERA-Interim/Land is an improved version of the ERA-Interim reanalysis (Dee et al. 2011) produced with an improved version of HTESSEL, run offline, using a rescaling of monthly precipitation totals with GPCP v2.2 (Huffman et al. 2009; Balsamo et al. 2010). Offline HTESSEL simulations, such as the OFFLINE experiment in this study, are uncoupled from the atmosphere, without the LDAS and forced with near-surface meteorological input data such as temperature, specific humidity, wind speed, surface pressure, radiative fluxes, and water fluxes. Offline land-surface-only simulations are an affordable way of achieving land surface improvements, and this offline research methodology has been used in numerous studies with HTESSEL in the last few decades (e.g., Agustí-Panareda et al. 2010; Dutra et al. 2010, 2011; Haddeland et al. 2011).

f. ERA5 reanalysis

The fifth generation global climate reanalysis (succeeding ERA-Interim) at ECMWF is ERA5 (Hersbach and Dee

2016). ERA5 is a key contribution to the EU-funded Copernicus Climate Change Service (C3S). ERA5 will cover the period from 1950 to present and is in production with 2008–17 already officially released. The release of the remaining period is foreseen by end of 2018. ERA5 will then continue running in (nonquality assured mode) near-real time with only a few days' delay. The data are open access and free to download for all uses (<https://climate.copernicus.eu/>).

ERA5 uses the IFS cycle 41r2 and it relies on land surface model and assimilation configuration that are consistent with those used for operational NWP with coupled land-atmosphere simulations and the latest soil moisture and snow assimilation (see sections 2a and 2b above). ERA5 has a high-resolution component at ~31 km which is used in this study (hereafter called ERA5-HRES). In ERA5-HRES, variables (analysis and short-range forecasts generated at 0600 and 1800 UTC) are available hourly. Variables that are valid for a period, for example, precipitation or runoff with an accumulation time, are provided as hourly forecasts.

At the time of writing, approximately 28 years of ERA5-HRES data were available in the ECMWF MARS data archive in three separate periods: 1985–87, 1989–95, and 1999–2016. The first years (1985, 1989, and 1999) were used as spinup years, so in total 25 years of daily river discharge and other surface data could be processed for the analysis (hereafter called ERA5-D25).

g. Experimental setup

In the ONLINE experiment, the operational ERA5-HRES reanalysis data were used directly from all three ERA5-HRES periods for land surface variables, including runoff, produced by coupled land-atmosphere model with LDAS and an open water budget (Fig. 1). In the OFFLINE experiment, on the other hand, three stand-alone HTESSEL runs were set up, one for each of the periods, to reproduce the land surface variables in land surface only mode without the impact of coupling and LDAS, but with a closed water budget. As ERA5 has a recent model cycle (41r2), the same HTESSEL version could be used in the offline experiment as in the operational ERA5.

In the ECMWF NWP system, there is no option currently to run the land-atmosphere coupling and LDAS separately. Either both are active as in ONLINE, or neither of them as in OFFLINE. It would be interesting to separate the impact of these two contributing modeling options, but as they are too strongly interwoven the separation would require a very large effort, which is outside of the scope of this study.

In the OFFLINE experiment, the offline HTESSEL model was forced with hourly ERA5-HRES atmospheric

data, wherever it was possible on the lowest model level, with an hourly model time step. The model was run on the original horizontal resolution of ERA5-HRES (~31 km). For precipitation, temperature, specific humidity, wind speed, and surface pressure the hourly analysis fields were applied, while for radiation and precipitation fluxes the first 12-h period of the 0600 and 1800 UTC short-range forecasts were used to cover each 24-h periods.

The river discharge was generated by routing the runoff using CaMa-Flood for both the ONLINE and OFFLINE datasets over the ~25 km river network. CaMa-Flood was run with a 1-h time step and a 24-h output frequency to match the 24-h reporting frequency of the river discharge observations.

h. River discharge observations

In this study, daily river discharge observations used in the GloFAS system are selected. These are mostly from the Global Runoff Data Centre (GRDC) archive, an international depository of river discharge observations and associated metadata.

The observations consist of a network of approximately 900 river gauging stations with upstream areas over 10000 km², selected from the catchments used in Zsoter et al. (2016). After visual inspection those catchments that showed a clear nonrealistic behavior and/or influence of dams were excluded. A minimum of 9 years, with at least 330 days in each of those calendar years, was selected as criteria for the stations to be included in the river discharge analysis. This is quite a short period, but due to the limited availability in more recent years, it was accepted as a compromise. In total 590 stations could be processed globally leaving large blank areas mostly in Asia and Africa (Fig. 3).

i. Annual peak river discharge

For the river discharge verification, the annual peak river discharges from the two ERA5-HRES simulations were determined in each calendar year as the highest value in the ±30-day window around the observed annual maximum river flow. The 30-day window was defined as a safeguard to avoid detecting high skill with similar peaks in observation and simulation of completely different flood waves at very different periods of the year.

j. Water budget increments

This study focuses on the impact of the water budget closure on river discharge. To analyze this, the daily (0000–0000 UTC) water budget error term dA was computed as

$$dA = P - E - R - dS, \quad (1)$$

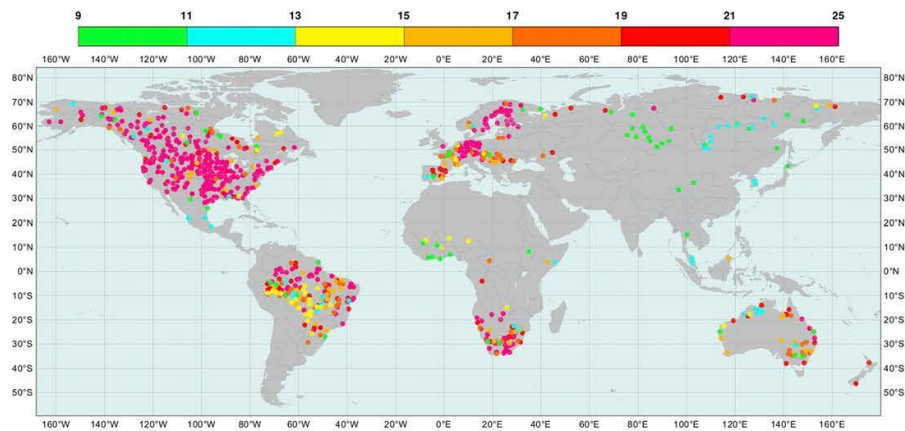


FIG. 3. Geographical distribution of river discharge observations with sufficient record length selected for the analysis. Colors indicate the length of the available data in years (from 9 to 25).

where P is precipitation, E is evapotranspiration, and R is runoff, all taken as the sum of the hourly forecast values (24 in total) in the ONLINE experiment from the 0000–0000 UTC period, and dS is the change in the storage term (water content in the soil including all four layers and also in the snow cover) computed as the difference between the two subsequent 0000 UTC analysis values in ONLINE (representing the change in the water content during the 24-h period). Even though the water budget error is zero in OFFLINE (the water budget is closed), the contributing variables can help identifying the behavior of the surface processes in both the ONLINE and OFFLINE simulations.

The imbalance in the amount of water that is not accounted for in the ONLINE water budget effectively comes from the snow depth and soil moisture increments in LDAS which remove or add water in the system. The daily increments (valid for a 0000–0000 UTC 24-h period) are computed as the sum of two increment values at 0600 and 1800 UTC (each day). Both of these increments are computed as the ERA5-HRES analysis value minus the corresponding 12-h ERA5-HRES forecast value (initialized 12 h earlier).

k. Daily 2-m temperature and snow depth

The in situ surface synoptic observations (SYNOP) were used to verify 2-m temperature and snow depth for both the OFFLINE and ONLINE experiments. The

observing stations were filtered according to the station altitude difference to the model orography and only those were used which had less than 150-m discrepancy, as orography has control on both variables and large differences would make the comparison unreliable. This maximum orography difference value was chosen in accordance with the general practice at ECMWF, where 100 m is used to filter stations in the 2-m temperature verification. For our study, a less stringent compromise value was preferred in order to increase the sample size and still guarantee good match between model and real orography.

The 2-m temperature was verified for around local noon (Table 1), while for snow depth the first measurement of the calendar day was evaluated in case of subdaily records. In total, observations from about 4000 stations for 2-m temperature and 1500 stations for snow depth were available for verification. For each catchment, a representative daily observation was also determined for both variables. For catchments with more than one SYNOP station available, these were calculated as the arithmetic average of the stations within the catchment. It has to be acknowledged that the observation network available was not dense enough to represent the full spatial variability of these surface variables, especially snow depth, which vary dramatically in space from one point to another (Molotch and Bales 2005). However, for a global study on the hydrological impacts it is expected to be sufficient.

TABLE 1. Criteria for selecting daytime 2-m temperature.

Longitude band	30°W–60°E	60°–150°E	150°E–180°	120°W–180°	30°–120°W
Approx. local noon	1200 UTC	0600 UTC	0000 UTC	0000 UTC	1800 UTC

TABLE 2. List of verification scores used in the analysis with a short description and also the areas where they were applied.

Score	Description	Used for
ME	Mean error	Daily river discharge, snow depth, and 2-m temperature
MAE	Mean absolute error	Daily river discharge, snow depth, and 2-m temperature
NSE	Nash–Sutcliffe efficiency	Daily river discharge time series
<i>R</i>	Pearson correlation coefficient	Daily river discharge time series
PMnE	Percentage sample mean error	Whole river discharge sample
PMnAe	Percentage sample mean absolute error	Whole river discharge sample
PStE	Percentage sample standard deviation error	Whole river discharge sample
PStAe	Percentage sample standard deviation absolute error	Whole river discharge sample
PkTiMe	Peak timing mean error	Annual river discharge peaks
PkTiMae	Peak timing mean absolute error	Annual river discharge peaks
PPkMgMe	Percentage peak magnitude mean error	Annual river discharge peaks
PPkMgMae	Percentage peak magnitude mean absolute error	Annual river discharge peaks

l. Climatologies

Daily climatologies were used for river discharge and other surface variables in this work for both observations and the two simulations. These datasets were produced with all potentially available 25 years of data in ERA5-D25, always matching the number of available nearly complete calendar years (with minimum 330 river discharge observations) for all the catchments. For each day of the year a 21-day window, centered over the day, was used, which provided a minimum of about 180 values in the climate sample (with the 9 years minimum criteria). The only exceptions are 2-m temperature and snow depth, where a fixed shorter period of 2000–07 was used without the criteria of nearly complete years. As the 2-m temperature and snow depth observation availability is much better in more recent periods and also less prone to missing values than river discharge, a shorter fixed period (when ERA5-HRES was available) is sufficient.

m. Verification statistics

A number of statistics were applied to evaluate the overall performance of the two climatological simulations in ERA5-D25 (Table 2). Several scores were selected in order to give a more representative description of the general behavior including the differences between the ONLINE and OFFLINE experiments. This is recommended, for example, by Legates and McCabe (1999) as different scores demonstrate different aspects of the model attributes ultimately providing a more complete picture.

The climatological daily time series were compared to the observed data using mean error (ME), mean absolute error (MAE), Nash–Sutcliffe model efficiency (NSE; Nash and Sutcliffe 1970), and also Pearson correlation coefficient *R* (Pearson 1896) in order to measure the fit between model and observations. In addition, the mean and standard deviation of the observed

and modeled values were analyzed with four additional indices, the percentage sample mean error, the percentage sample mean absolute error, the percentage sample standard deviation error, and the percentage sample standard deviation absolute error.

Another very important aspect of hydrological model verification is the ability of the systems to correctly predict the extremes, as these events can cause the highest impact. To measure this, the timing and magnitude errors of the annual peaks were considered. Both the ME and MAE measures (mean of all years in the sample) were computed for the timing and for the percentage magnitude errors using the annual peaks over the 25 analyzed years (for details on how the annual peaks were computed, see section 2i). For the analysis of the data assimilation impact on 2-m temperature and snow depth the ME and MAE scores were used. In this study verification was conducted on homogeneous samples across all compared scores for all the verified surface variables.

3. Results

The river discharge behavior provides a useful indication of the hydrological differences between the ONLINE and OFFLINE simulations. However, in order to understand the underlying processes better, the coupling and LDAS impact was also analyzed globally and regionally based on the water budget and the related surface variables.

a. Snow depth and 2-m temperature impact

The LDAS is designed to provide adequate initial surface conditions to the NWP forecasts. The impact on the hydrology could be demonstrated on two important surface variables: 2-m temperature and snow depth (at least in snow impacted areas) which are relatively well observed variables and can be used to analyze the impact of the land–atmosphere coupling

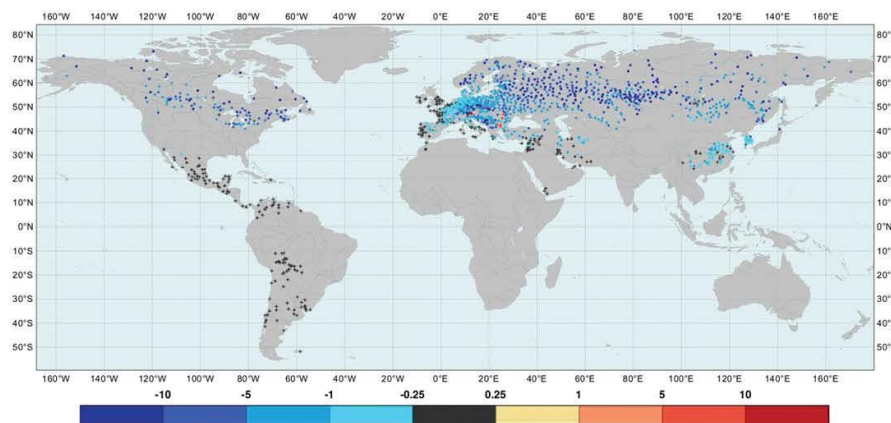


FIG. 4. Difference in the snow depth mean absolute errors between ONLINE and OFFLINE for January based on observations in 2000–07 (cm). Points are shown where observations are available. Blue colors indicate lower errors in the ONLINE experiment.

and LDAS on the surface globally in the two experiments. For details on how the observations were used please see section 2k.

The picture for 2-m temperature is rather mixed geographically with an overall MAE improvement in ONLINE of around 0.3° – 0.4° C as a global average up to 1° – 2° C locally (not shown). This corresponds to about 20%–30% decrease in MAE on average in ONLINE, with the impact of coupling and LDAS, compared to OFFLINE.

The improvement in the snow depth, which has much larger direct impact on the hydrology, is more pronounced, based on the stations used in this study. The errors in ONLINE are significantly reduced with most stations showing below ± 1 – 2 cm of ME (not shown), and decrease of MAE by as much as 10–20 cm in some of the snow dominant locations in the 50° – 70° latitude band (Fig. 4). This is a very large improvement in ONLINE by removing 70%–80% (as global average) of the errors found in the OFFLINE experiment. Countries of Central America, including Mexico, Venezuela, and Columbia, tend to provide snow information in their SYNOP observations. In these regions both the model and the in situ stations mostly indicate snow free conditions, leading to very low MAE as shown in Fig. 4. Although the improvements are large, this does not necessarily mean that the simulation is generally better. In situ snow observations are associated to potential representativeness issues, particularly in mountainous areas. When assimilating a nonrepresentative dataset at a coarse spatial scale, the results can potentially degrade, even though the match to the actual observations is better (Molotch and Bales 2005). As the 2-m

temperature and snow depth observations used in this study for verification were also assimilated in ERA5, the result will favor to some extent the ONLINE experiment.

b. Global water budget analysis

The water budget is closed in OFFLINE by design, while in ONLINE the LDAS increments can add or remove water, which could potentially lead to large errors in the budget over a long period. The first aspect that was important to check is the amount of water that is lost or gained in a day on average in the hydrological cycle.

Figure 5 shows the average daily water budget errors (Fig. 5a) and the related snow water equivalent (Fig. 5b) and soil water content (Fig. 5c) increments (for the definition of these terms please see section 2j). In Fig. 5, negative values (red) indicate water removal by LDAS, while positive values (blue) show where water is added to the hydrological cycle.

The three figures highlight significant biases in the ONLINE experiment as these water budget errors represent generally $\pm 10\%$ – 25% of the total precipitation with locally even higher ratios (not shown). In addition, at latitudes higher than 50° N the dominant pattern is a negative water budget error (Fig. 5a). The major contributing factor to the clearly negative errors in this area is the correction of snowpack with LDAS removing snow to account for possible inaccuracies in the HTESSEL snow scheme (Fig. 5b). On average snow water increments are negative almost everywhere where snow is present. The only notable exception is in Canada, where some central areas have positive water

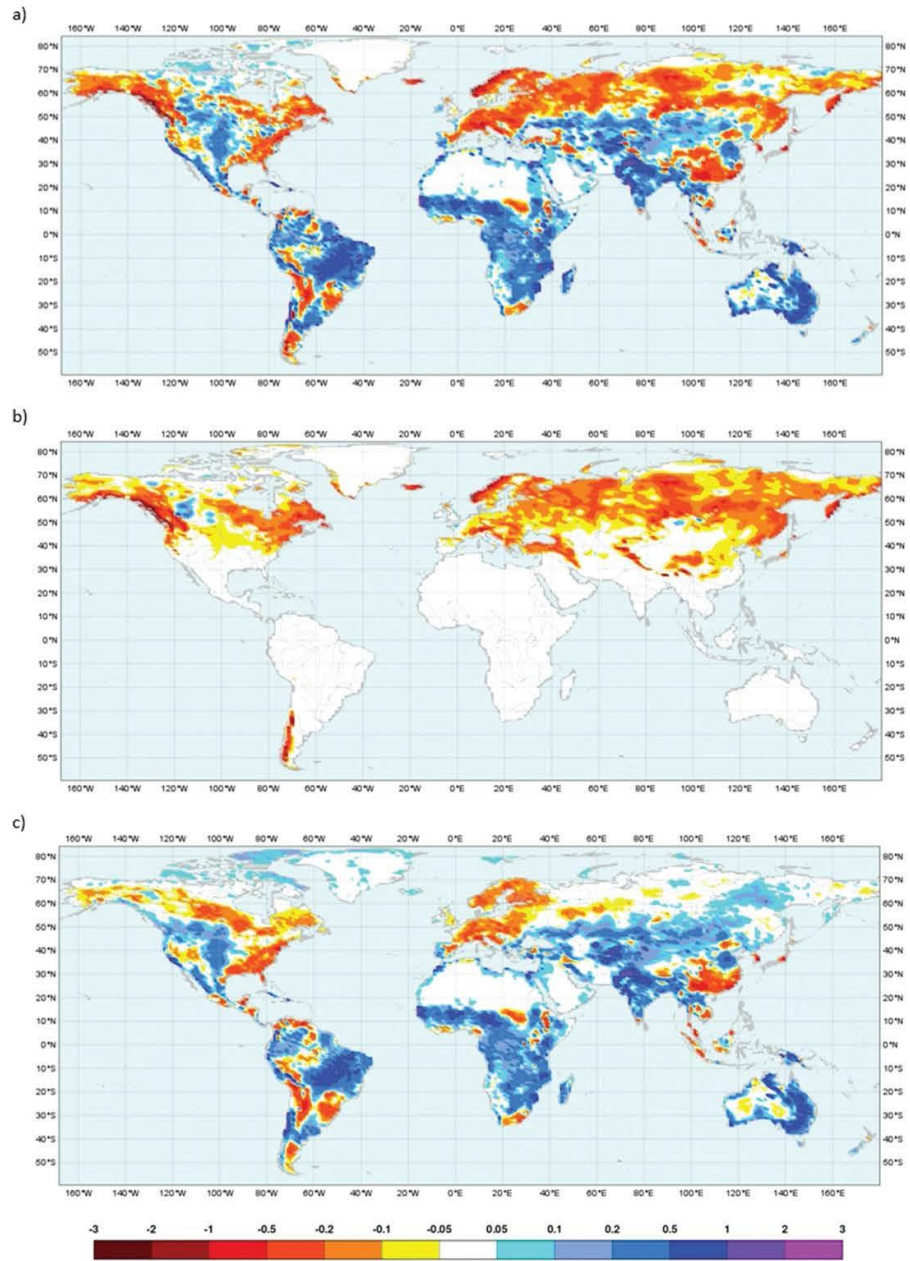


FIG. 5. Average daily water budget analysis (mm day^{-1}) of the ONLINE experiment based on the ERA5-D25 dataset for (a) the total 24-h water budget errors, (b) the 24-h snow water equivalent increments, and (c) the 24-h soil water content increments. Negative values (red) indicate water removal by LDAS, while positive values (blue) show where water is added to the hydrological cycle.

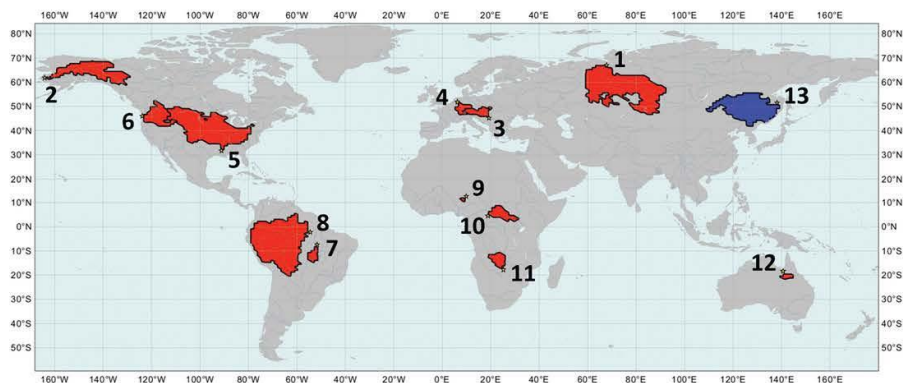


FIG. 6. Map of the catchments analyzed in section 3c (Fig. 7), where the catchment-level process is examined over the Amur River (blue area, 13), and in section 3d (Fig. 8), where the simplified representation of the annual water cycle is shown for some selected regional catchments of the world (red areas, 1–12). The catchment details are provided in Table 3.

budget errors which could possibly come from a negative precipitation bias that needs to be compensated by LDAS.

Other areas of the world—the central United States, most of Amazonia, Africa, South Asia with India, and large parts of Australia—show positive errors in Fig. 5a, where extra water is added by LDAS. However, the positive errors are not exclusive, as large parts of China, the southeastern United States, and areas in central South America experience negative water budget errors in these mostly warm climatic conditions. Most of these increments come from the soil moisture assimilation impact (Fig. 5c). The soil moisture assimilation can generally compensate for precipitation or 2-m temperature biases. For example, if the 2-m temperature is too low, the assimilation will remove water, therefore reducing evaporative cooling which subsequently increase the temperature in general.

c. Catchment-level process examination

To demonstrate how HTESSSEL handles the land surface processes with and without coupling and LDAS, an in-depth case study analysis of the annual water budget cycle was performed for an example catchment on the Amur River in east Russia (see Fig. 6, catchment 13). This catchment is heavily snow impacted during winter and can demonstrate nicely the important aspects of the hydrological cycle behavior with the LDAS in action.

In the HTESSSEL hydrological cycle representation the input precipitation combined with the melted part of the snowpack (snowmelt) is distributed into evapotranspiration, runoff (as sum of surface and subsurface runoffs), snow water storage (falling snow part of the

precipitation) and soil water storage (soil moisture in the four soil layers). The daily water budget error, computed as in Eq. (1) (without the snowmelt separated), is zero in OFFLINE, while ONLINE can show errors due to the increments adding or removing water. Figure 7 summarizes the annual cycle of all the water budget contributing variables.

The displayed variables are daily climatological means calculated as described in section 2I. The following variables are shown in Fig. 7: simulated precipitation (same for both experiments), evapotranspiration, runoff, soil water, and snow water storage terms [in Eq. (1)] for both ONLINE and OFFLINE; snow and soil water content increments for ONLINE; simulated snowmelt, snow depth, and river discharge for both the ONLINE and OFFLINE experiments; and finally the corresponding river discharge and snow depth observations.

Figure 7 shows that for the Amur the ONLINE simulation significantly improves the representation of snow depth, but as consequence, by the snow assimilation removing a lot of snow, it drastically reduces the river discharge peak seen during the snowmelt season. The explanation of this conclusion with detailed analysis of the evolution of the different surface variables in the different seasons is given in the following:

- **Winter:** During December–February there is relatively little activity. The little amount of precipitation falls mostly as snow, building the snowpack. Some snow is removed by the assimilation through the small negative snow increments. Water leaves the bottom of the soil as subsurface runoff with hardly any surface runoff. The OFFLINE simulation is generally

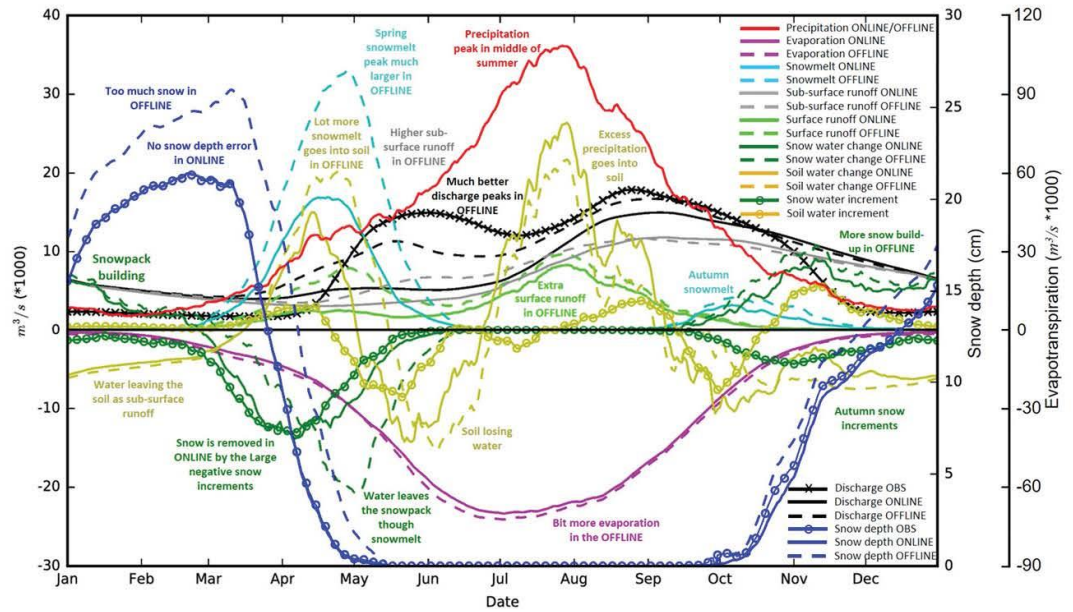


FIG. 7. Average daily water budget cycle for a catchment on the Amur River in Russia at Komsomolsk. It includes the following parameters: precipitation (red line), snow (green line with markers), and soil (mustard line with markers) water content increments for the ONLINE simulation; surface runoff (light green), subsurface runoff (gray), evapotranspiration (magenta), snowmelt (cyan), and soil (mustard) and snow (green) water storage daily changes for both ONLINE (solid lines) and OFFLINE (dashed lines); snow depth (blue); and river discharge (black) for the ONLINE (solid lines) and OFFLINE (dashed lines) experiments and observations (lines with markers). The snow depth values are based on 2000–07 while all other displayed daily climatological means are based on the ERA5-D25 dataset (for more detail on the computation of these values, see sections 2k and 2l).

similar to ONLINE, but snow depth bias shows increasingly positive values in OFFLINE due to the extra amount of water going into the snowpack in the OFFLINE experiment from snowfall (especially during first half of the winter).

- Spring: From March, there is a pronounced snowmelt period in the model, peaking at the end of April, lasting until the middle of June (with virtually zero snowpack in catchment average after middle of May). The increased precipitation in this spring period, with the large amount of snowmelt, increases the soil water content, and also results in larger surface runoff output in both experiments. However, the snowmelt is much smaller in ONLINE during April–May as a direct consequence of the large negative snow increments (peaking early April) removing snow in the ONLINE experiment. Similarly, due to the smaller amount of available water in ONLINE, the surface runoff is also significantly smaller mainly in April/May. The snow depth errors peak in middle of March by about 5 cm in OFFLINE with no errors in ONLINE (as catchment average). The data assimilation rightly corrects this substantial positive snow bias, however,

the removed snow will be missing from the water cycle, as highlighted by the unnoticeable spring peak river flow, which is higher in the OFFLINE simulation mainly due to the extra snowmelt.

- Snowmelt problem: This behavior of HTESSSEL with LDAS is rather surprising, and at first it might sound like a contradiction. How can the correct snow conditions lead to such poor river discharge in the ONLINE experiment? A possible explanation could be the representativeness issue of some of the snow observations, which can potentially cause local degradation in some of the catchments. It can also be explained by the HTESSSEL’s tendency to melt the snow too slowly (Dutra et al. 2012). In its simple, single layer snow scheme, too much snow accumulates into the snowpack and then that snow melts too slowly. For example, during a 20-mm mixed snow/rain forecast event (10mm liquid and 10mm solid) the snow scheme will accumulate most of the 10 mm solid (snow) part of the precipitation into the snowpack regardless of the temperature conditions and melt only a little of this 10mm. However, in reality a lot of that rain, sleet, or wet snow would not accumulate on the ground, and instead most

of it would melt straightaway. It seems the OFFLINE simulation gets the river discharge right mainly for the wrong reasons. Although the snowpack is clearly more poorly represented, the better timing with the delayed snowmelt (through the too slow melting) and the extra water in the snowpack, the OFFLINE experiment gets the runoff peak more correct.

- Summer: The water budget is balanced between precipitation and evapotranspiration with some soil water increments. During early summer water is taken out of the soil to cover the higher evapotranspiration. In OFFLINE more water leaves the soil which increases the runoff and also evapotranspiration. By August, however, the excess water from precipitation over evapotranspiration goes again into the soil, which is more pronounced in ONLINE where the soil is drier. The end of summer river discharge peak is present in both simulations, with the OFFLINE showing a better peak due to more water in the soil and subsequently higher surface and subsurface runoff during all summer. The OFFLINE river discharge exceeds the ONLINE values all summer and the two will level out by September, when the runoffs become similar in the two experiments.
- Autumn: From the middle of September there is another smaller snowmelt period starting with the falling temperatures and bringing some negative snow increments in the ONLINE simulation. The snow accumulates into the snowpack in both experiments, but again with a higher rate in OFFLINE, and also with larger snowmelt amounts in OFFLINE.

d. Regionally representative catchments

In the previous section the LDAS response was highlighted for an important weakness of HTESSEL with significant consequences on river discharge. In the following, the land–atmosphere coupling and LDAS impact is now demonstrated with a simplified representation of the annual water cycle in different geographical areas and also various climatic conditions for a selection of the world's catchments in Fig. 8. The displayed variables are simulated snowmelt, evapotranspiration, and river discharge in both the ONLINE and OFFLINE experiments, the snow and soil water increments for ONLINE, and finally the river discharge observations. All values are daily climatological mean values as in Fig. 7. The location of the catchments is provided in Fig. 6.

In Fig. 8, 12 catchments are selected to represent all main areas of the world where river discharge observations are available. Many of them are very large rivers, some of the catchments are dominated with mixed snow and soil moisture influence from the Northern Hemisphere, while others, mainly in the tropics, are only soil

moisture impacted. In Table 3, the main catchment details are provided (following the numbering from Fig. 6), complemented with the NSE and the percentage peak magnitude ME and MAE values for the catchments. Bold numbers denote the better score of ONLINE and OFFLINE.

Figure 8 suggests that the decreased snowmelt is a general feature in ONLINE across the Northern Hemisphere as predicted already by Fig. 5b. All displayed catchments have generally lower river discharge in ONLINE, either concentrated over the high river discharge season [e.g., Ob (1) and Yukon (2)], or elongated over most of the year [e.g., Danube (3) and Rhine (4)]. The snowmelt is universally smaller in the ONLINE simulation, with the LDAS removing snow at different periods of the year, which seems to be the driving force behind the river discharge differences.

The decreased amount of water has a mixed river discharge skill impact. For some catchments [Ob (1), Yukon (2), Columbia (6), and the case study catchment on the Amur (13)] the change during the high river discharge season is disadvantageous in ONLINE, confirmed by mostly negatively impacted scores, such as the NSE and the percentage peak magnitude MAE values in Table 3. On the other hand, for the Mississippi (5), Danube (3), and Rhine (4), it is rather beneficial as the daily climatological mean river discharge is closer to the corresponding observations during the high season, accompanied with mainly positive skill changes in the ONLINE experiment as both NSE and percentage peak magnitude MAE improves (Table 3), except the Rhine catchment (4), where the percentage peak magnitude MAE deteriorates.

In the warm climate, however, where soil water dominates the land surface processes [Xingu, Amazon, Hadejia, Ubangi, Zambesi, and Flinders (7–12)], the land–atmosphere coupling and LDAS impact on river discharge seems to be smaller than for the snow-influenced catchments, and on evapotranspiration it tends to be larger. There are large biases over five of the six highlighted tropical catchments (the only exception being the Flinders River in Australia), where both the ONLINE and OFFLINE experiments show significant mismatch with the observed values for the total river discharge volume and also for the annual peaks. For example, as displayed in Table 3, on the Hadejia River in Nigeria the percentage peak magnitude ME is 297% (the simulation is almost three times higher than the observation) in ONLINE, which is significantly better than OFFLINE (the improvement is 139% in the percentage peak magnitude MAE). This points to the fact that even though the river discharge differences are smaller in relative terms, it can still lead to noticeable change in the scores for some of these highlighted catchments (Table 3).

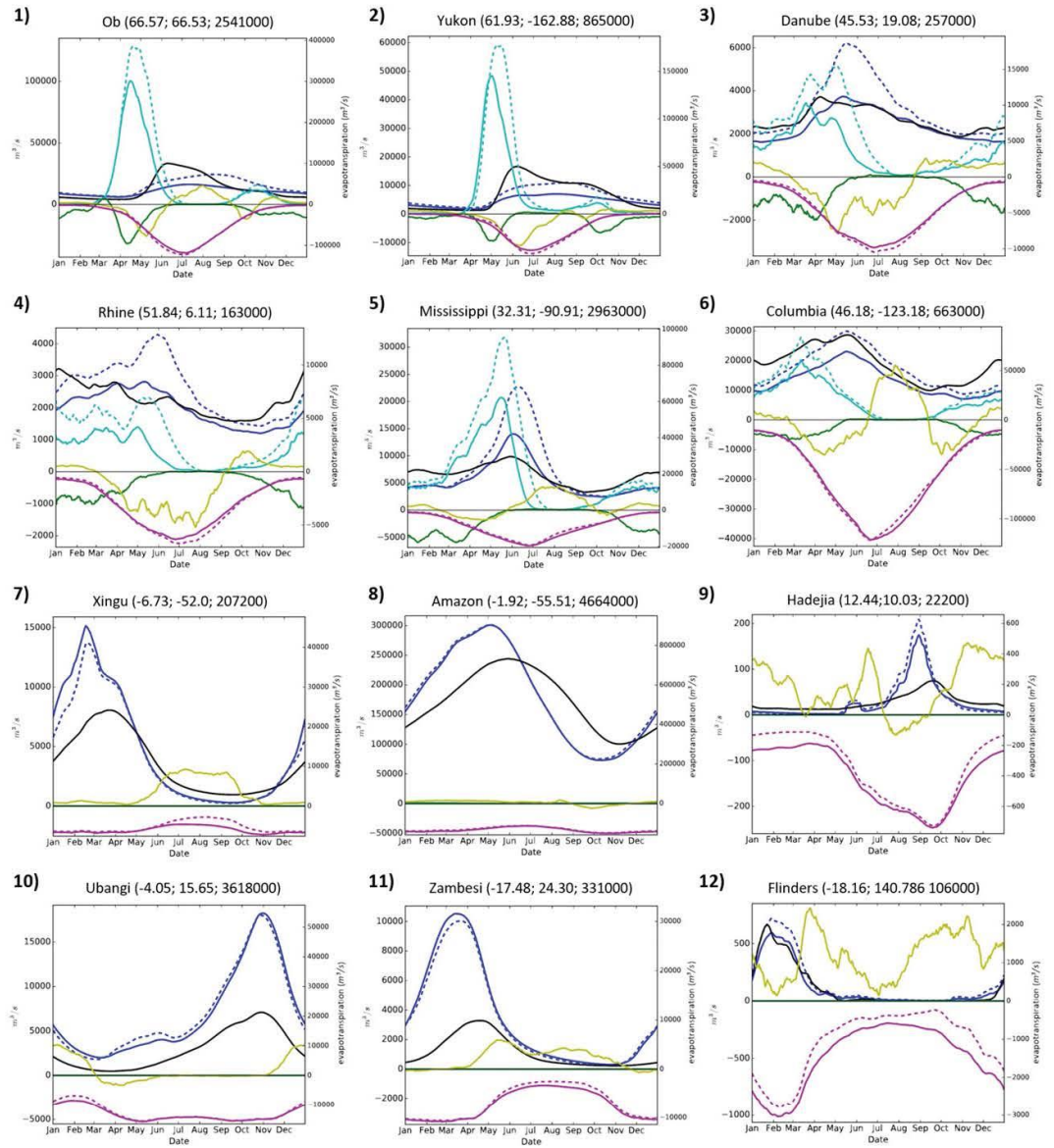


FIG. 8. The annual cycle of water budget variables for a selection of catchments worldwide numbered from 1 to 12 (see Fig. 6). The displayed variables are the snowmelt (cyan), evapotranspiration (magenta), and river discharge (blue) for both the ONLINE (solid lines) and OFFLINE (dashed lines) experiments; the snow (green) and soil (mustard) increments for ONLINE; and the river discharge observations (black line). All values are daily climatological averages based on the ERA5-D25 dataset (for details on the computation of these values, see section 21). The river names, the gauge coordinates, and the upstream area values are displayed in the subplot titles. The catchment descriptions with the main verification score values for the ONLINE and OFFLINE simulations are provided in Table 3. In addition, the catchment area contours are provided in Fig. 6. The evapotranspiration scale is provided on the secondary vertical axis, while the scale for all other parameters is shown on the main vertical axis.

TABLE 3. Details of the 13 catchments analyzed in Fig. 7 (13) and Fig. 8 (1–12) with the NSE, percentage peak magnitude ME (PPkMgMe) and percentage peak magnitude MAE (PPkMgMae) score values for the ONLINE and OFFLINE experiments based on the ERA5-D25 dataset. Bold scores denote better performance. For further details on the scores see section 2m.

Catchment No.	Station	River	Area ($\times 1000 \text{ km}^2$)	NSE		PPkMgMe (%)		PPkMgMae (%)	
				ONLINE	OFFLINE	ONLINE	OFFLINE	ONLINE	OFFLINE
1	Salekhard	Ob	2541	0.40	0.52	−55.0	−40.7	55.0	40.7
2	Pilot station	Yukon	865	0.31	0.64	−64.7	−50.7	64.7	50.7
3	Boogojevo	Danube	257	0.47	−0.43	−3.5	29.1	19.8	32.4
4	Lobith	Rhine	163	0.45	0.05	−39.1	−14.8	39.1	18.5
5	Viicksburg	Mississippi	2963	−0.02	−2.69	1.6	31.4	17.7	43.5
6	Quincy	Columbia	663	0.25	0.54	−24.0	−7.6	27.5	20.2
7	Boa Sorte	Xingu	207	−1.53	−0.85	159.0	147.9	159.0	147.9
8	Obidos-Linografo	Amazon	4664	−0.17	−0.21	26.6	26.9	26.6	26.9
9	Hadejia	Hadejia	22	−9.01	−11.85	297.1	436.1	297.1	436.1
10	Bangui	Ubangi	496	−5.72	−6.17	162.8	159.1	162.8	159.1
11	Katima Mulilo	Zambesi	331	−7.97	−6.70	196.6	183.0	196.6	183.0
12	Walkers bend	Flinders	106	0.66	0.62	−24.5	−11.4	46.9	45.9
13	Komsomolsk	Amur	1846	0.43	0.68	−33.5	−18.7	33.5	18.7

Even though there is no clear systematic difference between the exclusively soil moisture and the mixed (snow and soil moisture) catchments in terms of river discharge skill impact, the snow clearly looks to carry a more direct influence on the river discharge volume and also on the river discharge skill.

e. Global river discharge analysis

In the previous sections it could be shown that the water budget is out of balance in the ONLINE simulation over large parts of the world leading to significant impact on the river discharge for the analyzed list of catchments. As an extreme example, it was demonstrated that the snowmelt-driven spring river discharge peak was almost completely missed in a large catchment in east Russia in ONLINE. After the individual catchment examples, a systematic analysis of the river discharge quality in the ONLINE and OFFLINE experiments is provided based on all available catchments globally.

Although a large number of scores was computed in this study, this section will focus only on the annual peak flow scores. The timing and magnitude of the high river discharges are both crucial aspects of river discharge simulations in any flood prediction system such as GloFAS. The accurate simulation of the river discharge peaks is essential to get the best possible guidance for the potentially most damaging floods. The analyzed performance of the annual peak river flows should give a good indication on the general ability of the two experiments to predict peaks.

Figure 9a highlights a large systematic percentage peak magnitude ME in the ONLINE simulation. Many catchments show over 50% error (either positive or negative) of the annual river discharge peaks on average.

The majority of the Northern Hemispheric higher latitudes is overwhelmingly underpredicted, while Amazonia, the western United States, and many catchments in Africa are overpredicted in the ONLINE experiment. The geographical pattern in Fig. 9a is rather similar to the one seen in Fig. 5a. Most of the catchments with significant negative values over the Northern Hemisphere and positive ones mainly in lower latitudes do resemble well the water budget error pattern seen in Fig. 5a.

The water budget imbalance, caused by the increments in LDAS, is only one of the many potential contributing factors to peak river flow errors (and in fact to general river discharge errors); atmospheric forcing biases, imperfect river routing, and observation errors could also lead to large inaccuracies (Zhao et al. 2017).

The impact of the land–atmosphere coupling and LDAS seems to decrease the amount of water overwhelmingly in the rivers (decreased sample mean river discharge, not shown). The sample average river discharge increased only in the southern half of Brazil, in the central part of Canada, and one or two catchments in Africa, East Asia, and South Australia (not shown). It is expected that the decreased average river discharge in ONLINE should generally also result in lower annual peak river flows over most of the globe. Figure 9b shows that this decreasing tendency of the annual peaks in the ONLINE experiment coincides with widespread, quite large deterioration in the percentage peak magnitude MAE score (increase of the annual peak magnitude errors) especially in Asia and Europe and the northwestern part of North America, where the majority of the catchments show significant negative bias in Fig. 9a. On the other hand, quite a few catchments seem to benefit from the coupling and LDAS as the annual peak errors decrease, especially in the western parts

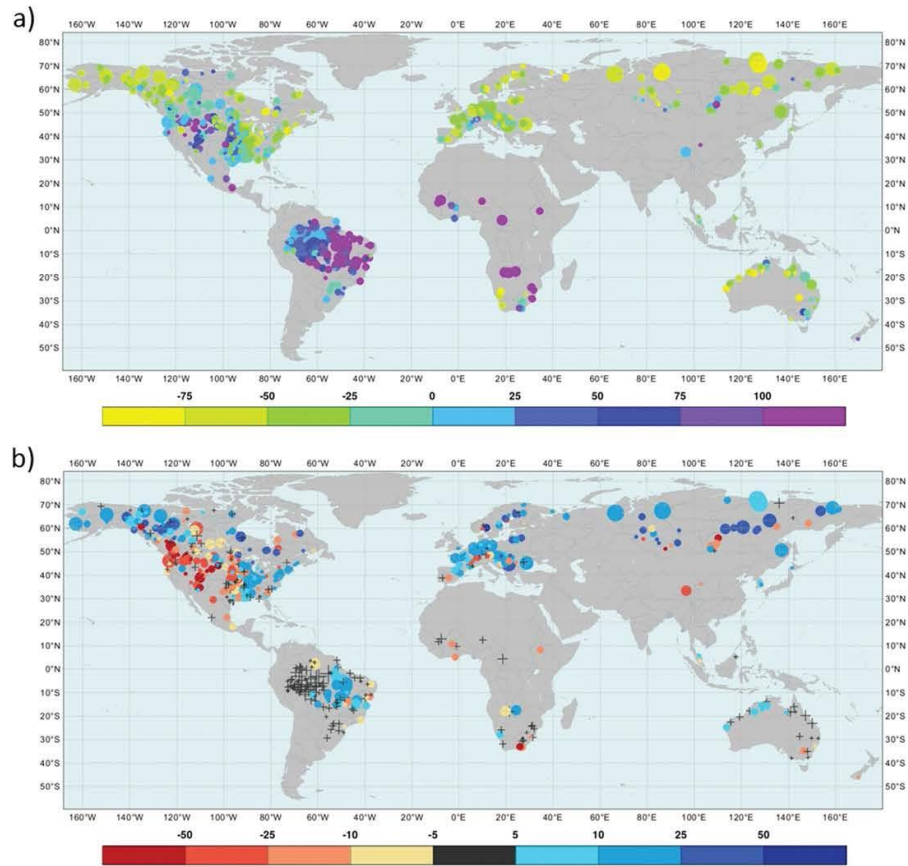


FIG. 9. River discharge percentage peak magnitude (a) ME (%) of the ONLINE experiment and (b) change in the percentage peak magnitude MAE (%) between ONLINE and OFFLINE based on the ERA5-D25 dataset. Positive error differences in (b) indicate deterioration (blue) while negative changes show improvement (red) in the ONLINE simulation compared with OFFLINE. The catchments are displayed with different marker sizes representing the size of the catchment area. Near-zero differences are shown by black crosses, while all other categories are displayed by circles.

in North America, where there is a large cluster of catchments with noticeably smaller percentage peak magnitude MAE.

The river discharge peak timing bias in the ONLINE simulation is dominantly positive (peaks are too late) in the Northern Hemisphere and mainly negative (peaks too early) in the tropics (not shown). However, the coupling and LDAS do not seem to have any systematic impact on this aspect of the peak river flows. There are noticeable differences, but they have no distinguishable geographical pattern (not shown). It seems the short time series (9–25 annual values only) were not sufficient to extract any representative timing differences between the two experiments.

In addition to the analysis of the annual river discharge peak performance, the general fit between modeled and observed daily river discharge time series is also extensively measured by several scores. Table 4 shows a global summary giving an indication on the overall performance of the two experiments. The scores are calculated as global averages weighted by the square root of the catchment area size. This way a more representative picture can be provided by giving more emphasis on the larger catchments.

The generally decreasing amount of water leads to larger differences for most of the volume-related bias scores. The percentage sample ME, the percentage sample standard deviation error, and the percentage

TABLE 4. List of global average scores for the ONLINE and OFFLINE experiments based on the ERA5-D25 dataset. Each value is a mean of scores from 590 catchments (where a minimum of 9 years of river discharge observations was available) weighted by the square root of the catchment area sizes. For further details on the scores, see section 2m. Bold numbers denote the better score of ONLINE and OFFLINE. The following scores are displayed: ME, MAE, NSE, R, percentage sample mean error (PMnE), percentage sample mean absolute error (PMnAe), percentage sample standard deviation error (PSIE), percentage sample standard deviation absolute error (PSIAe), peak timing ME (PKTME), peak timing MAE (PKTMae), percentage peak magnitude ME (PPKMgMe), and percentage peak magnitude MAE (PPKMgMae).

Score	ME ($\text{m}^3 \text{s}^{-1}$)	MAE ($\text{m}^3 \text{s}^{-1}$)	NSE	R	PMnE (%)	PMnAe (%)	PSIE (%)	PSIAe (%)	PKTME (day)	PKTMae (day)	PPKMgMe (%)	PPKMgMae (%)
ONLINE	-264	3017	-0.29	0.67	-2.6	29.0	9.6	48.3	-0.95	11.8	6.3	61.3
OFFLINE	236	2954	-0.53	0.70	16.9	27.2	34.2	52.1	-0.81	11.8	27.3	59.2

peak magnitude ME scores all decrease significantly in the ONLINE simulation, bringing the global biases closer to zero. The only exception is the discharge ME score, which changes from a positive value to a negative one with similar magnitude. The better biases, however, do not necessarily help improve the river discharge skill globally; the scores presented in Table 4 provide a mixed picture, with some favoring the ONLINE while others favoring the OFFLINE simulation. This agrees with the mixed scores shown in Table 3 for the regional example catchments. In general, the MAE, R, the percentage sample MAE, and the percentage peak magnitude MAE values are all slightly better for OFFLINE, while the NSE and percentage sample standard deviation absolute error show improvement for ONLINE. And finally, the peak timing ME is slightly better for the OFFLINE experiment, while there is no difference in the global average peak timing MAE.

4. Discussion

In section 3, the land-atmosphere coupling and LDAS impact on hydrology, including river discharge and the related water budget variables, was analyzed. The river discharge scores showed a mixed picture between the ONLINE and OFFLINE simulations with relatively similar global performance. Larger differences could be highlighted in certain regions, such as many of the snow-dominant catchments in the Northern Hemisphere, where over many areas a large amount of water is missing from the hydrological cycle and causing downstream issues in river discharge especially during the snowmelt season in ONLINE.

The general decrease in the volume of water in the ONLINE experiment, mainly coming from the snow-dominated areas where the assimilation removes snow, seems to be the primary impact on the hydrology. In soil moisture-dominated areas the river discharge seems to be less impacted by the increments and the evapotranspiration rate holds a more important role.

Data assimilation is a very important component of any NWP system with a lot of effort and research concentrated on the use of observations to correct for random (day-to-day) errors. Data assimilation systems are not there to correct for systematic biases. The fact that LDAS produces consistent negative increments in snow covered areas in this study is pointing toward an apparent snow model bias. In contrast, a model affected by random errors only, would lead to data assimilation increments of both signs with close to zero annual mean values.

Other studies have also highlighted significant snow assimilation impacts on the water balance. For example, De Lannoy et al. (2012) showed that on a small catchment

in Colorado (United States) the season averaged snowpack water content is largely decreased by the snow water equivalent assimilation in the Noah land surface model, and could only be overcome by scaling applied (to anomalies) to the observations prior to assimilation. Similarly, [Arsenault et al. \(2013\)](#) found that assimilating MODIS snow cover fraction observations into the CLM land surface model by a simple rule-based direct insertion and the one-dimensional ensemble Kalman filter methods, lead to substantial snowpack removal (without melting, thus causing negative bias in runoff), by both methods in Colorado and Washington.

In the ECMWF system, the snow increments are correcting for the systematic overestimation of the current HTESEL snow scheme that melts the snow too slowly. [Dutra et al. \(2012\)](#) highlighted that although the current snow scheme provides a significant improvement over the previous one, it does not yet improve on the short-duration melting events during late winter and spring. They argued that the experimental multilayer snow scheme was able to reproduce, at least partially, those snowmelt episodes thanks to the top snow layer having a reduced thermal inertia.

The findings in this work are specific to the NWP configuration at ECMWF with the HTESEL land surface model and the processes within. However, any LSM's ability to support hydrological simulations can be limited by inadequate handling of the processes, potentially causing a similar problem downstream in the hydrology. The areas highlighted here for ECMWF's HTESEL in supporting the flood forecasting activities can be improved by some potential developments in the future. Some of the areas where substantial improvements could be achieved are described:

- A new multilayer snow scheme is currently being tested at ECMWF, which is similar to the one evaluated in [Dutra et al. \(2012\)](#). This improved snow scheme is expected to represent better the snowmelt processes and therefore reduce the snow increments that currently remove a significant amount of water from the hydrological cycle. The hydrological context developed in this study will be used to aid this development of the new scheme.
- Another potential way of improving HTESEL performance for hydrological applications would be to modify the LDAS by special handling of the snow increments in order to retain the water in the hydrological cycle during the data assimilation. For example, [Zaitchik and Rodell \(2009\)](#) proposed an interesting approach using near-future, snow-covered area observations to adjust the air temperature and precipitation forcing data in order to preserve the local hydrological

balance. In another study, [Pan and Wood \(2006\)](#) developed a constrained ensemble Kalman filter method to assure closure of the water balance when assimilating hydrological observations. These types of studies rely on uncoupled systems, and they would be difficult to implement in an operational, real-time environment. However, they provide some insight on water budget closure in data assimilation, and they should be further investigated and adapted to coupled land-atmosphere NWP systems. In the longer term, further coupling between NWP and hydrological forecasting systems will be considered, thereby opening the possibility for coupled land-hydrology data assimilation. In this context, joint assimilation of land surface and river discharge observations will consistently correct the different components of the Earth system.

- In addition, the land surface development methodology including data assimilation techniques and process representation is continuously improved at ECMWF. The future inclusion of the LDAS scheme in the offline HTESEL is in development. It will create an environment where the offline research work, including the reanalysis improvements (e.g., ERA5), could be done in a consistent way with the real-time forecast generation. In parallel to these developments, addressing the water budget closure in land-atmosphere data assimilation systems should be a priority in the future to ensure consistent high-quality coupled NWP and hydrological forecasts.

GloFAS is one of the few existing flood forecasting systems that utilizes an LSM (HTESEL) for representing the hydrology ([Emerton et al. 2016](#)). Although we acknowledge that in some cases a simple routing model, initialized from observed upstream river levels (either from river gauges or satellite measurements), could be a simpler alternative to simulate downstream discharge on large rivers a few days in advance, for example, in [Hossain et al. \(2014\)](#); in other cases where forecasts are required further in advance or where observations are unavailable or of too low quality, a more complex modeling configuration, which represents hydrological fluxes, becomes essential. Regardless of some limitations (e.g., the one highlighted in the ECMWF NWP configuration), these complex models play crucial roles in harnessing the available predictability in the land-atmosphere system.

5. Conclusions

Understanding the impacts of both the data assimilation and land surface process representation in land

surface models on simulated hydrological variables is very important, not only for improving the weather and climate forecasts, but specifically for supporting flood forecasting and other hydrological applications such as drought forecasting, and also for giving feedback about the Earth system. In this paper, the influence of land–atmosphere coupling and land data assimilation on global hydrological simulations from LSMs was evaluated. Two river discharge simulations from two climatological reanalyses (based on ERA5) were compared: one operational set which includes land–atmosphere coupling and LDAS with an open water budget, and also an offline HTESSEL set with a closed water budget and no LDAS.

It was found that while the ONLINE version of the model largely improves the 2-m temperature and snow depth conditions, it is causing poor representation of peak river flow in snowmelt-dominated areas, particularly in the high latitudes. However, there are also localized improvements to peak river flow, such as in the western United States. The LDAS increments remove or add water even on an annual average scale which inevitably leads to systematic water budget errors and subsequently contribute to significant errors in river discharge during times of peak flow downstream, something that is critical during times of flooding.

a. Implications for hydrological forecasting

This study has highlighted the impact of using land data assimilation in reanalysis products. Where data assimilation is adjusting snowpack in forecasting mode then there will also be important implications for hydrological predictions. Future studies should address how far ahead the impact of data assimilation propagates in hydrological forecasts. In addition, hydrological forecasting systems often use initial river conditions derived from climatology. In these circumstances using climatological products derived using data assimilation methodologies could lead to issues with the hydrological forecasts. There are also related issues for forecasting systems such as GloFAS that compare model output to climatology to provide early awareness of extreme events—consistency between operational and climatological configurations goes some way to bypass this problem, and this conclusion has directly influenced the design of the new GloFAS-seasonal system (Emerton et al. 2018).

b. Implications for land surface modeling and data assimilation

Data assimilation is designed to compensate for noise errors and not systematic bias. In the case of the current HTESSEL snow assimilation scheme it is doing

the latter—compensating for system deficiencies such as the slow snowmelt process. This paper has discussed potential ways of addressing water budget deficiencies in land surface approaches, for example, including multiple layers within the HTESSEL snow scheme or moving toward data assimilation that conserves the water budget.

Without addressing such issues there will never be confidence in using LSMs for hydrological forecasting applications across the globe. This type of analysis should be used to diagnose where improvements need to be made; considering the whole Earth system in data assimilation and coupling developments is critical for moving toward the goal of holistic Earth system approaches.

Acknowledgments. Ervin Zsoter's PhD is supported by the Wilkie Calvert Co-Supported PhD Studentships at the University of Reading. Ervin Zsoter was supported by the Copernicus Emergency Management Service - Early Warning Systems [CEMS-EWS (EFAS)]. Hannah Cloke is supported by the TENDERLY project: Towards END-to End flood forecasting and a tool for Real-time catchment susceptibility U.K. NERC Flooding From Intense Rainfall (FFIR) programme, NE/K00896X/1. Elisabeth Stephens and Hannah Cloke are supported by the FATHUM project: Forecasts for Anticipatory HUMANitarian Action funded by U.K. NERC as part of their Science for Humanitarian Emergencies and Resilience (SHEAR) programme, NE/P000525/1. We are also grateful to The Global Runoff Data Centre, 56068 Koblenz, Germany, for providing the observation dataset for our river discharge analysis. Ervin Zsoter designed the experiment, produced the ERA5-D25 datasets, carried out the river discharge data analysis, and led the writing of the manuscript. Hannah Cloke and Liz Stephens assisted with posing the research question and designing the analysis, Patricia de Rosnay and Joaquin Muñoz-Sabater helped with the scientific analysis of data assimilation and coupling issues, and Christel Prudhomme and Florian Pappenberger helped design the research methodology. All authors assisted with writing the manuscript.

REFERENCES

- Agusti-Panareda, A., G. Balsamo, and A. Beljaars, 2010: Impact of improved soil moisture on the ECMWF precipitation forecast in West Africa. *Geophys. Res. Lett.*, **37**, L20808, <https://doi.org/10.1029/2010GL044748>.
- Albergel, C., P. de Rosnay, G. Balsamo, L. Isaksen, and J. Muñoz Sabater, 2012: Soil moisture analyses at ECMWF: Evaluation using global ground-based in situ observations. *J. Hydrometeorol.*, **13**, 1442–1460, <https://doi.org/10.1175/JHM-D-11-0107.1>.

- 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>.
- Andreadis, K., and D. Lettenmaier, 2006: Assimilating remotely sensed snow observations into a macroscale hydrology model. *Adv. Water Resour.*, **29**, 872–886, <https://doi.org/10.1016/j.advwatres.2005.08.004>.
- Arsenault, K. R., P. R. Houser, G. J. M. De Lannoy, and P. A. Dirmeyer, 2013: Impacts of snow cover fraction data assimilation on modeled energy and moisture budgets. *J. Geophys. Res. Atmos.*, **118**, 7489–7504, <https://doi.org/10.1002/jgrd.50542>.
- 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. Hydrometeorol.*, **10**, 623–643, <https://doi.org/10.1175/2008JHM1068.1>.
- , S. Boussetta, P. Lopez, and L. Ferranti, 2010: Evaluation of ERA-Interim and ERA-Interim-GPCP-rescaled precipitation over the U.S.A. ERA Rep. Series 5, ECMWF, 10 pp., <https://www.ecmwf.int/node/7926>.
- , 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>.
- , and Coauthors, 2015: ERA-Interim/Land: A global land surface reanalysis data set. *Hydrol. Earth Syst. Sci.*, **19**, 389–407, <https://doi.org/10.5194/hess-19-389-2015>.
- Bélair, S., L.-P. Crevier, J. Mailhot, B. Bilodeau, and Y. Delage, 2003: Operational implementation of the ISBA land surface scheme in the Canadian Regional Weather Forecast Model. Part I: Warm season results. *J. Hydrometeorol.*, **4**, 352–370, [https://doi.org/10.1175/1525-7541\(2003\)4<352:OIOTIL>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)4<352:OIOTIL>2.0.CO;2).
- Beljaars, A. C. M., P. Viterbo, M. Miller, and A. K. Betts, 1996: The anomalous rainfall over the United States during July 1993: Sensitivity to land surface parameterization and soil anomalies. *Mon. Wea. Rev.*, **124**, 362–383, [https://doi.org/10.1175/1520-0493\(1996\)124<0362:TAROTU>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<0362:TAROTU>2.0.CO;2).
- Blyth, E., D. B. Clark, R. Ellis, C. Huntingford, S. Los, M. Pryor, M. Best, and S. Sitch, 2011: A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale. *Geosci. Model Dev.*, **4**, 255–269, <https://doi.org/10.5194/gmd-4-255-2011>.
- Brasnett, B., 1999: A global analysis of snow depth for numerical weather prediction. *J. Appl. Meteor.*, **38**, 726–740, [https://doi.org/10.1175/1520-0450\(1999\)038<0726:AGAOSD>2.0.CO;2](https://doi.org/10.1175/1520-0450(1999)038<0726:AGAOSD>2.0.CO;2).
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- De Lannoy, G. J. M., R. H. Reichle, K. R. Arsenault, P. R. Houser, S. Kumar, N. E. C. Verhoest, and V. R. N. Pauwels, 2012: Multi-scale assimilation of Advanced Microwave Scanning Radiometer-EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado. *Water Resour. Res.*, **48**, W01522, <https://doi.org/10.1029/2011WR010588>.
- de Rosnay, P., M. Drusch, D. Vasiljevic, G. Balsamo, C. Albergel, and L. Isaksen, 2013: A simplified extended Kalman filter for the global operational soil moisture analysis at ECMWF. *Quart. J. Roy. Meteor. Soc.*, **139**, 1199–1213, <https://doi.org/10.1002/qj.2023>.
- , G. Balsamo, C. Albergel, J. Muñoz-Sabater, and L. Isaksen, 2014: Initialisation of land surface variables for Numerical Weather Prediction. *Surv. Geophys.*, **35**, 607–621, <https://doi.org/10.1007/s10712-012-9207-x>.
- , L. Isaksen, M. Dahoui, 2015: Snow data assimilation at ECMWF. ECMWF Newsletter, No. 143, ECMWF, Reading, United Kingdom, 26–31, <https://doi.org/10.21957/lkpxq6x5>.
- Drusch, M., and P. Viterbo, 2007: Assimilation of screen-level variables in ECMWF's Integrated Forecast System: A study on the impact on the forecast quality and analyzed soil moisture. *Mon. Wea. Rev.*, **135**, 300–314, <https://doi.org/10.1175/MWR3309.1>.
- , D. Vasiljevic, and P. Viterbo, 2004: ECMWF's global snow analysis: Assessment and revision based on satellite observations. *J. Appl. Meteor.*, **43**, 1282–1294, [https://doi.org/10.1175/1520-0450\(2004\)043<1282:EGSAAA>2.0.CO;2](https://doi.org/10.1175/1520-0450(2004)043<1282:EGSAAA>2.0.CO;2).
- Dutra, E., G. Balsamo, P. Viterbo, P. Miranda, A. Beljaars, C. Schär, and K. Elder, 2010: An improved snow scheme for the ECMWF land surface model: Description and offline validation. *J. Hydrometeorol.*, **11**, 899–916, <https://doi.org/10.1175/2010JHM1249.1>.
- , C. Schär, P. Viterbo, and P. M. A. Miranda, 2011: Land-atmosphere coupling associated with snow cover. *Geophys. Res. Lett.*, **38**, L15707, <https://doi.org/10.1029/2011GL048435>.
- , P. Viterbo, P. M. A. Miranda, and G. Balsamo, 2012: Complexity of snow schemes in a climate model and its impact on surface energy and hydrology. *J. Hydrometeorol.*, **13**, 521–538, <https://doi.org/10.1175/JHM-D-11-072.1>.
- Emerton, R., and Coauthors, 2016: Continental and global scale flood forecasting systems. *Wiley Interdiscip. Rev.: Water*, **3**, 391–418, <https://doi.org/10.1002/wat2.1137>.
- , H. Cloke, E. Stephens, E. Zsoter, S. Woolnough, and F. Pappenberger, 2017: Complex picture for likelihood of ENSO-driven flood hazard. *Nat. Commun.*, **8**, 14796, <https://doi.org/10.1038/ncomms14796>.
- , and Coauthors, 2018: Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0. *Geosci. Model Dev.*, **11**, 3327–3346, <https://doi.org/10.5194/gmd-11-3327-2018>.
- Haddeland, I., and Coauthors, 2011: Multimodel estimate of the global terrestrial water balance: Setup and first results. *J. Hydrometeorol.*, **12**, 869–884, <https://doi.org/10.1175/2011JHM1324.1>.
- Helfrich, S. R., D. McNamara, B. Ramsay, T. Baldwin, and T. Kasheta, 2007: Enhancements to, and forthcoming developments in the interactive multisensor snow and ice mapping system, (IMS). *Hydrol. Processes*, **21**, 1576–1586, <https://doi.org/10.1002/hyp.6720>.
- Helmert, J., and Coauthors, 2018: Review of snow data assimilation methods for hydrological, land surface, meteorological and climate models: Results from a COST HarmoSnow survey. *Geosciences*, **8**, 489, <https://doi.org/10.3390/geosciences8120489>.
- Hersbach, H., and D. Dee, 2016: ERA5 reanalysis is in production. *ECMWF Newsletter*, No. 147, ECMWF, Reading, United Kingdom, 7, <https://www.ecmwf.int/sites/default/files/elibrary/2016/16299-newsletter-no147-spring-2016.pdf>.
- Hossain, F., A. H. Siddique-E-Akbor, S. Biancamaria, H. Lee, and C. K. Shum, 2014: Proof of concept of an altimeter-based river forecasting system for transboundary flow inside Bangladesh.

- IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, **7**, 587–601, <https://doi.org/10.1109/JSTARS.2013.2283402>.
- Huffman, G. J., R. F. Adler, D. T. Bolvin, and G. Gu, 2009: Improving the global precipitation record: GPCP version 2.1. *Geophys. Res. Lett.*, **36**, <https://doi.org/10.1029/2009GL040000>.
- Kauffeldt, A., S. Halldin, F. Pappenberger, F. Wetterhall, C.-Y. Xu, and H. L. Cloke, 2015: Imbalanced land surface water budgets in a numerical weather prediction system. *Geophys. Res. Lett.*, **42**, 4411–4417, <https://doi.org/10.1002/2015GL064230>.
- Legates, D. R., and G. J. McCabe Jr., 1999: Evaluating the use of “goodness-of-fit” Measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.*, **35**, 233–241, <https://doi.org/10.1029/1998WR900018>.
- Le Vine, N., A. Butler, N. McIntyre, and C. Jackson, 2016: Diagnosing hydrological limitations of a land surface model: Application of JULES to a deep-groundwater chalk basin. *Hydrol. Earth Syst. Sci.*, **20**, 143–159, <https://doi.org/10.5194/hess-20-143-2016>.
- Mahfouf J-F., P. Viterbo, H. Douville, A.C.M. Beljaars, and S. Saarinen, 2000: A revised land-surface analysis scheme in the Integrated Forecasting System. *ECMWF Newsletter*, No. 88, ECMWF, Reading, United Kingdom, 8–13, <https://www.ecmwf.int/node/14636>.
- Mengelkamp, H.-T., K. Warrach, C. Ruhe, and E. Raschke, 2001: Simulation of runoff and streamflow on local and regional scales. *Meteor. Atmos. Phys.*, **76**, 107–117, <https://doi.org/10.1007/s007030170042>.
- Molotch, N. P., and R. C. Bales, 2005: Scaling snow observations from the point to the grid element: Implications for observation network design. *Water Resour. Res.*, **41**, W11421, <https://doi.org/10.1029/2005WR004229>.
- Nash, J. E., and J. V. Sutcliffe, 1970: River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.*, **10**, 282–290, [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6).
- Overgaard, J., D. Rosbjerg, and M. B. Butts, 2006: Land-surface modelling in hydrological perspective—A review. *Biogeosciences*, **3**, 229–241, <https://doi.org/10.5194/bg-3-229-2006>.
- Pan, M., and E. Wood, 2006: Data assimilation for estimating the terrestrial water budget using a constrained ensemble Kalman filter. *J. Hydrometeorol.*, **7**, 534–547, <https://doi.org/10.1175/JHM495.1>.
- Pearson, K., 1896: Mathematical contributions to the theory of evolution: III. Regression, heredity and panmixia. *Philos. Trans. Roy. Soc. London*, **187A**, 253–318, <https://doi.org/10.1098/rsta.1896.0007>.
- Rabier, F., 2005: Overview of global data assimilation developments in numerical weather prediction centres. *Quart. J. Roy. Meteor. Soc.*, **131**, 3215–3233, <https://doi.org/10.1256/qj.05.129>.
- , H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, **126**, 1143–1170, <https://doi.org/10.1002/qj.49712656415>.
- Slater, A. G., and M. P. Clark, 2006: Snow data assimilation via an ensemble Kalman filter. *J. Hydrometeorol.*, **7**, 478–493, <https://doi.org/10.1175/JHM505.1>.
- Stephens, E., J. J. Day, F. Pappenberger, and H. Cloke, 2015: Precipitation and floodiness. *Geophys. Res. Lett.*, **42**, 10316–10323, <https://doi.org/10.1002/2015GL066779>.
- van der Knijff, J. M., J. Younis, and A. P. J. de Roo, 2010: LIS-FLOOD: A GIS-based distributed model for river basin scale water balance and flood simulation. *Int. J. Geogr. Inf. Sci.*, **24**, 189–212, <https://doi.org/10.1080/13658810802549154>.
- Wang, L. L., D. H. Chen, and H. J. Bao, 2016: The improved Noah land surface model based on storage capacity curve and Muskingum method and application in GRAPES model. *Atmos. Sci. Lett.*, **17**, 190–198, <https://doi.org/10.1002/asl.642>.
- Wu, H., R. F. Adler, Y. Tian, G. J. Huffman, H. Li, and J. Wang, 2014: Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model. *Water Resour. Res.*, **50**, 2693–2717, <https://doi.org/10.1002/2013WR014710>.
- Yamazaki, D., S. Kanae, H. Kim, and T. Oki, 2011: A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resour. Res.*, **47**, W04501, <https://doi.org/10.1029/2010WR009726>.
- Zaitchik, B. F., and M. Rodell, 2009: Forward-looking assimilation of MODIS-derived snow-covered area into a land surface model. *J. Hydrometeorol.*, **10**, 130–148, <https://doi.org/10.1175/2008JHM1042.1>.
- Zhao, F., and Coauthors, 2017: The critical role of the routing scheme in simulating peak river discharge in global hydrological models. *Environ. Res. Lett.*, **12**, 075003, <https://doi.org/10.1088/1748-9326/aa7250>.
- Zsoter, E., and Coauthors, 2016: Building a multi-model flood prediction system with the TIGGE archive. *J. Hydrometeorol.*, **17**, 2923–2940, <https://doi.org/10.1175/JHM-D-15-0130.1>.



A2: Hydrological impact of the new ECMWF multi-layer snow scheme

This paper presents the published version of chapter 5 of this thesis, with the following reference:

Zsoter, E., G. Arduini, C. Prudhomme, E. Stephens, H. Cloke, 2022: Hydrological Impact of the New ECMWF Multi-Layer Snow Scheme, *Atmosphere*, 13(5):727, doi:10.3390/atmos13050727

Article

Hydrological Impact of the New ECMWF Multi-Layer Snow Scheme

 Ervin Zsoter ^{1,2,*} , Gabriele Arduini ¹ , Christel Prudhomme ¹, Elisabeth Stephens ^{3,4} and Hannah Cloke ^{2,3,5,6}
¹ European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, UK; gabriele.arduini@ecmwf.int (G.A.); christel.prudhomme@ecmwf.int (C.P.)

² Department of Geography and Environmental Science, University of Reading, Reading RG6 6AB, UK; h.l.cloke@reading.ac.uk

³ Department of Meteorology, University of Reading, Reading RG6 6BB, UK; elisabeth.stephens@reading.ac.uk

⁴ Red Cross Red Crescent Climate Centre, 2593 HT The Hague, The Netherlands

⁵ Department of Earth Sciences, Uppsala University, 752 36 Uppsala, Sweden

⁶ Centre of Natural Hazards and Disaster Science (CNDS), 752 36 Uppsala, Sweden

* Correspondence: ervin.zsoter@ecmwf.int

Abstract: The representation of snow is a crucial aspect of land-surface modelling, as it has a strong influence on energy and water balances. Snow schemes with multiple layers have been shown to better describe the snowpack evolution and bring improvements to soil freezing and some hydrological processes. In this paper, the wider hydrological impact of the multi-layer snow scheme, implemented in the ECLand model, was analyzed globally on hundreds of catchments. ERA5-forced reanalysis simulations of ECLand were coupled to CaMa-Flood, as the hydrodynamic model to produce river discharge. Different sensitivity experiments were conducted to evaluate the impact of the ECLand snow and soil freezing scheme changes on the terrestrial hydrological processes, with particular focus on permafrost. It was found that the default multi-layer snow scheme can generally improve the river discharge simulation, with the exception of permafrost catchments, where snowmelt-driven floods are largely underestimated, due to the lack of surface runoff. It was also found that appropriate changes in the snow vertical discretization, destructive metamorphism, snow-soil thermal conductivity and soil freeze temperature could lead to large river discharge improvements in permafrost by adjusting the evolution of soil temperature, infiltration and the partitioning between surface and subsurface runoff.

Keywords: land-surface modelling; snow scheme; hydrological processes; river discharge; surface runoff; permafrost



Citation: Zsoter, E.; Arduini, G.; Prudhomme, C.; Stephens, E.; Cloke, H. Hydrological Impact of the New ECMWF Multi-Layer Snow Scheme. *Atmosphere* **2022**, *13*, 727. <https://doi.org/10.3390/atmos13050727>

Academic Editor: Stephan De Wekker

Received: 8 March 2022

Accepted: 28 April 2022

Published: 2 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Land-surface models (LSMs) are vital tools for simulating water and energy fluxes at the land–atmosphere interface of the Earth [1]. Although LSMs were originally designed to provide lower-boundary conditions to the atmosphere [2], with the improving realism of these models they are increasingly used for simulating the hydrological cycle [3] and supporting hydrological applications (e.g., [4–6]).

However, there are still significant limitations in the representation of the hydrological cycle in LSMs, as important processes can still be inadequately modelled or neglected for runoff generation, for instance groundwater simulation, snow-vegetation interactions, representation of frozen soil and lateral flow between adjacent grid cells among others [1,5,7–11].

Simulating the extent and variability of the snow cover is a crucial aspect of land-surface modelling, as it strongly influences the energy and water balances [12,13]. Snow schemes have various complexities in the representation of the snow physics [14–18],

differing largely in their handling of the snowpack and the creation of snowmelt, which in turn impacts runoff generation and river flow in snow dominated areas [8,19].

The snow scheme currently used operationally in the ECLand land-surface model at the European Centre for Medium-Range Weather Forecasts (ECMWF) [20,21] and which is used in the production of the ERA5 [22] and ERA5-Land [23] reanalysis datasets, is a single-layer snow scheme (SLS hereafter) with an additional snow layer on top of the soil [17]. The use of only one layer limits the handling of the temporal evolution of the snow, as changes on multiple time scales (i.e., diurnal to seasonal) cannot be accurately represented. This has a significant impact on the quality of the snow depth, mainly during periods of accumulation and ablation, which then impacts the soil freezing, the snowmelt and ultimately the hydrological cycle [8,24].

Snow schemes using multiple layers to represent the snowpack offer significant improvement on the single-layer schemes with better handling of the snow processes. For hydrological and climate applications, so-called “intermediate complexity” snow schemes are generally used, following the terminology introduced by Boone and Etchevers [14]. Such schemes include a description of snowpack properties, such as density and temperature, using a limited number of vertical layers, as opposed to detailed snow physics models, which aim at simulating the microstructure properties of the snowpack as well (see for instance Vionnet et al. [25]). Such intermediate complexity schemes have been implemented in various LSMs used in Earth system modelling during the last decade. Examples are ECLand [8,26], Noah [24], JULES [27], ISBA [14,16] and the ORCHIDEE [18] land-surface models.

The multi-layer snow scheme (MLS hereafter), introduced experimentally in ECLand, is an intermediate complexity scheme representing the vertical structure and evolution of snow temperature, density, liquid water content and surface snow albedo with a maximum of five layers [26]. It has been shown to increase the realism of the snow representation, including decreasing snow depth and snowmelt timing errors, and has been shown to largely improve 2-metre temperature in coupled forecasts, especially in clear sky conditions [26].

The more realistic snowpack representation, with better snow water equivalent, snow depth and snowmelt, is expected to improve the hydrological cycle and thus have a positive impact on river flow simulations [28]. Preliminary studies have demonstrated improvements in localized settings [8,18] or combined with other land-surface improvements [29]. However, the hydrological impact analysis of MLSs, focusing on river flow, has not been done at regional or global scales.

Areas in which snow plays an important role are predominantly found in the Northern Hemisphere over higher latitudes. A large fraction of these areas is permafrost [30], where soil freezing/thawing conditions play a major role in controlling the hydrological processes [31,32]. Representing permafrost in LSMs is important for better understanding of the hydrological variability and the impacts of climate change [32–34].

In this study, the hydrological impact of the MLS implemented in ECLand is analyzed on more than 400 catchments globally, with over a third located partially or entirely in permafrost areas. To achieve this, ECLand experiments forced with ERA5 over the period 1979–2018 are coupled to the Catchment-based Macro-scale Floodplain model (CaMa-Flood; [35]) to generate river discharge, allowing direct comparison with gauged observations. Different sensitivity experiments are conducted to evaluate the impact of the more physically complex snow scheme on the terrestrial hydrological processes, with particular focus on permafrost, where complicated error dynamics arise from different land-surface processes. Two main questions are posed:

- How does the MLS impact the simulated hydrological processes and river discharge, especially in the snowmelt-driven flood season?
- How sensitive is the hydrological representation of permafrost to the snow and soil parametrization?

2. Materials and Methods

In this section, the data set, models and methods used will be described.

2.1. ECLand Land-Surface Model and Offline Methodology

The hydrological core of the analyzed data sets was provided by the ECLand land-surface model, formerly known as HTESEL (The Hydrology-Tiled ECMWF Scheme for Surface Exchange over Land [20,21,36]). ECLand is part of the Integrated Forecasting System (IFS) at ECMWF and used in coupled land-atmosphere simulations for describing the evolution of soil, vegetation and snow conditions over land, at various spatial resolutions, from short- to seasonal-range.

In ECLand, up to six tiles are present over land (bare ground, low and high vegetation, intercepted water, shaded and exposed snow) and three over water (inland, open and frozen water) that provide the interface between the atmosphere and the one-dimensional soil column, with all tiles having their separate energy and water balances.

The snowfall (the solid fraction of precipitation) is collected in the snowpack, which overlays the soil [20]. The fraction of the soil that is covered by snow (snow cover fraction) is parametrized as a linear function of snow depth [17]. This assumes that a model grid-box is fully covered for snow depth greater than 10 cm. The same parametrization is used for exposed and shaded snow (i.e., snow under high vegetation) tiles in ECLand. The soil is divided into four layers with fixed layer depths (0–7, 7–28, 28–100 and 100–289 cm). Runoff is generated as fast (surface) and slow (subsurface) components at each grid point [20,21].

Snowmelt occurs when the temperature of the snow is high enough, contributing to surface runoff, soil infiltration and evaporation. Some part of rain and snowmelt will be removed as surface runoff. This surface runoff fraction depends on the standard deviation of the sub-grid scale orography (a measure of unresolved orographic features), the soil texture and the soil water content. Subsurface runoff is the water leaving the soil column at the bottom. It depends on the infiltration and surface evaporation as top boundary conditions, while water can be extracted by roots in each soil layer where vegetation is present [20,21].

ECLand can be used in a stand-alone mode, when the model runs uncoupled from the atmosphere, usually with hourly time step, forced with near-surface meteorological input data of temperature, specific humidity, wind speed, surface pressure, radiative fluxes (downward solar and thermal radiation) and water fluxes (liquid and solid precipitation), without land data assimilation. This offline research methodology provides an affordable way of testing land-surface improvements and has been used in various applications (e.g., [26,37–39]). The ERA5-Land dataset is a prime example of this methodology, which was produced as an offline ECLand simulation with downscaled ERA5 meteorological forcing on higher resolution, including an elevation correction for the thermodynamic near-surface state [23].

2.2. ERA5 Reanalysis

The meteorological forcing for the offline ECLand simulations was taken from ERA5, the latest global climate reanalysis of ECMWF [22]. ERA5 is a key contribution to the EU-funded Copernicus Climate Change Service (C3S) and is open access and free to download for all users (<https://cds.climate.copernicus.eu/> (accessed on 24 October 2021)). It covers the period 1979 to present, with a preliminary version also available from 1950. It includes a high-resolution component (~31 km) and a lower resolution (~62 km) ensemble component with 10 members. In this study, the high-resolution component (hereafter referred to as ERA5) was used from 1979 with ~31 km horizontal resolution and hourly output frequency.

2.3. CaMa-Flood River-Routing

The hydrodynamics to produce river discharge from the ECLand runoff output were simulated by CaMa-Flood [35], a global river-routing model, which is part of ECLand since IFS cycle 47r1 [21]. CaMa-Flood routes runoff generated by land-surface models to

oceans or inland seas. The model calculates river and floodplain water storages, discharge, water depth, as well as flood inundation. CaMa-Flood does not currently include the representation of dams and permanent lakes and wetlands are only treated as part of the floodplain storages. CaMa-Flood is computationally cheap to run and has been used widely in global climatological research studies, such as Emerton et al. [40], Dottori et al. [41] and Zsoter et al. [39].

2.4. ECLand Snow and Soil Freezing Schemes

The current SLS, used operationally in ECLand, is a basic energy balance model describing the temporal evolution of the heat and mass contents of the snowpack [17]. The MLS, used experimentally in ECLand, is an intermediate complexity snow scheme [14], which represents the vertical structure and time evolution of snow temperature, density, liquid water content and surface snow albedo with up to five active snow layers. In this section, the important model features are described that are necessary to understand the scheme variations tested in this paper (described later in Section 2.8). Further details of MLS, including a detailed comparison of snowpack properties to SLS, are given in Arduini et al. [26].

2.4.1. Snow Vertical Discretization

The number of active snow layers and their associated thicknesses are defined diagnostically at each time step before the updating of the other snow variables. The number of active layers (N) is dependent on the snow depth D_{sn} and varies from one layer to a maximum of five (N_{max}). The topmost snow layer, in contact with the atmosphere, is assumed to be the first one, whereas layer N is the one in contact with the soil. N is determined as the lowest number that satisfies the following inequality for $N = 1, \dots, N_{max}-1$:

$$\sum_{j=1}^{N+1} D_{min,j} > D_{sn}, \quad (1)$$

where $D_{min,j}$ is the minimum snow depth allowed for layer j . It is by default set to 0.05 m for all layers (denoted hereafter as D_{min} for all layers). The depth of the first layer is defined as:

$$D_{sn,1} = \begin{cases} D_{sn}, & \text{if } D_{sn} < 2D_{min} \\ D_{min}, & \text{if } D_{sn} > 2D_{min} \end{cases}. \quad (2)$$

Note that with this choice, MLS has only one active snow layer for $D_{sn} < 0.1$ m. For the remaining layers, the vertical discretization is defined as:

$$D_{sn,i=2,\dots,N_{max}} = \begin{cases} 0, & \text{if } D_{sn} < \sum_{j=1}^{j=i} D_{min,j} \\ \min\left[\frac{D_{sn}-D_{sn,1}}{N-1}, D_{max,i}\right], & \text{if } D_{sn} > \sum_{j=1}^{j=i} D_{min,j} \end{cases}, \quad (3)$$

where $D_{max,i}$ is the maximum snow depth allowed for the i th active snow layer. This effectively means, the snow is evenly divided into the remaining layers, as long as the maximum layer depths allow it. By default, these maximum values are 0.05, 0.10, 0.20, ∞ and 0.15 for layers $1-N_{max}$, respectively. This definition of the maximum layer depths means that when all N_{max} snow layers are active, the $N_{max} - 1$ layer is used as the accumulation layer for thick snowpacks. This layering allows a relatively high vertical resolution both at the interfaces to the atmosphere above and to the soil underneath the snowpack. Take $D_{sn} = 1.25$ m as an example, for this depth the snowpack is discretized into 5 layers with thicknesses of 0.05, 0.1, 0.20, 0.75 and 0.15 m from top to bottom.

2.4.2. Destructive Metamorphism of the Snow

The density of freshly fallen snow can vary rapidly with time due to metamorphic processes, that is, the change in shape and size of snow grains once they settle in the snowpack. The rate of change of snow density due to destructive metamorphic processes

of the snow (ζ) is parametrized in both SLS and MLS using the formulation introduced by Anderson [42]:

$$\frac{1}{\rho_{sn,i}} \left[\frac{\partial \rho_{sn,i}}{\partial t} \right]_{\zeta} = a_{\zeta} \exp \left[-b_{\zeta} (T_f - T_{sn,i}) - c_{\zeta} \max(0, \rho_{sn,i} - \rho_{\zeta}^*) \right], \quad (4)$$

where $T_{sn,i}$ and $\rho_{sn,i}$ are the snow temperature and snow density for each snow layer (for SLS, $i = 1$) and $a_{\zeta} = 2.8 \times 10^{-6} \text{ s}^{-1}$, $b_{\zeta} = 4.2 \times 10^{-2} \text{ K}^{-1}$, $c_{\zeta} = 460 \text{ m}^3 \text{ kg}^{-1}$, $T_f = 273.16 \text{ K}$ and $\rho_{\zeta}^* = 150 \text{ kg m}^{-3}$.

2.4.3. Snow-Soil Thermal Conductivity

In SLS and MLS, the thermal coupling between the snowpack and the soil underneath is described using a thermal conductance between the two media (λ_b), thus the heat flux (G_b) can be written as:

$$G_b = \lambda_b (T_{sn,N} - T_{so}), \quad (5)$$

where $T_{sn,N}$ is the snow temperature of the bottom active snow layer ($N = 1$ for SLS) and T_{so} is the temperature of the topmost soil layer. Given that the heat resistances are in series, the sum of the inverse of the conductance of each medium yields the total conductance of the snow-soil system, that is:

$$\lambda_b^{-1} = \frac{l_b \Delta z_{sn,N}}{\lambda_{sn,N}} + \frac{l_b \Delta z_{so}}{\lambda_{so}}, \quad (6)$$

where $\Delta z_{sn,N}$ and $\lambda_{sn,N}$ are the thickness and conductivity, respectively, of the bottom active snow layer, and Δz_{so} and λ_{so} are the thickness and conductivity of the topmost soil layer. The parameter l_b is set as 0.5 in SLS, whereas it was changed to 1 in MLS, as described in Arduini et al. [26], to account for the additional insulation effect due to organic material or vegetation between the snow layer and the soil.

2.4.4. Soil Freezing Scheme and Relationship to Runoff Generation

Frozen soil is characterized by very different thermal and hydrological properties compared to unfrozen (wet) soil. For instance, precipitation infiltrates less into a frozen soil, thus more water goes into runoff than into the soil. In ECLand, a simplified representation of unfrozen soil was introduced by Viterbo et al. [43] to account for the latent heat release/absorption of soil water around 0 °C, reducing a pronounced cold bias in 2-metre temperature in the ECMWF forecasts. In this simplified approach, the frozen water fraction in each soil layer ($\theta_{ice,i}$, $i = 1, \dots, 4$) is given as a function of the soil temperature ($T_{so,i}$) as:

$$\theta_{ice,i} = \begin{cases} 0 & \text{for } T_{so,i} > T_{Th} \\ 0.5 \left(1 - \sin \left(\frac{\pi (T_{so,i} - 0.5(T_{Th} + T_{Fr}))}{T_{Th} - T_{Fr}} \right) \right) & \text{for } T_{Fr} \leq T_{so,i} \leq T_{Th} \\ 1 & \text{for } T_{so,i} < T_{Fr} \end{cases} \quad (7)$$

where $T_{Th} = +1 \text{ }^\circ\text{C}$ (thaw temperature) and $T_{Fr} = -3 \text{ }^\circ\text{C}$ (freeze temperature) are the soil temperature thresholds for which soil water is totally unfrozen or frozen, respectively, and $[T_{Th}, T_{Fr}]$ is the temperature interval for which phase change can occur. Both SLS and MLS use this soil freezing scheme.

The runoff generation in EC-Land follows the formulation described in Balsamo et al. [20], which includes a dependency of surface runoff on soil textures as well as sub-grid scale orography features not resolved at the resolution used in the simulation. The surface runoff can be as large as 50% of the available precipitation and snowmelt, for large standard deviation of the sub-grid scale orography and finer soil textures.

When the soil is partially frozen, the surface runoff is enhanced, as less water infiltrates and percolates within the soil column. This soil freezing mechanism is represented by

reducing the soil hydraulic conductivity and diffusivity. This is done by computing the soil hydraulic conductivity and diffusivity as a weighted average of the values for the unfrozen soil water fraction and for the frozen water fraction [20].

As previously stated, the frozen water fraction parametrization described in Equation (7) has been developed to address temperature errors in weather forecasting applications. Such simplified approach can have limitations for hydrological applications in cold regions, where the interaction between frozen soil and runoff is key in modulating river streamflow (see for instance [44]).

2.5. River Catchment Selection

The river catchments were selected for this study only if they experience regular snowfall and they have adequate river discharge observations available. Observations are selected from the Global Runoff Data Centre (GRDC; <https://www.bafg.de/GRDC/> (accessed on 20 November 2019)) supplemented by additional data collected by the Copernicus Emergency Management Service for floods in the Global flood Awareness System (GloFAS), as described in Harrigan et al. [6]. The catchments and associated river discharge observations are selected with the following set of criteria:

- Minimum 8 years of river discharge observations in 1980–2018. Gaps are not considered a problem, as long as the climatological mean can be computed for each day of the year (see Section 2.7);
- Stations in snow impacted climate, defined by the percentage ratio of ERA5 snowfall and total precipitation being at least 10%, based on the 1979–2018 mean for each catchment;
- Catchment area of at least 5000 km² (e.g., minimum of 8 river pixels);
- Good general quality. After visual inspection of the river discharge time series, the catchments that showed observation errors, problems with station metadata (wrong or uncertain location, etc.) or visible influence of dams and lakes were excluded. To help with identifying reservoir and lake influence, the Global Reservoir and Dam Database (GRAND; [45]) and the Global Lakes and Wetlands Database (GLWD; [46]) were used as visual tools.

Out of the 2119 stations in the Copernicus Emergency Management Service GloFAS station database with at least 1 year of observations, 1913 met the criteria of at least 8 years of data, 889 catchments had also at least 10% snowfall ratio and 849 were additionally over 5000 km² area. In total, 453 catchments were selected after the quality checks, almost entirely in the Northern Hemisphere (Figure 1). For the sensitivity analysis in the permafrost, a specific area was defined in the 60–80 N belt containing parts of northern Siberia, Alaska and western Canada. It focuses on the coldest parts of the permafrost. The eastern area of Canada was omitted due to the large number of lakes in the area to avoid unduly influencing of the overall results. Figure 1 shows the bounding box, together with the permafrost, as defined by areas of below 0 °C climatological annual mean temperature in the lowest soil layer of ECLand in ERA5 for 1st of June. Within ECLand, the land use of the catchments in the permafrost sensitivity area are mainly a mixture of tundra and boreal forest vegetation, whereas the soil textures are medium-coarse and coarse soil types, which are characterized by a relatively low field capacity and higher hydraulic conductivity (for details on land use and soil hydraulic properties see [20,21]).

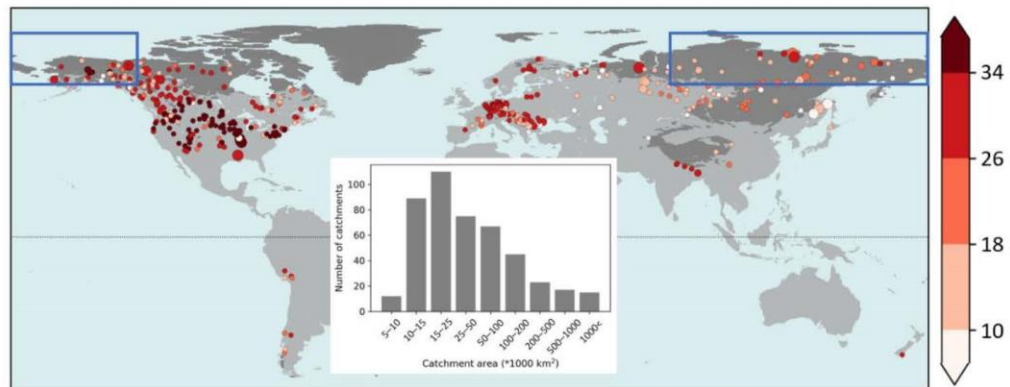


Figure 1. Stations used in this study with the number of river discharge observation years available (8 years is the minimum). In total, 453 catchments worldwide. The darker grey shading indicates the permafrost areas (defined as the area where the lowest soil layer's temperature is below 0 °C in the ERA5 climate mean on 1st of June, based on 1980–2018), while the blue rectangle shows the sensitivity area defined in the permafrost. The distribution of the catchment upstream area values is provided in the inset table (please note the area values are divided by 1000).

2.6. Verification Statistics

Hydrological performance is assessed with the modified Kling–Gupta efficiency (kge ; [47,48]). The kge is increasingly considered as the standard performance metric in hydrology [49,50]. It can be decomposed into three components, measuring the correlation, bias and variability errors, which makes it an easy to interpret metric, ideal for assessing hydrological dynamics:

$$kge = 1 - \sqrt{(pcorr - 1)^2 + \left(\frac{\mu_s}{\mu_o} - 1\right)^2 + \left(\frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} - 1\right)^2}. \quad (8)$$

In the kge decomposition, $pcorr$ is the Pearson correlation coefficient between daily simulation (s) and observation (o) time series, measuring the temporal errors; μ is the mean and σ is the standard deviation of the time series. In Equation (1), $\frac{\mu_s}{\mu_o}$ is the bias ratio, while $\frac{\sigma_s/\mu_s}{\sigma_o/\mu_o}$ is the variability ratio, which highlight how close the mean and the variability (normalized by the means) are in the simulated and observed time series. The kge and its three components are all dimensionless. The correlation ranges from -1 to $+1$, with $+1$ showing perfectly strong linear relationship and 0 no linear relationship (-1 being perfect inverse relationship). In the kge definition, the bias and variability ratios both range from 0 to infinity, with 1 being the optimal value. The bias and variability errors, used in this study, were defined by the bias and variability ratio components of the kge as follows:

$$bias = \frac{\mu_s}{\mu_o} - 1 \text{ and } var = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} - 1, \quad (9)$$

This way, the optimal score value transforms to 0 , highlighting the direction of the biases more intuitively, with negative values showing underprediction, while positive ones overprediction. To aid comparison between different experiments, the absolute version of the $bias$ ($abias$) and var ($avar$) errors are also used. Change in these metrics can directly highlight improvement or deterioration, while they also share the optimal value of zero.

2.7. Daily Climatology Computation

The land-surface contribution to the water budget is diagnosed qualitatively at specific catchments by using daily climatological mean time series of simulated and observed river discharge, runoff (surface and subsurface components), snowpack water content and snowmelt, evaporation, soil temperature and water content (at different layers) and precipitation, computed from daily values in the 1980–2018 period. All water related variables are converted into catchment totals as the sum of all the grid point values in the catchment, in order to compare them directly to river discharge. For temperature variables, the catchment averages are used.

These climatological means are computed for every day of the year (1 January–31 December), by applying a 21-day window, centered over the day of the year. To aid direct comparison with observed river discharge, only those days of 1980–2018 are considered for the climatological mean computation, which have river discharge observations available. For the climatological mean, the minimum data length was lowered to 4 years, in order to maximize the likelihood of being able to compute the mean for every day of the year. With this choice, the climate sample size could range from 84 values (4 years, 21 values each) to over 800 (most years in 1980–2018) to compute the daily climate mean.

2.8. Experimental Setup

In this study, the hydrological impact of the MLS is analyzed on ECLand/CaMa-Flood coupled experiments. The runoff is produced by ECLand, while the CaMa-Flood model is used to produce river discharge by routing the runoff over the 15 arcmin (~25 km on the Equator) river network, which is an appropriate horizontal resolution for the related meteorological forcing of ERA5.

CaMa-Flood uses a 1-h time step and a 24-h output frequency to match the 24-h reporting frequency of the river discharge observations. All experiments are generated for the ERA5 period of 1979–2018, while for the hydrological analysis 1980–2018 is used with 1979 omitted to account for the spin-up in the simulations. Preliminary analysis showed that a 1-year spin up period is appropriate, as a longer spin up period did not have a large influence on results but considerably reduced the sample size.

In total, 13 experiments are produced and compared (Figure 2). Two experiments use SLS, while the other 11 are produced with variations of MLS and the soil freezing parametrization in ECLand, for analyzing the hydrological sensitivity, focusing on permafrost areas. The very high computational cost of running these experiments meant that it was not possible to run all permutations of the schemes and parameters tested, instead possible modifications build on each other incrementally.

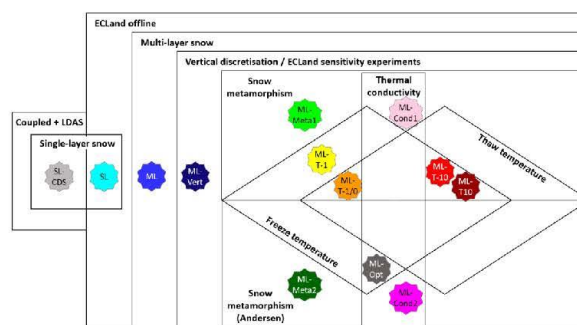


Figure 2. Experiments analyzed in this study, describing the main simulation features including the snow and soil scheme modifications for sensitivity analysis in permafrost. The experiments are displayed with the star shapes, using a short name and a color to help identifying them throughout the paper.

Single-layer, online, fully coupled with land data assimilation: SL-CDS

The first experiment involves a single CaMa-Flood run, using the original ERA5 runoff data (downloaded from the Copernicus Climate Data Store), which is produced online with land-atmosphere coupling, including atmospheric and land data assimilation and SLS (SL-CDS).

Offline experiments

All other experiments that follow include a surface only (offline) ECLand simulation, without land-atmosphere coupling and land data assimilation, to produce runoff, which is then routed with CaMa-Flood. This is because online experiments with coupling and data assimilation would be infeasible to run due to the very high computational cost. The offline experiments are initialized from the ERA5 state on 1 January 1979 and forced with ERA5 near-surface meteorological data on ~31 km horizontal resolution and hourly output frequency (see Section 2.1 for further details).

Single-layer snow scheme: SL

In order to compare the online and offline modelling approaches directly, one of the offline experiments is run with the SLS (SL), to be compared with SL-CDS (see Zsoter et al. [39] for further details on online/offline comparison).

Multi-layer snow scheme: ML

The first MLS experiment uses the default snow parametrization, introduced by Arduini et al. [26] and default ECLand soil freezing parametrizations (see Section 2.4) and can be considered the default multi-layer experiment (ML).

ECLand sensitivity experiments

The following 10 experiments are variations of the ECLand snow and soil freezing schemes, designed to evaluate the river discharge sensitivity in permafrost.

Vertical snow discretization: ML-Vert

The first change is for the vertical snow discretization in MLS (ML-Vert). It introduces thicker snow layers over complex terrains with deep snowpack and reduces issues related with excessive melting over mountainous regions [21]. Complex terrain is defined as areas with standard deviation of the sub-grid scale orography over 50 m. For flat terrain and for $D_{sn} < 0.25$ m over complex terrain, the same vertical discretisation is used as in ML (described previously in Section 2.4). For $D_{sn} > 0.25$ m over complex terrain, $D_{max,i}$ and $D_{min,i}$ can vary with snow depth in Equations (1)–(3) as follows:

$$D_{min,i} = \min(0.25, 0.10 + \alpha_0(D_{sn} - 0.25)) \text{ for } i = 1 \quad (10)$$

$$D_{max,i} = \begin{cases} \min(0.25, 0.10 + \alpha_0(D_{sn} - 0.25)) & \text{for } i = 1 \\ \min(0.30, 0.15 + \alpha_0(D_{sn} - 0.25)) & \text{for } i = 2, \dots, N_{max} \end{cases} \quad (11)$$

where $\alpha_0 = 0.1$ is a predefined parameter. For the example of $D_{sn} = 1.25$ m (as in Section 2.4), the thickness of the 5 layers changes to 0.20, 0.25, 0.25, 0.30 and 0.25 m, which means the snowpack is more evenly distributed for this large depth.

Destructive metamorphism of the snow: ML-Meta1 and ML-Meta2

The second group of changes is for the snow metamorphism. The parametrization for the destructive metamorphism of the snow (Equation (4)), as used in SL, SL-CDS and ML with the default value of $c_{\bar{z}} = 460 \text{ m}^3 \text{ kg}^{-1}$, implies that the rate of snow density changes due to the 2nd density-dependent term in the exponential, and is active only for $\rho_{sn,i} < 150 \text{ kg m}^{-3}$, that is for relatively fresh snowpack. Cao et al. [51] pointed out, while evaluating soil temperature and snow characteristics of ERA5-Land in permafrost regions, that this can partly explain the underestimation of the snow density of ERA5-Land. They have argued that the underestimation of the snow density could lead to an overestimation

of the thermal decoupling between the atmosphere and the soil underneath. This could contribute to the warm bias of soil temperature over permafrost regions as less heat is diffused from the soil towards the colder atmosphere above.

To address this, the impact of the representation of the snow density on river discharge is explored in two experiments, by changing the value of the parameter c_{ξ} . In the first experiment (ML-Meta1), the parameter c_{ξ} is varied for the five snow layers, using values closer to the $0.046 \text{ m}^3 \text{ kg}^{-1}$, reported in Anderson [42], as follows:

$$c_{\xi} = (0.112, 0.152, 0.192, 0.288, 0.488) \text{ m}^3 \text{ kg}^{-1}. \quad (12)$$

These values are chosen as they were the best compromise in terms of land-surface and atmospheric impact (i.e., in particular 2-metre temperature) in coupled land-atmosphere forecast experiments, which were conducted for the foreseen implementation of the MLS in operational weather forecasts at ECMWF. The decreasing values towards the first (top) snow layer imply that the destructive metamorphic process of the snow is more active for the top snow layers, whereas for the settled snow at the bottom of the snowpack it is less active, by this making the compaction process slower.

Another experiment (ML-Meta2) uses the original value of $c_{\xi} = 0.046 \text{ m}^3 \text{ kg}^{-1}$ for all snow layers, as reported by Anderson [42], which is a commonly used approximation in land-surface models. This latter unified and low value option of c_{ξ} gives the opportunity to explore the maximum sensitivity to the destructive metamorphic process of the snow in the snow density representation, without considering the implication for the snow-atmosphere coupling (e.g., atmospheric scores), given that the feedback to the atmosphere is not considered in the offline experiments presented in this work. In coupled experiments, the impact of this change would be expected to be substantial on the near surface variables, like 2 m temperature, possibly requiring an additional tuning of other land-atmosphere coupling parameters. Both experiments build on ML-Vert by adding the snow metamorphism corrections to the vertical snow depth discretization adjustment.

Snow-soil thermal conductivity: ML-Cond1 and ML-Cond2

The next two experiments use the parameter $l_b = 0.5$ in the snow-soil thermal conductivity computation (Equation (6)), the same value as in SLS and reduced by half compared with MLS. Organic material distribution is highly variable both in horizontal and vertical (within the soil) scales and its handling requires more sophisticated parametrizations [16]. For this reason, in these two experiments we relax the hypothesis of $l_b = 1$ in MLS, removing the additional thermal insulation effects caused by organic material, effectively considering half of the topmost soil layer in the computation of the conductance between the two media (see Equation (6)), consistently with what is done for the other snow-free land-surface tiles of ECLand. Effectively, this increases the thermal coupling between the snow and soil and thus also increases the heat flux between the two media in these experiments. Both experiments include the vertical discretization change (ML-Vert), while ML-Cond1 is run together with the first snow metamorphism change in ML-Meta1 and ML-Cond2 with the change in ML-Meta2.

Soil freeze and thaw temperatures: ML-T-1, ML-T-1/0, ML-T10 and ML-T-10

The next step in the experiment design is to test the sensitivity of river discharge simulations to the fraction of frozen water in the soil. Soil freezing and thawing is particularly important during the snowmelt season, when the thermal coupling of the topmost soil layers to the atmosphere increases quickly as the snow melts and phase changes of the soil water can occur. Out of the two temperature parameters, the freeze temperature has a more direct impact on runoff generation during the spring period, whereas the thaw temperature could be more important for energy fluxes and the coupling to the atmosphere in the fall/winter season. Gouttevin et al. [34] suggested that a shorter temperature interval for phase change works better for permafrost.

These soil temperature experiments are run including the changes in ML-Vert and ML-Meta1. First, while keeping the T_{Th} at the default value of $+1$ °C in Equation (7), the T_{Fr} parameter is increased from the default -3 to -1 °C (ML-T-1). Similarly, in another experiment, while keeping T_{Fr} at the higher level of -1 °C, T_{Th} is decreased from the default $+1$ to 0 °C (ML-T-1/0).

In addition, the extreme boundaries of the soil freezing contribution to river discharge are also tested, by setting the $[T_{Fr}, T_{Th}]$ phase change interval unreasonably high at (ML-T10) and also unreasonably low at $[-10.5, -10]$ (ML-T-10). These extreme temperature thresholds allow the soil to remain almost always or never frozen, which consequently should increase or decrease the amount of infiltration to the soil and thus the amount of surface runoff to an extreme level.

Optimal combination: ML-Opt

Finally, a prospective experiment is defined with combinations of some evaluated changes that are expected to work best for permafrost. ML-Opt combines incremental changes in ML-Vert, ML-Meta2, ML-Cond2 and ML-T-1 into one experiment. This experiment aims at exploring the interactions among the proposed changes, as feedbacks between different processes can be highly non-linear. For instance, testing the proposed changes in combination can indicate if singular modifications that improve the river discharge simulation, actually (over-) compensate for other sources of errors.

3. Results

3.1. Default Multi-Layer vs. Single-Layer Snow Schemes

The default MLS generally improves on the SLS, mainly through better bias and variability, with the exception of some parts of the permafrost, where the multi-layer simulation is suboptimal for river discharge.

The impact of the MLS is analyzed first in this section by comparing the default parametrization option ML with the single-layer SL (Figure 3). The ML improves the river flow predictions for the majority of stations over the midlatitudes (about two thirds of them), with a larger cluster of improved catchments (with higher kge) present in western/central North America (Figure 3a). The kge mean, computed across 453 catchments, is 0.43 for ML while 0.40 for SL, highlighting a small overall improvement. However, in some higher latitude areas, especially in the northern half of Siberia in Asia and also near Alaska in North America (coinciding with the blue box permafrost sensitivity area in Figure 3), the river discharge performance is deteriorated in ML, with many catchments showing a drop of at least 0.05–0.1 (some even above 0.3) in kge .

The improvements in midlatitudes seem to come mainly from the smaller bias errors ($abias$ decreasing), especially pronounced in central North America (Figure 3c) and to a lesser extent from the smaller variability errors ($avar$ decreasing; Figure 3d) and higher correlation (Figure 3b). On the other hand, the deterioration in the higher latitudes over the permafrost seems to relate to much higher variability errors (many catchments with an increase of at least 0.3) and to somewhat higher bias errors and lower correlation in ML.

3.2. ML Struggles in Permafrost

Analysis in Siberia demonstrates that the deterioration of the daily river discharge representation and the largely missed snowmelt-driven flood wave is caused by too low surface runoff, which primarily comes from warmer soil in ML, allowing more water infiltrating into the soil and thus reducing surface runoff.

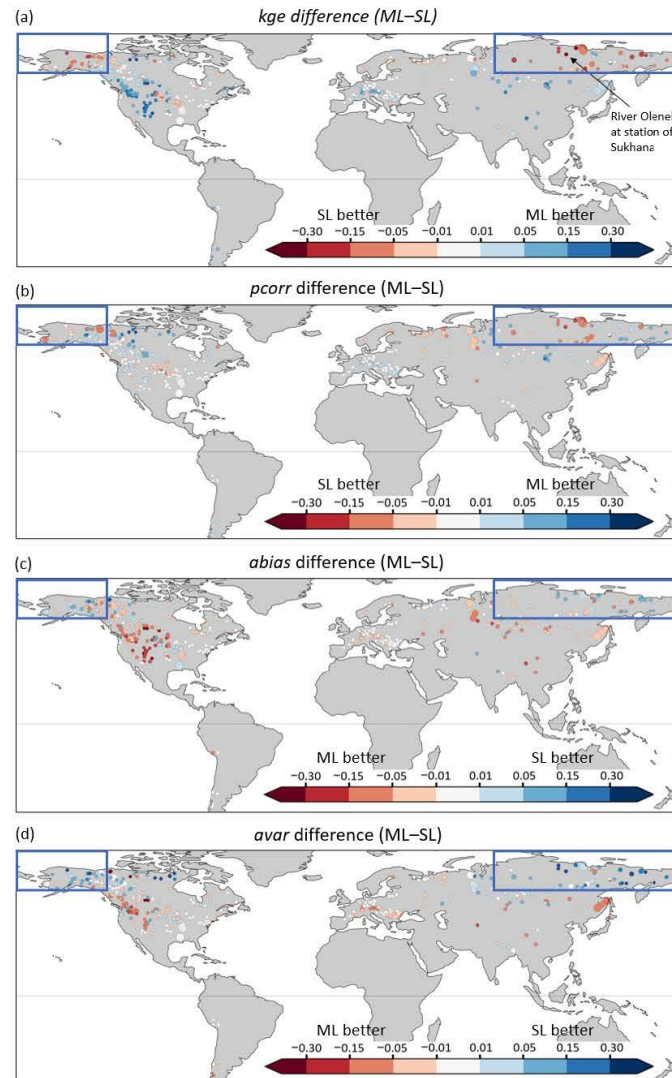


Figure 3. Difference of performance metrics between the default multi-layer (ML) and the single-layer (SL) snow scheme experiments, across all 453 stations, calculated on daily river discharge over 1980–2018. (a) Modified Kling–Gupta efficiency (*kge*) and (b) Pearson correlation (*pcorr*), (c) absolute bias ratio (*abias*) and (d) absolute variability ratio (*avar*). Improvements in ML are indicated by blue dots in (a,b), while by red dots in (c,d). Size of the dots represent the catchment area. The sensitivity area in the permafrost is shown by blue rectangles, while the test catchment on river Olenek at station Sukhana is indicated by black arrow in (a).

A test catchment on the Olenek river in Siberia (Sukhana station from within the permafrost sensitivity area; indicated in Figure 3) is used to demonstrate this large negative impact through analyzing the daily climate time series of some key water budget related land-surface variables (snowmelt, surface and subsurface runoff, soil water content and soil temperature) for SL-CDS, SL and ML (Figure 4).

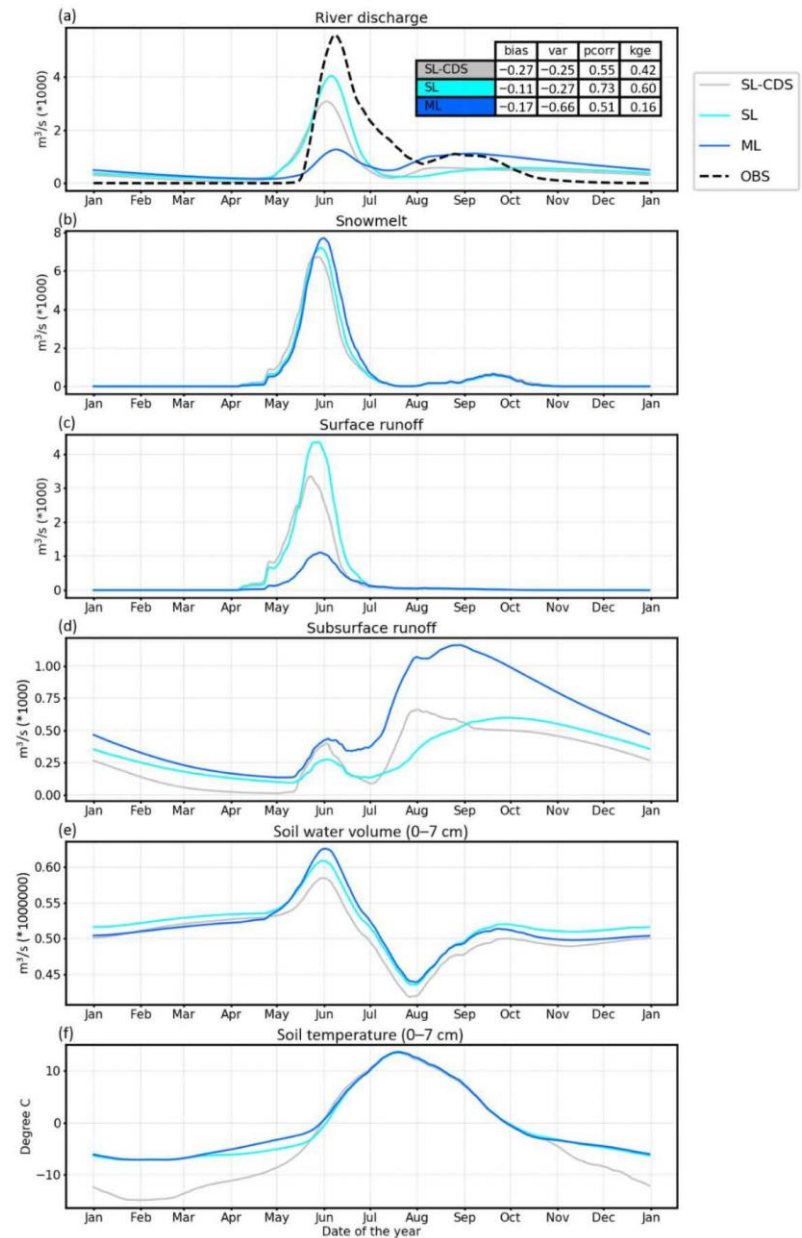


Figure 4. Daily climatological mean time series of (a) river discharge, (b) snowmelt, (c) surface runoff, (d) subsurface runoff, (e) water content and (f) temperature in the top 7 cm of the soil from SL-CDS, SL and ML experiments for the Olenek river at the station of Sukhana in eastern Siberia (with area of 127,000 km²). All water related variables are displayed as catchment totals in order to compare them directly to river discharge, while for soil temperature catchment averages are shown (please note the values are divided by 1000 for river discharge, snowmelt and subsurface runoff and by 1,000,000 for soil water volume). The *kge* and its *bias*, *var* and *pcorr* component scores are provided in an inset table for river discharge.

The ML-produced river discharge is clearly inferior compared with both SL and SL-CDS, as it shows a much larger underestimation of the observed flood peak in May–June (Figure 4a). Even though the secondary flood period is better represented by ML in August–September, due to the generally higher river discharge with the exception of May–June. The deterioration of the flow is reflected in the ML scores being the lowest across all three experiments (inset table in Figure 4a). The only exception is the bias ratio which is lowest for SL-CDS. This very low negative bias is a consequence of the snow data assimilation that removes water from the rivers in high latitudes. It is related to the current single layer snow model's tendency to melt the snow too slowly, which is compensated by the assimilation system during snowmelt periods, as documented in Zsoter et al. [39].

The snowmelt peak in May–June is slightly higher and delayed in ML, compared with SL and even more so with SL-CDS (Figure 4b). At the same time, the soil is better insulated by the multi-layer snow in ML during March–May, as the temperature is higher by up to 2 °C than in SL (Figure 4f). As the snow is better insulated in ML, the melting will start later during spring, which then delays the faster soil temperature increase in ML, occurring in end of May. The surface runoff is highest in SL and lowest in ML during April–June, also highlighting a similar delay seen on the snowmelt peaks (Figure 4c).

The higher soil temperature in ML implies that a smaller fraction of the water is frozen in the soil, compared to SL, during the crucial snowmelt period and more water can infiltrate into the soil, which then reduces the amount of water that can runoff directly. The higher infiltration will increase the water content in the soil during May–July (Figure 4e), which will result in increased subsurface runoff (Figure 4d). This happens with a delay, as the water needs to reach the bottom of the soil to leave as subsurface runoff, producing a delayed peak to around end of July–September (Figure 4d).

The missing river discharge in ML is clearly related to the too low surface runoff during the April–June snowmelt, which period is clearly better represented in both SL and SL-CDS. The small differences in snowmelt cannot explain the lower surface runoff amounts. If anything, it should likely contribute to an increase in ML. In addition, the deficit in surface runoff is offset by the higher subsurface runoff in ML. However, the excess water from subsurface runoff spreads out over most parts of the year and thus cannot compensate for the missing surface runoff during the snowmelt flood period. Instead, the crucial aspect of the changes in ML is the higher soil temperature. This soil temperature difference is most critical in the middle of May, when the snowmelt is rapidly increasing in the catchment.

3.3. Improving the Multi-Layer Snow Scheme Performance in Permafrost

In this section, the ECLand sensitivity experiments (Figure 2 and Section 2.8) are evaluated for their ability to improve the river discharge simulation, focusing on permafrost.

3.3.1. Impact of the ECLand Experiments on a Test Catchment in Siberia

The ECLand permafrost sensitivity experiments are demonstrated to increase the surface runoff in a test catchment in permafrost, by primarily making the soil colder through a series of incremental changes in the snow and soil freezing parametrizations, including modifications of the snow vertical discretization, snow density metamorphism, snow-soil thermal conductivity and soil freeze temperature.

In the Sukhana test catchment, the modifications are clearly effective in altering the amount of water distributed amongst different parts of the water budget (Figure 5). Generally, the impact on snowmelt (Figure 5b) is the smallest (maximum of a few days delay and a slight change in magnitude), while all other variables show much larger variability between the experiments (Figure 5a,c–f).

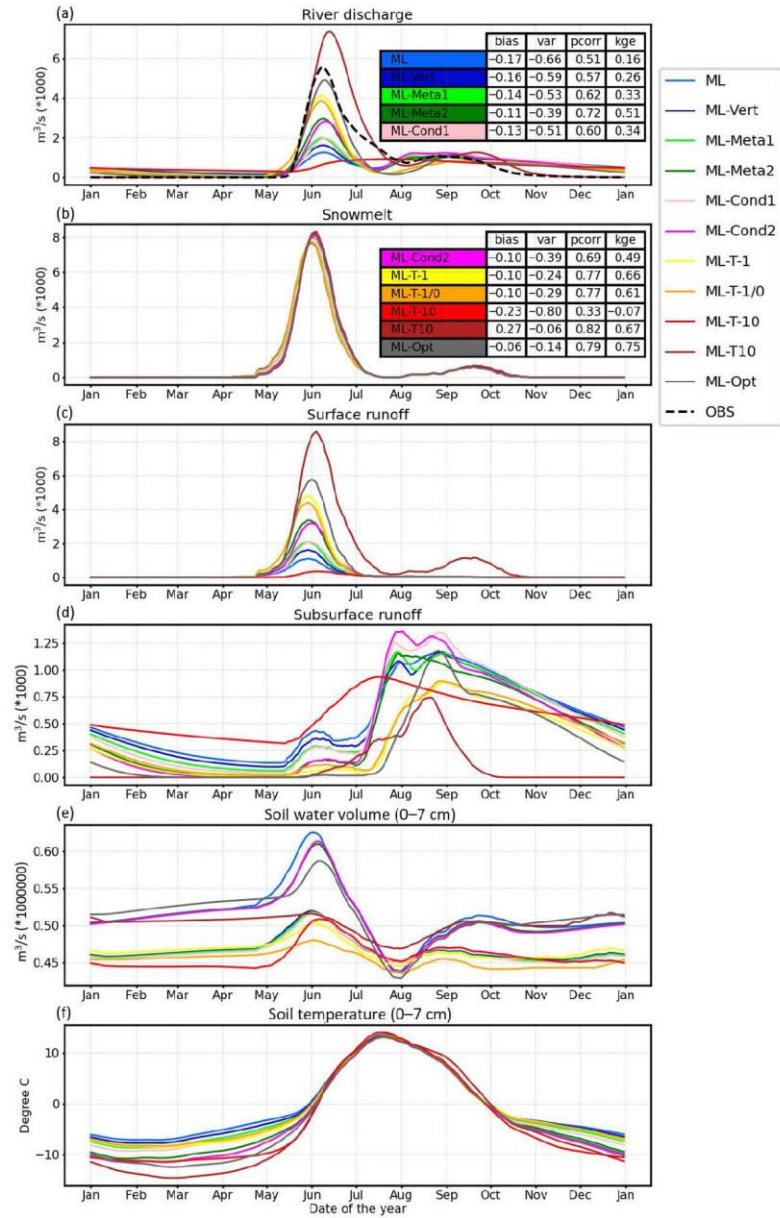


Figure 5. Daily climatological mean time series of (a) river discharge, (b) snowmelt, (c) surface runoff, (d) subsurface runoff, (e) water content and (f) temperature in the top 7 cm of the soil from ML, ML-Vert, ML-Meta1, ML-Meta2, ML-Cond1, ML-Cond2, ML-T-1, ML-T-1/0, ML-T-10, ML-T10 and ML-Opt experiments for the Olenek river at the station of Sukhana in eastern Siberia (with area of 127,000 km²). All water related variables are displayed as catchment totals in order to compare them directly to river discharge, while for soil temperature catchment averages are shown (please note the values are divided by 1000 for river discharge, snowmelt and subsurface runoff and by 1,000,000 for soil water volume). The *kge* and its *bias*, *var* and *pcorr* component scores are provided in an inset table for river discharge.

The surface runoff shows very large sensitivity to the ECLand changes in May–June, coinciding with the main snowmelt period (Figure 5c). The river discharge behavior (Figure 5a) is directly determined by the variability in surface runoff in this period. The vertical discretization adjustment (ML-Vert), which is active for about 25% of the test catchment area, usually during October–May, and the combination with the first snow metamorphism change (ML-Meta1) both add small increase. However, ML-Vert combined with the second snow metamorphism setting (ML-Meta2) produces a much larger impact on surface runoff. On the other hand, the change in surface conductivity does not appear to be effective as both ML-Meta1 with ML-Cond1 and ML-Meta2 with ML-Cond2 (both pairs sharing the same snow metamorphism change respectively) show similar levels of surface runoff.

The experiments with the largest impact on surface runoff (and thus river discharge) are those including changes of the soil freezing. The experiments of ML-T-1 and ML-T-1/0 help reaching the surface runoff level of SL by resetting the freeze temperature to $-1\text{ }^{\circ}\text{C}$ and the thaw temperature to $0\text{ }^{\circ}\text{C}$. ML-Opt shows even further improvement with a surface runoff peak that closely matches the shape and magnitude of the observed river discharge until middle of June. This is achieved by the combined impact of the vertical discretization, second snow metamorphism, surface conductivity and freeze temperature changes. On the extreme end of the spectrum, the ML-T10 experiment shows too high surface runoff, by setting the freeze and thaw temperatures to an unrealistically high level (at around $+10\text{ }^{\circ}\text{C}$), consequently partitioning much of the runoff into surface runoff.

The soil temperature changes are in agreement with the surface runoff behavior described above. The soil is warmest in ML, while it gradually gets colder by the ECLand modifications (Figure 5f). The soil is coldest for ML-Opt and the two extreme temperature experiments, ML-T-10 and ML-T10, during the main snowmelt season in March–May. The cooling soil means, the temperature will be more and more likely closer or below the freeze temperature threshold (T_{Fr}), which progressively decreases the infiltration into the soil. This leads to generally reduced soil water content (Figure 5e) and decreased subsurface runoff (Figure 5d), although with large variability depending on the actual parametrization changes and the time of year.

The kge and the three component scores (inset tables in Figure 5a,b) confirm the gradually improving behavior of these experiments from ML to ML-Opt, through the introduction of the incremental ECLand parametrization changes. The $bias$ improves from -0.17 to -0.06 , var from -0.66 to -0.17 , $pcorr$ from 0.51 to 0.71 and finally the kge from 0.16 to 0.75 , representing a very large jump in skill.

3.3.2. Impact of the ECLand Experiments in Permafrost

The series of ECLand experiments are shown to provide widespread improvement for river discharge in the permafrost sensitivity area, based on all available catchments. They result mainly in reduced bias and variability errors, with optimal performance achieved by the ML-Opt experiment, including the combined changes of vertical discretization, second snow metamorphism, surface conductivity and freeze temperature.

The 69 catchments, found in the sensitivity area defined in the 60–80 N latitude belt over the permafrost (Figure 1), were collectively analyzed to see if the conclusions obtained from the test catchment in Siberia can be generalized to the permafrost. Figure 6 provides a visual summary of the overall performance in simulating river discharge, averaged over the sensitivity area, where the default configuration ML performs unfavorably. Each experiment is represented by a dot in a 3-dimensional graph showing $bias$ (x-axis), var (y-axis) and $pcorr$ (symbol size).

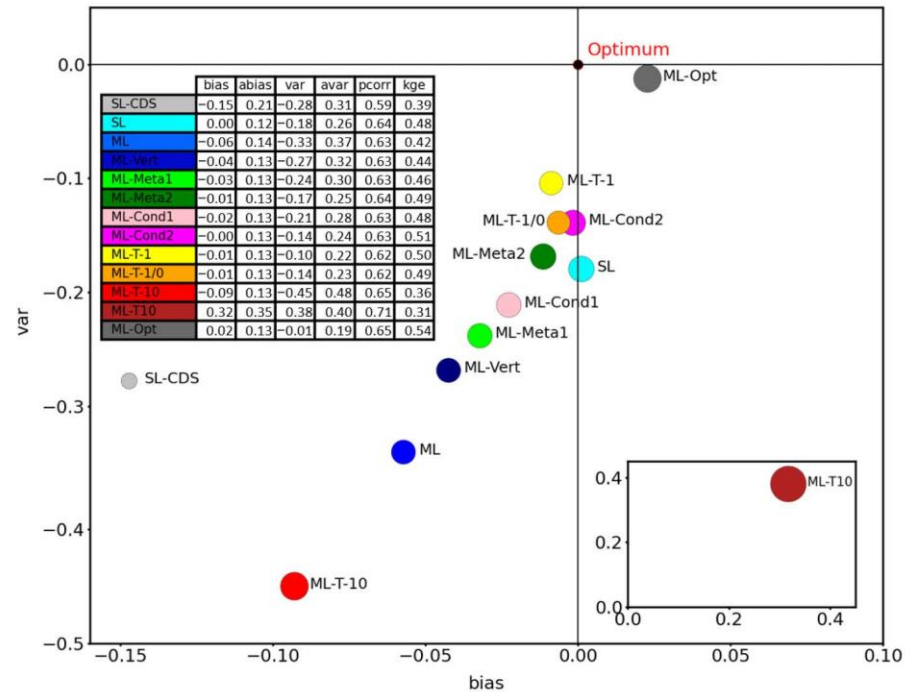


Figure 6. Scatter plot of area average scores for the 13 analyzed experiments. The scores are computed from 69 catchments in the 60–80 N belt of the permafrost (see Figure 1 for the area). The size of the dots represents *pcorr*. The outlier ML-T10 is displayed in an inset graph for better readability. The area average score values are provided for all analyzed experiments in the inset table.

The ECLand parametrization changes are effective to improve the simulation of river flow in the permafrost area, in particular the bias and variability errors, which improve from ML-T-10 to ML-Opt (see also the average scores of 69 catchments in an inset table in Figure 6). The correlation is less impacted by the modifications, the difference between all experiments is mostly within a few percent (the two extremes have average correlation of 0.59 (SL-CDS) and 0.71 (ML-T10)).

The furthest from the optimum $[0; 0]$ *bias/var* point are the ML-T-10 and ML-T10 extreme temperature experiments. Nevertheless, the highest average correlation is achieved with the positive extreme, when the soil remains mostly frozen. The fact that an unrealistically extreme-setup simulation, with excess surface runoff, produces the highest correlation shows that the current land-surface process representation in ECLand is still suboptimal in the permafrost, even after the parametrization changes. Another outlier is the online produced SL-CDS, which has lower scores, especially the very low negative *bias*, due to the snow data assimilation removing water from the rivers in the northern latitudes, as a compensation for the slow snowmelt in SLS [39].

Overall, there is a notable improvement achieved by the ECLand experiments, shown by the increase in *kge* from 0.42 (ML) to 0.54 (ML-Opt), while the single-layer experiments highlight lower *kge* of 0.48 (SL) and 0.39 (SL-CDS).

3.3.3. Global Impact of the ECLand Experiments

The changes in the snow and soil parametrization schemes in ECLand could improve the river discharge in the permafrost sensitivity area in eastern Siberia and Alaska. However,

in other areas these changes are dominantly suboptimal, as the difference between ML-Opt, the highest performing ECLand experiment, and ML highlights (Figure 7).

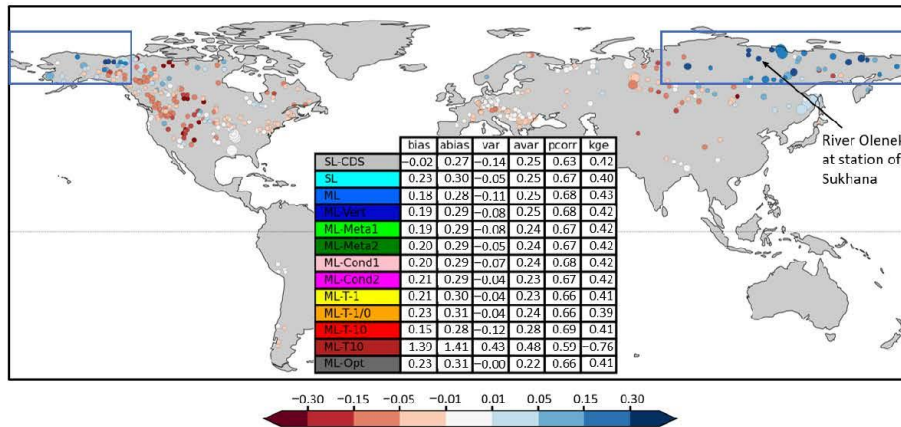


Figure 7. Difference of the modified Kling–Gupta efficiency (kge) between ML-Opt and ML. Blue areas show better performance in ML-Opt, while red indicates better performance in ML. The sensitivity area in the permafrost is shown by blue rectangles, while the test catchment on river Olenek at station Sukhana is indicated by black arrow. The area average scores are provided for all analyzed experiments in the inset table.

The worst ML-Opt performance is found in the western part of USA, Canada and catchments in central Asia, directly in those areas where ML could improve the most on SL (see Figure 2). Interestingly, the Canadian Arctic regions (Nunavut) do not show similarly improving results to other permafrost regions. One source of error, which can contribute to this, is the underlying physiographic data used to drive the simulations, e.g., lake cover. Boussetta et al. [21] showed that recent lake cover datasets are substantially different from those used in this work for the Nunavut region. The use of more realistic physiographic data will be explored in future work. Furthermore, unrepresented physical processes in the land-surface model can affect the hydrological cycle of these regions, e.g., wind-driven sublimation of the snowpack which can change the amount of snow mass available for melting at the end of the season. Similarly, the handling of frozen lakes with, e.g., snow not accumulating on top of lake ice in ECLand, may introduce compensating errors with the changes tested in ML-Opt.

Nevertheless, even though only about one third of the stations show improvement by ML-Opt, the globally averaged scores show that ML is only slightly better (inset table in Figure 7). Moreover, the score variability amongst all the experiments, in general, is quite small as well. The only real exception is ML-T10 that has very low scores throughout. The geographical maps of the kge differences between the experiments (from ML-Vert to ML-T10) and ML is provided in the Supplementary Materials (Figures S1–S9).

The deterioration of ML-Opt and the other ECLand experiments in milder climate areas is related to the surface runoff increase during the snowmelt period, demonstrated earlier. Especially in catchments where the *bias* in ML (and similarly in SL) was originally positive, e.g., around the Rockies (as suggested by Zsoter et al. [39]), any further increase in snowmelt-related surface runoff will dominantly be detrimental and contribute to further increased *bias* and thus lower kge .

4. Discussion

It was shown in this study that the MLS has an improved representation of the hydrology, with the notable exception of the coldest areas in permafrost, where the SLS is

superior. To improve the MLS hydrological representation in permafrost, modifications of the ECLand snow and soil freezing schemes were tested. It was shown that a series of incremental changes could noticeably improve the quality of the river discharge simulation over a large area in permafrost, primarily through decreasing the soil temperature and thus increasing the amount of surface runoff in the critical spring snowmelt period.

Improving the hydrological process representation

The results have demonstrated that the use of uniform parameters in ECLand in the snow and soil freezing schemes, currently applied in ECLand, are too simplistic and will not work for both the permafrost and non-permafrost areas in the snow-impacted climate. Spatially variable parametrization for variables, such as c_s in the destructive metamorphism of the snow, l_b in the snow-soil thermal conductivity and T_{Fz} freeze temperature in the soil freezing scheme, explored in this work, promise to bring a more balanced approach for delivering improved hydrological process representation.

However, this could technically be complex to implement in ECLand and needs a substantial amount of further research. As a first step, the parametrization part of ECLand has recently been refactored (by removing hard-coded parameter values) and the Multiscale Parameter Regionalization (MPR; [52]) has been implemented for estimation of spatially varying parameters. This change will make it easier in the future to work on aspects of ECLand, such as the one explored in this study for the snow and soil parametrization impact on hydrology.

This study has also contributed to the understanding of the hydrological importance of each tested change in the ECLand parametrization, which could be based on the average verification metrics in the permafrost (distance between dots in Figure 6), supported by the land-surface processes representation at the example catchment (Figure 5).

The thaw temperature adjustment from +1 to 0 brought deterioration, while all other changes could improve river discharge. The smallest improvement seems to come from the snow-soil thermal conductivity adjustment (updated l_b), closely followed by the first snow metamorphism change (variable c_s). The snow vertical discretization change looks to bring larger improvements (adjustment over complex terrain), while the second snow metamorphism change (uniform c_s) is even more beneficial in the permafrost. Finally, the biggest contribution appears to come from the adjustment of the lower temperature threshold in the soil freezing scheme (freeze temperature) from -3 to -1 °C.

Land-surface modelling challenges

Due to the presence of a unique soil column within a grid cell in EC-Land, the soil temperature may be largely affected by the repartitioning of the sub-grid tiled surfaces, as these can be characterized by different surface energy fluxes. In high latitude during transition seasons (e.g., snow accumulation and ablation), the tile fraction subdivision depends on the snow cover fraction, exerting a large control on the heat flux conducted through the soil ([19,53,54]). When the snow cover fraction is less than one, part of the soil directly interacts with the atmosphere above, without the thermal insulation effect of the snowpack. This in turn, may affect the soil temperature and therefore the fraction of frozen water within the soil. Future modelling work should evaluate the hydrological sensitivity to snow cover fraction parametrization in global land-surface simulations (see for example [55]), considering the combined effect that this can have on the soil below and the atmosphere above.

In addition, future modelling work should also evaluate the effect of unrepresented or poorly represented hydrological processes in ECLand. These include, for example wind-driven snow sublimation, snow interception by forests, as well as of a more physically-complex representation of frozen/unfrozen water phases in the soil column, which all promise to bring benefits to the hydrological cycle, especially in permafrost regions (see for instance [10,11,44]). Moreover, the impact of a deeper soil column, with additional vertical layers, has also been shown to work better for permafrost simulation in land-surface models ([56]). More complex models for hydrological applications in high-latitude regions

have been proposed in the literature (e.g., [44,56,57]). Their implementation in ECLand could be considered in the future.

Relevance for ECMWF

The next operational IFS cycle (48r1) of ECMWF will include multi-layer snow representation. Based on this study, the slightly modified version of ML-Meta1, was selected for operationalization, which includes the snow vertical discretization and first snow metamorphism changes on top of the default ML configuration. The results of this study highlighted that this new snow model did show moderate hydrological improvements in permafrost and could still retain most of the good performance of the default MLS in other areas.

Earth system modelling implications

Modelling improvements in the Earth system process representation, such as the use of any prospective MLS version explored in this study, require thorough testing in coupled forecast mode with data assimilation, similarly as in [26]. This is necessary to check the transferability of the offline-demonstrated hydrological improvements, and make sure other variables, such as 2-metre temperature and surface fluxes of sensible and latent heat, will not deteriorate.

This study has proven that further development of the snow and soil parametrizations in ECLand is crucial to achieve better hydrological performance everywhere globally; however, it could not be part of this study and will only be explored in the future as an important research area.

The success of helping the operational development of the ECLand model's snow component has shown that hydrological studies, such as the work presented in this paper, have great potential to help improve the land-surface realism in Earth system models and can contribute to improvements in not just the hydrological variables, such as river discharge, but potentially other components of the Earth system as well.

Limitations of the study

The authors acknowledge that even with the best care taken in the experimental setup of this study, some limitations remain. The network of analyzed catchments is still under-representative in some areas, also with the relatively short minimum observation length of eight years as a compromise. Similarly, this study was not designed to be a full sensitivity experiment for the land-surface processes in the permafrost, which would have required a much higher computational cost. In addition, the choice of testing ECLand changes in MLS was because of practical reasons, but most of the modifications would be expected to show improvements in a similar manner even if tested within the SLS.

5. Conclusions

The representation of snow is a crucial aspect of land-surface modelling, as it has a strong influence on the energy and water balances. Snow schemes with multiple layers can better represent the snowpack evolution and bring improvements on single-layer schemes in simulating the snow processes and contribute to better soil freezing and hydrological cycle.

In this paper, the hydrological impact of the MLS, implemented in ECLand, was analyzed globally with ERA5-forced experiments over the period of 1979–2018. The CaMa-Flood model was used to generate river discharge from the ECLand runoff output, allowing direct comparison with gauged observations over more than 400 snow-impacted catchments. Different sensitivity experiments were conducted to evaluate the impact of changes in the ECLand snow and soil freezing schemes on the terrestrial hydrological processes, with particular focus on permafrost.

It was found that while the default MLS generally improves the river discharge simulation, mainly through better bias and variability errors, the performance is suboptimal in large parts of the high latitude permafrost regions. The analysis of the climatological mean time series, in a test catchment in Siberia, demonstrated that the largely underestimated

snowmelt-driven floods in late spring to early summer are caused by too low surface runoff. The MLS better insulates the soil, which allows more water to infiltrate into the soil and thus surface runoff is reduced.

It was also found that the ECLand experiments provide widespread improvement for river discharge in the sensitivity area, defined in permafrost. The incremental changes of the snow vertical discretization, destructive metamorphism, snow-soil thermal conductivity and soil freeze temperature lead to gradually colder soil, which resulted in increased surface runoff and thus better river discharge simulation during the critical snowmelt-driven flood period. The ML-Opt experiment, as the combination of the best performing ECLand changes, has shown higher overall k_{ge} , mainly through reduced bias and variability errors.

The results, presented here, have directly influenced the MLS version that will be introduced in the next Integrated Forecasting System cycle of ECMWF (48r1). This has demonstrated that hydrological analyses, such as the work presented in this paper, can provide a useful platform to diagnose areas where improvements are needed in the land-surface representation of the Earth system models.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/atmos13050727/s1>, Figure S1: Difference of the Kling-Gupta efficiency (k_{ge}) between ML-Vert, the experiment with modified snow vertical discretisation over complex terrain with at least 25 cm snow depth, and ML, the default multi-layer experiment, across all 453 stations, calculated on daily river discharge over 1980–2018. Improvements in ML-Vert are indicated by blue dots. Size of the dots represent the catchment area; Figure S2: As Figure S1, but k_{ge} difference between ML-Meta1 (the 1st modification of the destructive metamorphism of the snow with variable c_{ξ} parameter values across the snow layers, together with ML-Vert) and ML; Figure S3: As Figure S1, but k_{ge} difference between ML-Meta2 (the 2nd modification of the destructive metamorphism of the snow with one c_{ξ} parameter value as in Anderson et al., 1976, together with ML-Vert) and ML; Figure S4: As Figure S1, but k_{ge} difference between ML-Cond1 (ML-Meta1 together with the snow-soil thermal conductivity computation with the revised parameter $l_b = 0.5$) and ML; Figure S5: As Figure S1, but k_{ge} difference between ML-Cond2 (ML-Meta2 together with the snow-soil thermal conductivity computation with the revised parameter $l_b = 0.5$) and ML; Figure S6: As Figure S1, but k_{ge} difference between ML-T-1 (change of the soil freeze parameter T_{Fr} from the default -3 to -1 °C added onto ML-Meta1) and ML; Figure S7: As Figure S1, but k_{ge} difference between ML-T-1/0 (change of the freeze temperature T_{Fr} to -1 °C and thaw temperature T_{Th} from the default $+1$ to 0 °C, added onto ML-Meta1) and ML; Figure S8: As Figure S1, but k_{ge} difference between ML-T-10 ($[T_{Fr}, T_{Th}]$ changed to $[-10.5, -10]$, added onto ML-Meta1) and ML; Figure S9: As Figure S1, but k_{ge} difference between ML-T10 ($[T_{Fr}, T_{Th}]$ changed to $[+10, +10.5]$, added onto ML-Meta1) and ML.

Author Contributions: Conceptualization, E.Z., G.A., C.P., E.S. and H.C.; Data curation, E.Z. and G.A.; Methodology, E.Z., G.A., C.P., E.S. and H.C.; Visualization, E.Z.; Writing—original draft, E.Z.; Writing—review and editing, E.Z., G.A., C.P., E.S. and H.C. All authors have read and agreed to the published version of the manuscript.

Funding: Ervin Zsoter's PhD is supported by the Wilkie Calvert Co-Supported PhD Studentships at the University of Reading. Christel Prudhomme and Ervin Zsoter were supported by the Copernicus Emergency Management Service—Early Warning Systems (CEMS-EWS). Hannah Cloke is supported by the EVOFLOOD project: The Evolution of Global Flood Risk UK NERC, NE/S015590/1. Elisabeth Stephens and Hannah Cloke are supported by the FATHUM project: Forecasts for Anticipatory HUMANitarian Action funded by UK NERC as part of their Science for Humanitarian Emergencies & Resilience (SHEAR) program, NE/P000525/1.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We are grateful to The Global Runoff Data Centre, 56068 Koblenz, Germany for providing observations for our river discharge analysis.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Fisher, R.A.; Koven, C.D. Perspectives on the Future of Land Surface Models and the Challenges of Representing Complex Terrestrial Systems. *J. Adv. Model. Earth Syst.* **2020**, *12*, e2018MS001453. [\[CrossRef\]](#)
2. Mengelkamp, H.-T.; Warrach, K.; Ruhe, C.; Raschke, E. Simulation of runoff and streamflow on local and regional scales. *Meteor. Atmos. Phys.* **2001**, *76*, 76–107. [\[CrossRef\]](#)
3. Prentice, I.C.; Liang, X.; Medlyn, B.E.; Wang, Y.-P. Reliable, robust and realistic: The three R's of next-generation land-surface modelling. *Atmos. Chem. Phys.* **2015**, *15*, 5987–6005. [\[CrossRef\]](#)
4. Bouilloud, L.; Chancibault, K.; Vincendon, B.; Ducrocq, V.; Habets, F.; Saulnier, G.-M.; Anquetin, S.; Martin, E.; Noilhan, J. Coupling the ISBA Land Surface Model and the TOPMODEL Hydrological Model for Mediterranean Flash-Flood Forecasting: Description, Calibration, and Validation. *J. Hydrometeorol.* **2010**, *11*, 315–333. [\[CrossRef\]](#)
5. Le Vine, N.; Butler, A.; McIntyre, N.; Jackson, C. Diagnosing hydrological limitations of a land surface model: Application of JULES to a deep-groundwater chalk basin. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 143–159. [\[CrossRef\]](#)
6. Harrigan, S.; Zsoter, E.; Alfieri, L.; Prudhomme, C.; Salamon, P.; Wetterhall, F.; Barnard, C.; Cloke, H.; Pappenberger, F. GloFAS-ERA5 operational global river discharge reanalysis 1979–present. *Earth Syst. Sci. Data* **2020**, *12*, 2043–2060. [\[CrossRef\]](#)
7. Overgaard, J.; Rosbjerg, D.; Butts, M.B. Land-surface modelling in hydrological perspective—A review. *Biogeosciences* **2006**, *3*, 229–241. [\[CrossRef\]](#)
8. Dutra, E.; Viterbo, P.; Miranda, P.M.; Balsamo, G. Complexity of snow schemes in a climate model and its impact on surface energy and hydrology. *J. Hydrometeorol.* **2012**, *13*, 521–538. [\[CrossRef\]](#)
9. Kauffeldt, A.; Halldin, S.; Pappenberger, F.; Wetterhall, F.; Xu, C.-Y.; Cloke, H.L. Imbalanced land surface water budgets in a numerical weather prediction system. *Geophys. Res. Lett.* **2015**, *42*, 4411–4417. [\[CrossRef\]](#)
10. Koren, V.; Smith, M.; Cui, Z. Physically-based modifications to the Sacramento Soil Moisture Accounting model. Part A: Modeling the effects of frozen ground on the runoff generation process. *J. Hydrol.* **2014**, *519*, 3475–3491. [\[CrossRef\]](#)
11. Krogh, S.A.; Pomeroy, J.W.; Marsh, P. Diagnosis of the hydrology of a small Arctic basin at the tundra-taiga transition using a physically based hydrological model. *J. Hydrol.* **2017**, *550*, 685–703. [\[CrossRef\]](#)
12. López-Moreno, J.I.; García-Ruiz, J.M. Influence of snow accumulation and snowmelt on streamflow in the Central Spanish Pyrenees. *Hydrol. Sci. J.* **2004**, *49*, 787–802. [\[CrossRef\]](#)
13. Griessinger, N.; Seibert, J.; Magnusson, J.; Jonas, T. Assessing the benefit of snow data assimilation for runoff modeling in Alpine catchments. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 3895–3905. [\[CrossRef\]](#)
14. Boone, A.; Etchevers, P. An intercomparison of three snow schemes of varying complexity coupled to the same land surface model: Local-scale evaluation at an Alpine site. *J. Hydrometeorol.* **2001**, *2*, 374–394. [\[CrossRef\]](#)
15. Best, M.J.; Pryor, M.; Clark, D.B.; Rooney, G.G.; Essery, R.L.H.; Ménard, C.B.; Edwards, J.M.; Hendry, M.A.; Porson, A.; Gedney, N.; et al. The Joint UK Land Environment Simulator (JULES), Model description—Part 1: Energy and water fluxes. *Geosci. Model Dev.* **2011**, *4*, 677–699. [\[CrossRef\]](#)
16. Decharme, B.; Brun, E.; Boone, A.; Delire, C.; Le Moigne, P.; Morin, S. Impacts of snow and organic soils parameterization on northern Eurasian soil temperature profiles simulated by the ISBA land surface model. *Cryosphere* **2016**, *10*, 853–877. [\[CrossRef\]](#)
17. Dutra, E.; Balsamo, G.; Viterbo, P.; Miranda, P.M.A.; Beljaars, A.; Schär, C.; Elder, K. An improved snow scheme for the ECMWF land surface model: Description and offline validation. *J. Hydrometeorol.* **2010**, *11*, 899–916. [\[CrossRef\]](#)
18. Wang, T.; Ottlé, C.; Boone, A.; Ciais, P.; Brun, E.; Morin, S.; Krinner, G.; Piao, S.; Peng, S. Evaluation of an improved intermediate complexity snow scheme in the ORCHIDEE land surface model. *J. Geophys. Res. Atmos.* **2013**, *118*, 6064–6079. [\[CrossRef\]](#)
19. Slater, A.G.; Schlosser, C.A.; Desborough, C.E.; Pitman, A.; Henderson-Sellers, A.; Robock, A.; Ya Vinnikov, K.; Entin, J.; Mitchell, K.; Chen, F.; et al. The representation of snow in land surface schemes: Results from PILPS 2(d). *J. Hydrometeorol.* **2001**, *2*, 7–25. [\[CrossRef\]](#)
20. Balsamo, G.; Beljaars, A.; Scipal, K.; Viterbo, P.; van den Hurk, B.; Hirschi, M.; Betts, A.K. A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System. *J. Hydrometeorol.* **2009**, *10*, 623–643. [\[CrossRef\]](#)
21. Boussetta, S.; Balsamo, G.; Arduini, G.; Dutra, E.; McNorton, J.; Choulga, M.; Agustí-Panareda, A.; Beljaars, A.; Wedi, N.; Muñoz-Sabater, J.; et al. ECLand: The ECMWF Land Surface Modelling System. *Atmosphere* **2021**, *12*, 723. [\[CrossRef\]](#)
22. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horanyi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 Global Reanalysis. *Q. J. R. Meteor. Soc.* **2020**, *146*, 1999–2049. [\[CrossRef\]](#)
23. Muñoz-Sabater, J.; Dutra, E.; Agustí-Panareda, A.; Albergel, C.; Arduini, G.; Balsamo, G.; Boussetta, S.; Choulga, M.; Harrigan, S.; Hersbach, H.; et al. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* **2021**, *13*, 4349–4383. [\[CrossRef\]](#)
24. Saha, S.K.; Sujith, K.; Pokhrel, S.; Chaudhari, H.S.; Hazra, A. Effects of multilayer snow scheme on the simulation of snow: Offline Noah and coupled with NCEP CFS v2. *J. Adv. Model. Earth Syst.* **2017**, *9*, 271–290. [\[CrossRef\]](#)

25. Vionnet, V.; Brun, E.; Morin, S.; Boone, A.; Faroux, S.; Le Moigne, P.; Martin, E.; Willemet, J.-M. The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2. *Geosci. Model Dev.* **2012**, *5*, 773–791. [[CrossRef](#)]
26. Arduini, G.; Balsamo, G.; Dutra, E.; Day, J.J.; Sandu, I.; Bousssetta, S.; Haiden, T. Impact of a multi-layer snow scheme on near-surface weather forecasts. *J. Adv. Model. Earth Syst.* **2019**, *11*, 4687–4710. [[CrossRef](#)]
27. Walters, D.; Baran, A.J.; Boutle, I.; Brooks, M.; Earnshaw, P.; Edwards, J.; Furtado, K.; Hill, P.; Lock, A.; Manners, J.; et al. The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations. *Geosci. Model Dev.* **2019**, *12*, 1909–1963. [[CrossRef](#)]
28. Magnusson, J.; Wever, N.; Essery, R.; Helbig, N.; Winstral, A.; Jonas, T. Evaluating snow models with varying process representations for hydrological applications. *Water Resour. Res.* **2015**, *51*, 2707–2723. [[CrossRef](#)]
29. Decharme, B.; Delire, C.; Minvielle, M.; Colin, J.; Vergnes, J.; Alias, A.; Saint-Martin, D.; Séférian, R.; Sénési, S.; Voldoire, A. Recent changes in the ISBA-CTRIIP land surface system for use in the CNRM-CM6 climate model and in global off-line hydrological applications. *J. Adv. Model. Earth Syst.* **2019**, *11*, 1207–1252. [[CrossRef](#)]
30. Romanovsky, V.E.; Burgess, M.; Smith, S.; Yoshikawa, K.; Brown, J. Permafrost temperature records: Indicators of climate change. *Eos Trans. Am. Geophys. Union* **2002**, *83*, 589–594. [[CrossRef](#)]
31. Koven, C.D.; Riley, W.J.; Stern, A. Analysis of permafrost thermal dynamics and response to climate change in the CMIP5 earth system models. *J. Clim.* **2013**, *26*, 1877–1900. [[CrossRef](#)]
32. Andresen, C.G.; Lawrence, D.M.; Wilson, C.J.; McGuire, A.D.; Koven, C.; Schaefer, K.; Jafarov, E.; Peng, S.; Chen, X.; Gouttevin, I.; et al. Soil moisture and hydrology projections of the permafrost region—A model intercomparison. *Cryosphere* **2020**, *14*, 445–459. [[CrossRef](#)]
33. Yokohata, T.; Saito, K.; Takata, K.; Nitta, T.; Satoh, Y.; Hajima, T.; Sueyoshi, T.; Iwahana, G. Model improvement and future projection of permafrost processes in a global land surface model. *Prog. Earth Planet. Sci.* **2020**, *69*. [[CrossRef](#)] [[PubMed](#)]
34. Gouttevin, I.; Krinner, G.; Ciais, P.; Polcher, J.; Legout, C. Multi-scale validation of a new soil freezing scheme for a land-surface model with physically-based hydrology. *Cryosphere* **2012**, *6*, 407–430. [[CrossRef](#)]
35. Yamazaki, D.; Kanai, S.; Kim, H.; Oki, T. A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resour. Res.* **2011**, *47*, W04501. [[CrossRef](#)]
36. Balsamo, G.; Pappenberger, F.; Dutra, E.; Viterbo, P.; van den Hurk, B. A revised land hydrology in the ECMWF model: A step towards daily water flux prediction in a fully-closed water cycle. *Hydrol. Proc.* **2011**, *25*, 1046–1054. [[CrossRef](#)]
37. Agustí-Panareda, A.; Balsamo, G.; Beljaars, A. Impact of improved soil moisture on the ECMWF precipitation forecast in West Africa. *Geophys. Res. Lett.* **2010**, *37*, L20808. [[CrossRef](#)]
38. Haddeland, I.; Clark, D.B.; Franssen, W.; Ludwig, F.; Voß, F.; Arnell, N.W.; Bertrand, N.; Best, M.; Folwell, S.; Gerten, D.; et al. Multimodel estimate of the global terrestrial water balance: Setup and first results. *J. Hydrometeorol.* **2011**, *12*, 869–884. [[CrossRef](#)]
39. Zsoter, E.; Cloke, H.; Stephens, E.; de Rosnay, P.; Muñoz-Sabater, J.; Prudhomme, C.; Pappenberger, F. How well do operational Numerical Weather Prediction configurations represent hydrology? *J. Hydrometeorol.* **2019**, *20*, 1533–1552. [[CrossRef](#)]
40. Emerton, R.; Cloke, H.; Stephens, E.; Zsoter, E.; Woolnough, S.; Pappenberger, F. Complex picture for likelihood of ENSO-driven flood hazard. *Nat. Commun.* **2017**, *8*, 14796. [[CrossRef](#)]
41. Dottori, F.; Szewczyk, W.; Ciscar, J.-C.; Zhao, F.; Alfieri, L.; Hirabayashi, Y.; Bianchi, A. Increased Human and Economic Losses from River Flooding with Anthropogenic Warming. *Nat. Clim. Chang.* **2018**, *8*, 781–786. [[CrossRef](#)]
42. Anderson, E.A. *A Point Energy and Mass Balance Model of a Snow Cover*; Stanford University: Stanford, CA, USA, 1976; 150p.
43. Viterbo, P.; Beljaars, A.; Mahfouf, J.F.; Teixeira, J. The representation of soil moisture freezing and its impact on the stable boundary layer. *Q. J. R. Meteorol. Soc.* **1999**, *125*, 2401–2426. [[CrossRef](#)]
44. Niu, G.Y.; Yang, Z.L. Effects of frozen soil on snowmelt runoff and soil water storage at a continental scale. *J. Hydrometeorol.* **2006**, *7*, 937–952. [[CrossRef](#)]
45. Lehner, B.; Reidy Liermann, C.; Revenga, C.; Vörösmarty, C.; Fekete, B.; Crouzet, P.; Döll, P.; Endejan, M.; Frenken, L.; Magome, J.; et al. High-resolution mapping of the world’s reservoirs and dams for sustainable river-flow management. *Front. Ecol. Environ.* **2011**, *9*, 494–502. [[CrossRef](#)]
46. Lehner, B.; Döll, P. Development and validation of a global database of lakes, reservoirs and wetlands. *J. Hydrol.* **2004**, *296*, 1–22. [[CrossRef](#)]
47. Gupta, H.V.; Kling, H.; Yilmaz, K.K.; Martinez, G.F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* **2009**, *377*, 80–91. [[CrossRef](#)]
48. Kling, H.; Fuchs, M.; Paulin, M. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.* **2012**, *424–425*, 264–277. [[CrossRef](#)]
49. Knoben, W.J.M.; Freer, J.E.; Woods, R.A. Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 4323–4331. [[CrossRef](#)]
50. Lin, P.; Pan, M.; Beck, H.E.; Yang, Y.; Yamazaki, D.; Frasson, R.; David, C.H.; Durand, M.; Pavelsky, T.M.; Allen, G.H.; et al. Global Reconstruction of Naturalized River Flows at 2.94 Million Reaches. *Water Resour. Res.* **2019**, *55*, 6499–6516. [[CrossRef](#)]
51. Cao, B.; Gruber, S.; Zheng, D.; Li, X. The ERA5-Land soil temperature bias in permafrost regions. *Cryosphere* **2020**, *14*, 2581–2595. [[CrossRef](#)]
52. Schweppe, R.; Thober, S.; Kelbling, M.; Kumar, R.; Attinger, S.; Samaniego, L. MPR 1.0: A stand-alone Multiscale Parameter Regionalization Tool for Improved Parameter Estimation of Land Surface Models. *Geosci. Model Dev.* **2021**, *15*, 859–882. [[CrossRef](#)]

53. Yang, Z.L.; Dickinson, R.E.; Robock, A.; Vinnikov, K.Y. Validation of the snow submodel of the biosphere–atmosphere transfer scheme with Russian snow cover and meteorological observational data. *J. Clim.* **1997**, *10*, 353–373. [[CrossRef](#)]
54. Bélair, S.; Brown, R.; Mailhot, J.; Bilodeau, B.; Crevier, L.P. Operational implementation of the ISBA land surface scheme in the Canadian regional weather forecast model. Part II: Cold season results. *J. Hydrometeorol.* **2003**, *4*, 371–386. [[CrossRef](#)]
55. Niu, G.Y.; Yang, Z.L. An observation-based formulation of snow cover fraction and its evaluation over large North American river basins. *J. Geophys. Res. Atmos.* **2007**, *112*. [[CrossRef](#)]
56. Melton, J.R.; Verseghy, D.L.; Sospedra-Alfonso, R.; Gruber, S. Improving permafrost physics in the coupled Canadian Land Surface Scheme (v.3.6.2) and Canadian Terrestrial Ecosystem Model (v.2.1) (CLASS-CTEM). *Geosci. Model Dev.* **2019**, *12*, 4443–4467. [[CrossRef](#)]
57. Clark, M.P.; Zolfaghari, R.; Green, K.R.; Trim, S.; Knoben, W.J.; Bennett, A.; Spiteri, R.J. The numerical implementation of land models: Problem formulation and laugh tests. *J. Hydrometeorol.* **2021**, *22*, 1627–1648. [[CrossRef](#)]

A3: Trends in the GloFAS-ERA5 river discharge reanalysis

This paper presents the published version of chapter 6 of this thesis, with the following reference:

Zsoter, E., H. L. Cloke, C. Prudhomme, S. Harrigan, P. de Rosnay, J. Munoz-Sabater and E. Stephens, 2020b: Trends in the GloFAS-ERA5 river discharge reanalysis, *ECMWF Technical Memoranda*, 871, Technical Report, ECMWF, doi:10.21957/p9jrh0xp

Technical Memo



871

Trends in the GloFAS-ERA5 river discharge reanalysis

E. Zsoter^{1,2}, H. Cloke^{2,3,4}, C. Prudhomme¹, S. Harrigan¹, P. de Rosnay¹, J. Muñoz-Sabater¹, E. Stephens²

- ¹ ECMWF Forecast Department
- ² University of Reading
- ³ Uppsala University
- ⁴ Centre of Natural Hazards and Disaster Science (CNDS)

September 2020

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our website under:

<http://www.ecmwf.int/en/publications>

Contact: library@ecmwf.int

© Copyright 2020

European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, UK

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. This publication is not to be reprinted or translated in whole or in part without the written permission of the Director-General. Appropriate non-commercial use will normally be granted under the condition that reference is made to ECMWF.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error or omission or for loss or damage arising from its use.

Abstract

The main objective of this study is to analyse the GloFAS-ERA5 river discharge reanalysis for any noticeable change (including gradual trends or discontinuities) in the annual mean time series across the 1979-2018 (40-year) period, and to evaluate how realistic these are compared with available observed river discharge time series.

These variabilities are quantified by linear regression in order to highlight any concerning features in the GloFAS-ERA5 time series.

This work is particularly important for GloFAS, as large trends, discontinuities or other similar features could have a major consequence on the GloFAS flood thresholds in around 50% of catchments, which are based on GloFAS-ERA5, and thus subsequently on the issuing of flood warnings.

In addition, this study also contributes to the understanding of the water cycle variable behaviour in ERA5 (driver of GloFAS-ERA5) and ERA5-Land (higher resolution land reanalysis forced by ERA5, produced offline) by exploring the linear trends in river discharge and related hydrological variables. In exploring the stability of the time series in ERA5, we seek to trigger potential further discussions and research studies, which subsequently should help with the planning and development for the next generation ECMWF reanalysis, ERA6.

1 Introduction

In November 2019, the GloFAS (Global Flood Awareness System) was upgraded to version 2.1 (<https://confluence.ecmwf.int/display/COPSRV/GloFAS+v2.1>). This included two important changes: the release of the official GloFAS-ERA5 river discharge reanalysis, and the revision of the flood threshold computation. The flood thresholds are predefined and specified as a magnitude of river flow for specific return periods; these are used operationally to highlight where upcoming flows may be severe and trigger alerts accordingly. They are computed from river discharge reanalysis by fitting an extreme value distribution onto the annual maxima time series. In the v2.1 upgrade, flood thresholds were recomputed using the 40-year (1979-2018) GloFAS-ERA5 river discharge reanalysis, which is an ERA5-forced hydrological simulation (Harrigan et al. 2020). In addition, the analysis of the v2.1 vs. v2.0 thresholds revealed that over large parts of the world the GloFAS-ERA5 river discharge time series has very noticeable linear trends across the 40-year period.

Linear trends can highlight noticeable change across the 40-year period, be that a gradual shift (i.e. a trend) or a discontinuity (i.e. a step change at one point in the time series). Any noticeable shift in the time series is particularly important as it can hinder the representativity of the GloFAS thresholds, through the characteristically different extreme flood event behaviour in different parts of the 40-year period.

In this study, the changes/shifts in the GloFAS-ERA5 annual mean time series are analysed and quantified by linear regression. The linear trend magnitudes, along the regression lines, are computed for river discharge, as well as for all ERA5 and ERA5-Land variables that directly affect the water budget: precipitation, snowfall, evaporation, 2m temperature, soil water content, runoff and snowmelt. The linear trends in the available river discharge observations are used as verifying truth and compared with the GloFAS-ERA5 river discharge (and also ERA5 runoff) trends.

2 Data and methods

2.1 Global Flood Awareness System

The Global Flood Awareness System (GloFAS; www.globalfloods.eu) is part of the Copernicus Emergency Management Service (CEMS) and has been developed in collaboration between ECMWF, the Joint Research Centre (JRC) of the European Commission with help from research institutions such as the University of Reading (UoR). It monitors and forecasts floods across the world. GloFAS has two complementary systems:

- GloFAS 30-day, that includes daily ensemble flood forecast predictions, out to 30 days ahead, updated daily, based on the ECMWF medium- and extended-range ensemble runoff as input forcing
- GloFAS Seasonal, that provides ensemble hydrological forecasts of unusually low or high flow for calendar weeks up to 16 weeks ahead, updated monthly, based on the ECMWF SEAS5 ensemble runoff as input forcing

GloFAS forecasts possible flood episodes and unusually high/low river flow for all major rivers of the world. It has been an operational service since April 2018 (following a pre-operational phase which started in 2011) with information shown on a dedicated web platform (www.globalfloods.eu; Figure 1).

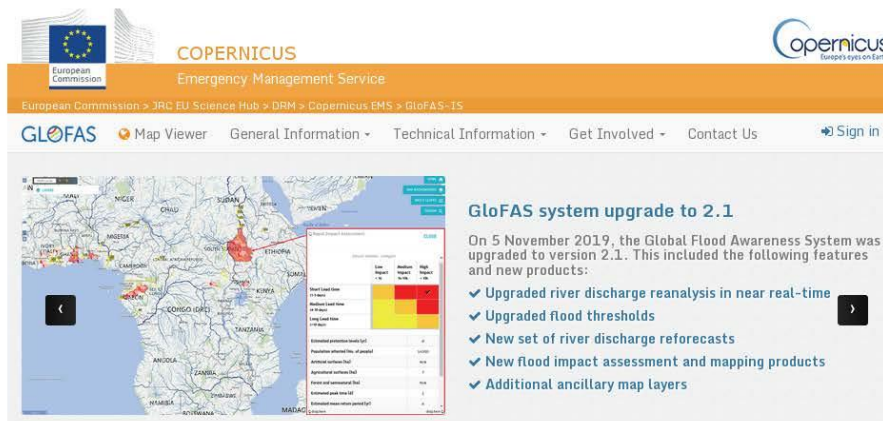


Figure 1. Snapshot of the GloFAS website's landing page.

GloFAS is designed to complement existing national and regional services, and to support national civil protection and international organisations in decision making before major flood events, particularly in large transnational river basins. Forecast information is used by a variety of decision makers, including national and regional water authorities, water resources managers, hydropower companies, civil protection and first line responders, and international humanitarian aid organisations. GloFAS only focuses on larger rivers (mainly over 10.000 km² catchment area) and does not provide real-time forecast information on flash flood risk, pluvial or coastal flooding.

GloFAS river discharge data is produced by coupling the LISFLOOD hydrological and channel routing model (van der Krijff et al., 2010) to the runoff output of the land-surface model of ECMWF NWP

system. The river routing runs with surface and sub-surface runoff inputs on a $0.1^{\circ} \times 0.1^{\circ}$ (~10 km resolution) global river network. The surface runoff component directly enters the river channel, while the sub-surface runoff first enters a groundwater module that outputs the water into the river channel after a time delay (Harrigan et al. 2020).

The river state of the real time ensemble GloFAS forecasts is initialised from a hydrological simulation, forced by the fast release version of ERA5 (ERA5T) up until when it is available (a few days behind real time) and then subsequently by the ECMWF ensemble control forecast.

A long term river discharge dataset is required in order to compute the return period flood thresholds in GloFAS 30-day: currently this is the GloFAS-ERA5 reanalysis.

Further description on the GloFAS system is available on the GloFAS internal website at (<https://confluence.ecmwf.int/display/COPSRV/Global+Flood+Awareness+System>) and in Harrigan et al 2020.

2.2 ERA5 and ERA5-Land

ERA5 is ECMWF's 5th generation global climate reanalysis (Hersbach et al., 2020). ERA5 covers the period January 1979 to near real time with a plan to extend to 1950. ERA5 includes one high-resolution component (~31 km horizontal resolution) and a lower resolution ensemble component with 10 members (~62 km horizontal resolution).

There are two flavours of ERA5 reanalysis available: the raw ERA5, based on consolidated, quality-checked data, lagging ~2-3 months behind real-time; and ERA5T, which is available 5 days behind real time, but is not fully quality-checked (Hersbach et al., 2020).

ERA5-Land (<https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land>) is also analysed for comparative purposes in this study. ERA5-Land is the offline land surface only improved version of ERA5.

ERA5 and ERA5-Land runoff are both produced by the HTESSEL (Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land; Balsamo et al., 2009) land-surface model of the ECMWF Integrated Forecasting System (IFS).

The HTESSEL scheme follows a mosaic (or tiling) approach where the grid boxes are divided into patches (or tiles), with up to six fractions over land (bare ground, low and high vegetation, intercepted water, shaded and exposed snow) and two extra tiles over water (open and frozen water) exchanging energy and water with the atmosphere. HTESSEL is used within the ECMWF IFS in coupled atmosphere-surface mode on time ranges from medium-range to seasonal forecasts.

ERA5 is a coupled application which includes the operational land data assimilation system of ECMWF to assimilate conventional in-situ and satellite observations for land surface variables for the analysis of soil moisture, soil temperature and snow fields (de Rosnay et al., 2014).

ERA5-Land, on the other hand, is a product of an offline HTESSEL simulation without atmosphere-land coupling and land data assimilation, forced by the atmospheric variables (e.g. air temperature and radiation). ERA5-Land is produced at 9 km spatial resolution using downscaled ERA5 atmospheric forcing and a vertical lapse rate correction on temperature. There is no direct coupling or land data assimilation in ERA5-Land (there is only an indirect impact through the ERA5 forcing) and this can

have a large impact on the hydrological cycle (Zsoter et al. 2019), and potentially also on the trends. Other major differences between ERA5 and ERA5-Land are the much higher resolution and the lapse-rate correction in ERA5-Land, which also can have a large impact on the water budget, especially in mountainous areas, changing the snow pack and snowmelt through the temperature differences.

ERA5-Land is analysed in a similar way to ERA5 for hydrological variables, apart from river discharge which is not yet produced for ERA5-Land. In addition, precipitation and snowfall are identical in ERA5 and ERA5-Land (apart from the downscaling) and therefore are not considered further within the ERA5-Land analysis.

2.3 GloFAS-ERA5 reanalysis

The GloFAS-ERA5 river discharge reanalysis is a product of the European Commission's Copernicus Emergency Management Service (CEMS) and is officially available in the Copernicus Climate Data Store free of charge after registration (<https://cds.climate.copernicus.eu/>).

GloFAS-ERA5 uses surface and sub-surface runoff input from the high-resolution component of the ERA5 reanalysis. It has a fast release version, GloFAS-ERA5T, forced by ERA5T, available 2-5-days behind real time, used in the initialisation of the GloFAS real time forecasts.

A schematic of the key components of the GloFAS-ERA5 and its potential extension GloFAS-ERA5-Land is given in Figure 2.

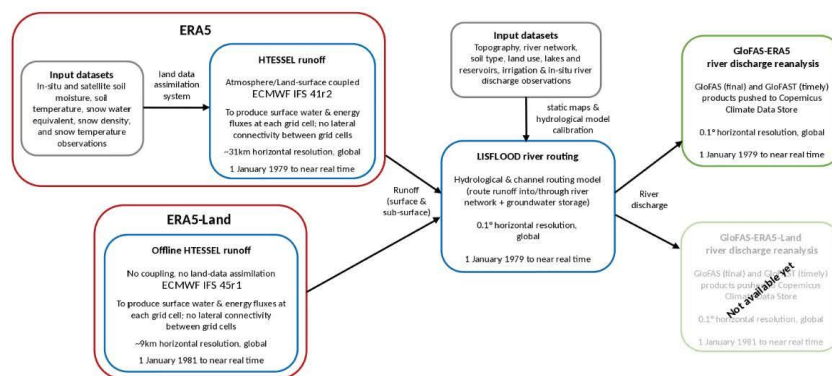


Figure 2. Schematic of the key components of the GloFAS-ERA5 river discharge reanalysis dataset production, including the ERA5 and ERA5-Land climate reanalyses. Adapted from Harrigan et al. 2020.

GloFAS verification studies often use GloFAS-ERA5, serving as a proxy for river discharge observations. More detailed information on GloFAS-ERA5 is available in Harrigan et al. (2020).

2.4 Trend analysis methodology

Although GloFAS-ERA5 starts in 1979, ERA5-Land data are only available from 1981, and thus the trend analysis is based on the common 1981-2018 period of 38 years. At the time of writing, no ERA5-Land river discharge is available, and instead annual averages of runoff are used as a proxy, which is a very good estimate of the river discharge in annual means.

Choice of catchments

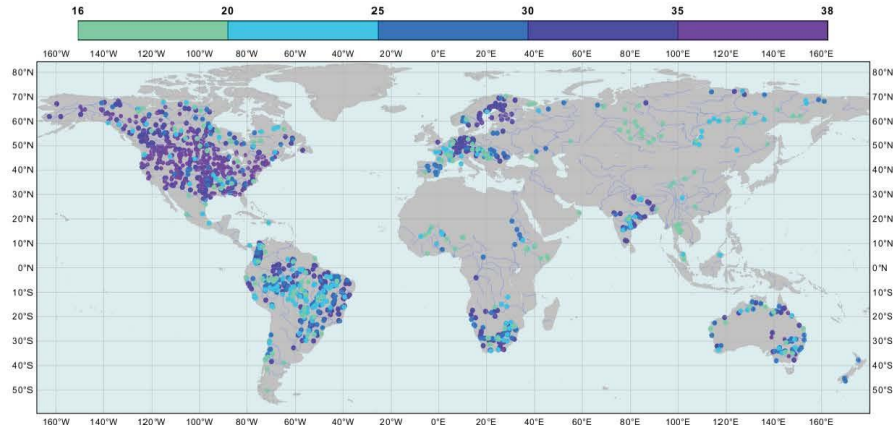


Figure 3. Length of observed river discharge records available for the trend analysis, represented at the catchment outlets. A total of 1324 stations with a minimum of 16 years was processed.

The GloFAS diagnostic catchments are used in this study to analyse the linear trends: a list of 6122 stations with catchment areas varying from about 1.000 km² to 5.4 million km².

On a subset of these catchments, where river discharge observations were available (collected and managed by the JRC), the GloFAS-ERA5 river discharge trends are compared to the equivalent trends, determined from observed river discharge, available in the 1981-2018 period (Figure 3). A total of 1324 stations were selected with a minimum of 16 years of observed daily data. Unfortunately, large parts of the world are poorly observed with large gaps in space and also in time in the 38-year period (Lavers et al., 2019; Rodda et al., 1993; Pavelsky et al., 2014)

Additionally, 33 major world river catchments with good observations are selected for providing detailed analysis of the annual mean time series and linear trend magnitudes (Figure 4). The catchments were selected to cover different parts of the world with the longest possible observation time series.

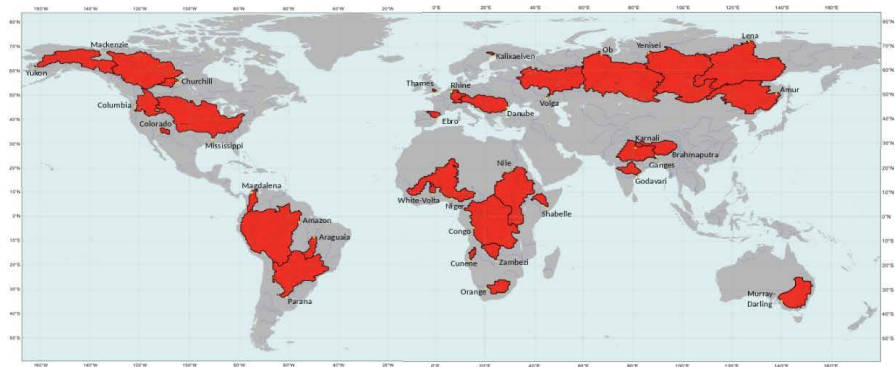


Figure 4. Global river catchments where the trends are provided in Table 2, 3 and A1, and where the hydrographs are shown in the Appendix C with the annual mean time series.

Choice of variables

In addition to the simulated and observed river discharge (Table 1), the surface variables that directly affect the water budget are analysed from ERA5 and ERA5-Land for linear trends. In the HTESSEL terrestrial water budget, precipitation (P) is the incoming water source. The water can stay in one of the water storage reservoirs or leave the land surface. Water reservoirs are the soil, the canopy interception and in solid form the snowpack (the water stored in the snowpack is the snow water equivalent, SWE). The interception accounts for only a very small fraction of the storage and thus it was left out of the analysis. From the soil, which has four layers in HTESSEL, two water reservoir versions are chosen to be analysed. The top layer (SWV7, 0-7cm), which provides an immediate impact to the atmosphere, and the combination of the top three layers (SWV100, 0-100cm) which represents the slower evolving part of the soil that is still more strongly connected to the atmosphere through the vegetation roots.

The snowfall, the solid part of the incoming precipitation (SF), contributes to building the snowpack. Some part of the rain (liquid part of precipitation) and the water from the melting snowpack (SMLT) leave the land surface system as surface runoff (SRO, the surface fraction of RO). Another fraction of the precipitation leaves as intercepted water evaporation. The remaining water (from the incoming rain and the snowmelt) enters the soil and contributes to the soil water reservoir. Some of this water evaporates, either directly from the soil or through the vegetation as transpiration. In total, evaporation in HTESSEL (E, where negative E means the land-surface losing water) is the sum of evaporation of the soil and the interception and also plant transpiration. Finally, some of the water drains from the soil at the bottom at layer 4 and leave the system as subsurface runoff (SSRO, the subsurface fraction of RO).

In order to compare directly with river discharge, the trend analysis is carried out on whole catchment values for each ERA5 and ERA5-Land water budget variable introduced above, after the values on the GloFAS grid are summed together in each of the catchments. This way, essentially the catchment average value is multiplied by the area for the water related variables. The only exception is 2m temperature which was analysed as area average.

Table 1. List of variables analysed for linear trends in this study with their short names, MARS archive codes (see <https://apps.ecmwf.int/codes/grib/param-db/>), the number of processed catchments and a short description.

Short name	Unit	MARS parameter	Number of catchments	Periods	Parameter description
DIS	m^3/s	240.024	5695	1981-2018, 1981-2003 and 2004-2018	GloFAS-ERA5 river discharge
OBS	m^3/s	-	1324	Only 1981-2018 (with variable length)	River discharge observations
DIS-match-OBS	m^3/s	240.024			River discharge observation dates matched GloFAS-ERA5 river discharge

RO-match-OBS	m^3/s	205.128			River discharge observation dates matched runoff outputs of HTESSEL
P	m^3/s	228.128	5698	1981-2018, 1981-2003 and 2004-2018	Precipitation
SF	m^3/s	144.128	3774	1981-2018, 1981-2003 and 2004-2018	Snowfall part of precipitation (same in ERA5 and ERA5-Land apart from downscaling)
RO	m^3/s	205.128	5694	1981-2018, 1981-2003 and 2004-2018	ERA5 sum of surface and sub-surface runoff outputs of HTESSEL
RO-Land	m^3/s	205.128		1981-2018, 1981-2003 and 2004-2018	ERA5-Land sum of surface and sub-surface runoff outputs of HTESSEL
SMLT	m^3/s	145.128	3789	1981-2018, 1981-2003 and 2004-2018	ERA5 snowmelt output of HTESSEL
SMLT-Land	m^3/s	145.128		1981-2018, 1981-2003 and 2004-2018	ERA5-Land snowmelt output of HTESSEL
E	m^3/s	182.128	5698	1981-2018, 1981-2003 and 2004-2018	ERA5 evaporation output of HTESSEL
E-Land	m^3/s	182.128		1981-2018, 1981-2003 and 2004-2018	ERA5-Land evaporation output of HTESSEL
P-E	m^3/s	-	5686	1981-2018, 1981-2003 and 2004-2018	ERA5 precipitation minus evaporation as net water flux to the land-surface in HTESSEL
P-E-Land	m^3/s	-		1981-2018, 1981-2003 and 2004-2018	ERA5-Land precipitation minus evaporation as net water flux to the land-surface in HTESSEL
SWV7	m^3/s	39.128	5698	1981-2018, 1981-2003 and 2004-2018	ERA5 water content in the top 7cm of the soil (layer 1) in HTESSEL
SWV7-Land	m^3/s	39.128		1981-2018, 1981-2003 and 2004-2018	ERA5-Land water content in the top 7cm of the soil (layer 1) in HTESSEL
SWV100	m^3/s	39-41.128	5698	1981-2018, 1981-2003 and 2004-2018	ERA5 water content in the top 1 meter of the soil (layers 1, 2 and 3 together) in HTESSEL
SWV100-Land	m^3/s	39-41.128		1981-2018, 1981-2003 and 2004-2018	ERA5-Land water content in the top 1 meter of the soil (layers 1, 2 and 3 together) in HTESSEL
T2	C°	167.128	5698	1981-2018, 1981-2003 and 2004-2018	ERA5 temperature output at 2 metres in HTESSEL
T2-Land	C°	167.128		1981-2018, 1981-2003 and 2004-2018	ERA5-Land temperature output at 2 metres in HTESSEL

Trend analysis presentation design

This analysis documents the linear trends in the GloFAS-ERA5 river discharge and in the main ERA5 and ERA5-Land water budget variables (Table 1). The trends are shown as maps for all these variables, where the catchment outlets are represented by circles, colour coded by the trend magnitudes, with the size representing the catchment area. In addition, the annual mean time series of all variables are also shown for some major world rivers in the Appendix C, supplemented by some trend related statistics.

As a preliminary check, before the linear trends are computed on the actual river catchments, the global land average annual mean time series of the water cycle variables are analysed for ERA5 (1979-2018) and ERA5-Land (1981-2018). All simulated variables are included in Table 1, extended by snow water equivalent (SWE), surface runoff (SRO) and subsurface runoff (SSRO). This step is carried out to identify if there is any major shift or discontinuity in the global time series. All land grid points are used in the averaging, with exception of the snow water equivalent (SWE), where a mask is applied to remove the glaciers, which are represented by a fixed value of 10 metre (of water equivalent) and thus would severely distort the average values.

The linear trends in the GloFAS-ERA5 river discharge simulations are also compared to the linear trends in the observed river discharge. The river discharge trend errors are assumed to be the difference between the trends in the simulation and the observation annual mean time series. To allow fair comparison, this error computation is done on a special subset of the GloFAS-ERA5 time series that matches the dates when the observations are available (DIS-match-OBS).

As trends in runoff and river discharge are expected to be very similar, the linear trend error, i.e. the difference between the runoff and observed river discharge trends, is calculated to assess how ERA5 and ERA5-Land estimate the changes in river discharge. For runoff trend errors, the observation matching time series are used for both ERA5 (RO-match-OBS) and ERA5-Land (RO-Land-match-OBS), similar to DIS-match-OBS. The impact by ERA5-Land is expressed as the difference between the absolute value of the ERA5-Land runoff linear trend error minus the ERA5 runoff linear trend error.

Even though reanalysis systems are designed with the intention of being independent of the changes in the observing system, it is inevitable that the 38-year reanalysis period has some inhomogeneities in the use of the different meteorological observations, including a known major change in the IMS snow observation use which was analysed further. The operational snow analysis was changed in the ECMWF IFS in 2004, when the 24-km Interactive Multi-Sensor Snow and Ice Mapping System (IMS) snow cover information was introduced, in addition to the SYNOP snow depth measurements. This led to a more realistic representation of the extent of snow cover in the operational analysis (Drusch et al., 2004). Details on ECMWF's snow assimilation can be found in an ECMWF Newsletter article (de Rosnay et al., 2015). In ERA5, the higher resolution (4 km) IMS snow products could be used from 2004, as the high-resolution data was reprocessed and made available to this date.

Moreover, the snow scheme in ECMWF's HTESSEL land-surface model is documented to melt the snow too slowly (Dutra et al. 2012). This, in combination with the use of the IMS snow cover data, could lead to a negative shift in ERA5 snowmelt from 2004, as the excess snow that is not melted by the model could then be removed by the assimilation in areas where in situ observations are not available.

This change is expected to produce a clear discontinuity in the snow related time series, and also contribute indirectly to creating potential shifts in other variables. This would make it possible that the 38-year-based linear trend would reflect mostly the discontinuity. Therefore, in order to see the trends not dependent on this discontinuity, the 38-year period is split in two parts, 1981-2003 (period1, with 23 years) and 2004-2018 (period2, with 15 years). The linear trends are then computed for both periods in the same way as for the whole 38-year period for all variables other than the observation related ones (see Table 1).

Criteria for catchment inclusion

The availability of the river discharge observations varies from catchment to catchment and can be as low as 4-5 years (which is the minimum criteria to include them in the GloFAS observation database). For the trend computation a longer minimum length is needed. The criteria of 16 minimum available years, with at least 330 days available in each year, is set as a compromise considering both the observation availability and the minimum record length for reliable regression analysis (Dai et al., 2009).

For all analysed variables (Table 1), only catchments which have at least 1 m³/s whole catchment value (or river discharge) as sample mean over the 38-year period (or shorter for the observations) and also the 2-year return period flood thresholds are above 20 m³/s are considered for trend analysis. This filters out the very dry catchments for which trends would not be necessarily meaningful or representative, and similarly filters out snowfall and snowmelt for catchments in the warm climate, where the whole catchment values for these variables will be very small. However, for some of the catchments in the tropical belt the trend could still be computed, for those which have some areas over higher orography and thus some snow contribution. These snow related trends should only be interpreted with caution, as especially in the large catchments, such as the Amazon, the snow has extremely small contribution to the total river discharge.

With these criteria, about 1300 catchments could be used for observed river discharge, roughly 3800 catchments for snowfall and snowmelt and about 5700 catchments for all other water cycle variables (see Table 1 for the exact numbers).

Regarding the gaps in the available river discharge observation time series, these were simply left out of the analysis, but the catchments that had at least 16 years of data in total (even if gaps in between) were used regardless of the gaps.

Definition of trend magnitudes

The trend magnitudes are defined after applying a linear regression to the annual mean time series sample. With this the trends are assumed to be linear. This assumption will not be true for all catchments, however, for the sake of this study, this is considered to be sufficient. Linear trends are expected to show if the time series is impacted by larger changes, discontinuities, etc., which could make the flood threshold computation problematic in GloFAS. To help the relative comparison between catchments and variables, the trend T is defined by (following Stahl et al. 2012):

$$T = (10 \cdot S) / M, \quad (1)$$

as the change over a fixed 10-year period, relative to the mean (M) of the n -year period. Here T stands for the trend magnitude at a catchment, S is the slope (annual change as a result of the linear regression), and $M = \text{Mean}(\text{Var}_1, \dots, \text{Var}_n)$, with Var_1 to Var_n denoting the annual mean values from year 1 to year n . The T value gives a measure of the change in a decade. For example, a value of ± 0.1 effectively means the variable increased (decreased) by 10% of the sample mean value over the course of the 10 years. The 10-year fixed period is chosen as a common ground to allow trend comparability for catchments with different observation length.

For 2m temperature, as the only non-water-related variable, the intercomparability with other variables is less important, and also the division by the mean (M) would cause problems (being near 0 in some areas of the world), therefore the trend magnitude is defined as the temperature change in 10 years along the linear regression line ($T=10*S$, where M is replaced by 1).

An alternative trend magnitude was also calculated for all variables, including 2m temperature, by calculating the linear regression on the standardised variable (each annual mean value divided by the standard deviation of the annual mean time series) and T defined by the $10*S$. The standardised trend improves the comparability across variables with very different value ranges (e.g. river discharge and evaporation) and suffer less from the potential issue of division by near 0 values (as it can happen in some isolated cases for P-E). However, the M -based trend (Eq. 1) is more intuitive when interpreting the size of the trends and therefore was selected as the focus of our analysis. In the rest of the paper the 'raw' trend, i.e. the original definition of the linear trend with the raw variable (in Eq. 1), is analysed and displayed, with only few exceptions where the standardised trend is mentioned.

In Eq. 1, n is either 38 (1981-2018), 23 (1981-2003) or 15 (2004-2018) for all variables other than the river discharge observation related variables (OBS, DIS-match-OBS, RO-match-OBS and RO-Land-match-OBS; see in Table 1) where it varies between 16 and 38 for the catchments, depending on observation availability.

3 Results

3.1 Global land average annual mean time series for ERA5 and ERA5-Land

The discontinuity in the use of the IMS product from 2004 is a known issue in ERA5. This is clearly present in the global land average annual mean time series of the snow, both for SWE and SMLT (Figure 5). For snowmelt, the change happens from 2004, while for SWE from mainly 2005. This discontinuity, however, seems to be mostly embedded in a general decreasing trend for both variables. ERA5-Land, which does not have land data assimilation as in ERA5, does not show any sign of this change in 2004.

Another very clear shift is seen for precipitation and all the runoff variables in the 1999-2004 period, when these variables suffer a very large drop (Figure 5). There is no such tendency in snowfall, which suggests that the source of this drop must come mainly from warm climate areas where snow is not dominant.

Finally, as expected, the strong upward tendency is present for 2m temperature, with over 1 degree difference between the start and end years of the 40-year period (Figure 5).

ERA5-Land is slightly warmer and generally has more water in the water cycle. It produces more runoff, even though as a balance of much less surface runoff and lot more subsurface runoff. There is more snow in the snowpack with consequently more snowmelt. The soil also has more water but finally less evaporation than in ERA5 (Figure 5).

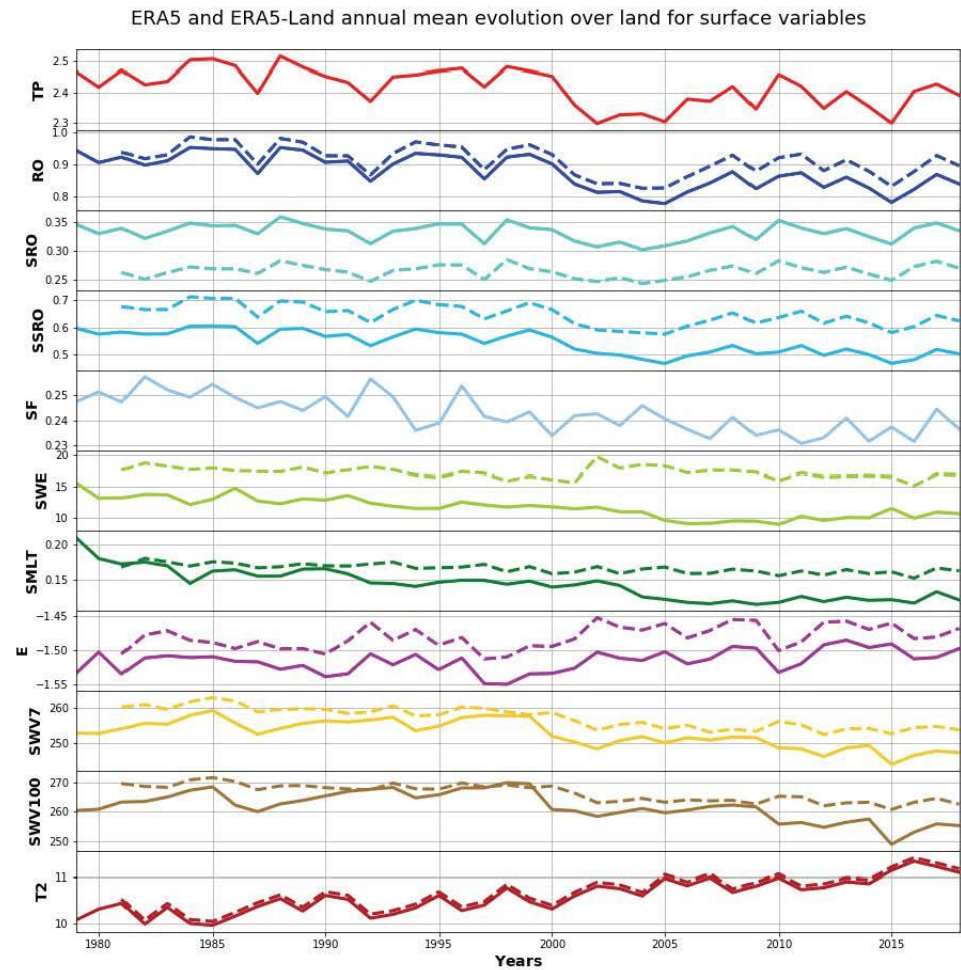


Figure 5. Annual mean time series of global land averages for precipitation (TP), runoff (RO), surface runoff (SRO), subsurface runoff (SSRO), snowfall (SF), snow water equivalent (SWE), snowmelt (SMLT), evaporation (E), soil water content in the top 7 cm (SWV7) and 100 cm (SWV100) and finally 2m temperature (T2) from ERA5 (solid lines) and ERA5-Land (dashed lines) for the 1979-2018 period (ERA5-Land is only available from 1981). TP and SF are the same in ERA5 and ERA5-Land (apart from downscaling to higher resolution in ERA5-Land), so only displayed for ERA5. All variables' unit is mm/day other than 2m temperature which has C.

3.2 Trends in GloFAS-ERA5 river discharge

Over the 1981-2018 period, GloFAS-ERA5 river discharge shows a dominantly negative raw linear trend with almost 80% of catchments exhibiting negative trend magnitudes and about 40% showing a negative value stronger than -0.1 (fraction/decade), i.e. decreasing at least by 10% of the 1981-2018 mean value across the 10-year reference period (Figure 6).

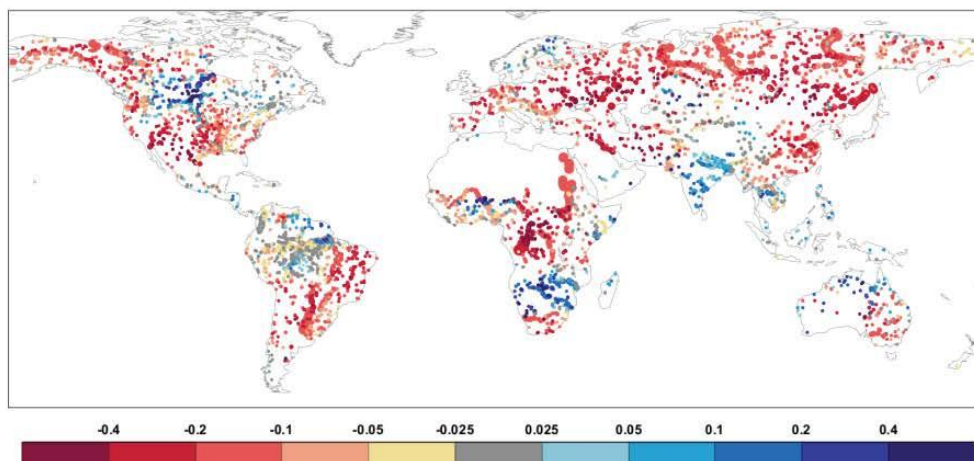


Figure 6. Raw linear trends (fraction/decade) at global river catchments for GloFAS-ERA5 river discharge, based on the 1981-2018 period. The circles represent the catchment outlets, while their size the catchment area.

The most negative river discharge trends are found in many of the larger world rivers such as: the Congo and the Nile in Africa; the Ob, Lena, Yenisei and Amur in Russia; the Dnieper and Volga in Europe; the Colorado, Mackenzie and the Yukon in North America; the Yellow and the Yangtze in China; the Tiger and Euphrates in the Middle East and the Sao Francisco, Tocantins and Paraguay in eastern South America.

Positive linear discharge trends can be seen in a few smaller areas, most notably around larger rivers like the Zambezi in South Africa, the Ganges in India and the Nelson in central Canada.

According to the DIS column in Table 2 (and also A1 with the standardised trends), only 8 out of 33 catchments show a positive trend (~25%), while 18 catchments (~55%) reveal larger trends at least with 10% decrease in the 10 years (or more than 1/3 of the standard deviation of the annual mean time series for the standardised trend in A1), making it a very substantial drop in GloFAS-ERA5 river discharge for the whole 1981-2018 period.

Table 2. Raw linear trends (fraction/decade) for selected catchments (see Figure 4) for 1981-2018, for GloFAS-ERA5 river discharge (DIS), ERA5 precipitation (P) and snowfall (SF) and both ERA5 and ERA5-Land snowmelt (SMLT), runoff (RO), evaporation (E), precipitation minus evaporation (P-E), soil moisture in the top 7 cm (SWV7) and 100 cm (SWV100) and 2m temperature (T2). Raw linear trends are also provided for observed river discharge (OBS) and the observation matched GloFAS-ERA5 river discharge (DISm) and ERA5 and ERA5-Land runoff (ROm). Differences in the absolute raw linear trend errors between ERA5-Land and ERA5 are also indicated (Imp). Empty cells correspond to cases for which trend computation was not possible. Coloured cells indicate negative (orange) and positive (blue) trends and also decreasing (green) and increasing (purple) trend errors in ERA5-Land (Imp column). Where there is no raw trend, defined for absolute values less than 0.025, cells are not coloured. Darkening shades show increasing trend magnitudes.

Rivers		DIS	P	SF	SMLT		RO		E		P-E	
		ERA5	ERA5	ERA5	ERA5	ERA5-Land	ERA5	ERA5-Land	ERA5	ERA5-Land	ERA5	ERA5-Land
North Asia	Ob	-0.15	-0.01	-0.03	-0.13	-0.04	-0.11	-0.06	-0.023	-0.018	-0.07	-0.06
	Yenisei	-0.16	-0.02	-0.03	-0.18	-0.03	-0.15	-0.07	-0.016	-0.017	-0.07	-0.08
	Lena	-0.14	0.00	-0.03	-0.24	-0.04	-0.11	-0.04	-0.019	-0.023	-0.04	-0.05
	Amur	-0.26	-0.04	-0.04	-0.39	-0.07	-0.20	-0.16	-0.007	-0.004	-0.16	-0.17
North America	Yukon	-0.14	-0.01	-0.03	-0.13	-0.03	-0.13	-0.03	-0.001	-0.014	-0.01	-0.03
	Mackenzie	-0.12	0.00	-0.01	-0.05	0.00	-0.11	-0.02	-0.008	-0.009	-0.02	-0.03
	Churchill	0.12	0.03	0.03	0.07	0.02	0.07	0.07	-0.009	-0.009	0.15	0.19
	Columbia	-0.11	-0.02	0.01	-0.06	0.01	-0.10	-0.01	0.038	0.021	0.01	-0.01
	Colorado	-0.42	-0.13	-0.16	-0.15	-0.16	-0.36	-0.43	0.143	0.120	-0.05	-1.05
Europe	Mississippi	-0.09	-0.02	-0.06	-0.08	-0.06	-0.08	-0.07	-0.005	0.006	-0.12	-0.07
	Kalixaelven	0.02	0.03	0.02	-0.06	0.02	0.01	0.05	-0.022	-0.024	0.04	0.04
	Thames	-0.11	0.01	-0.12	0.10	-0.10	-0.11	0.00	0.003	-0.013	0.05	-0.01
	Rhine	-0.12	-0.03	-0.08	-0.15	-0.07	-0.10	-0.08	-0.019	-0.023	-0.08	-0.09
	Danube	-0.09	0.01	-0.04	-0.13	-0.05	-0.08	-0.03	-0.019	-0.026	0.00	-0.02
South America	Volga	-0.27	-0.03	-0.03	-0.12	-0.03	-0.22	-0.13	-0.024	-0.013	-0.16	-0.14
	Ebro	-0.20	0.00	0.04	-0.17	0.04	-0.15	-0.02	0.059	-0.002	0.21	0.00
	Magdalena	-0.03	-0.02				-0.04	-0.03	0.001	-0.008	-0.03	-0.04
	Amazon	-0.02	0.00	-0.12	-0.12	-0.12	-0.02	-0.01	-0.001	-0.004	-0.01	-0.01
Africa	Araguaia	-0.15	-0.06				-0.13	-0.13	0.004	0.015	-0.18	-0.13
	Parana	-0.17	-0.05	-0.15	-0.16	-0.16	-0.14	-0.16	0.005	0.019	-0.21	-0.17
	Nile	-0.11	-0.07				-0.10	-0.12	0.029	0.051	-0.22	-0.15
	Niger	-0.06	-0.02				-0.05	-0.07	-0.028	0.007	-0.35	-0.06
	Shabelle	-0.04	-0.03				-0.04	-0.03	0.021	0.027	-0.08	-0.03
	White-Volta	0.14	-0.03				0.12	0.07	-0.044	0.038	-0.67	0.25
	Congo	-0.30	-0.08				-0.25	-0.28	0.005	0.008	-0.33	-0.30
	Zambesi	0.08	0.01				0.06	0.08	0.022	0.010	0.12	0.06
South Asia	Cunene	0.12	0.05				0.09	0.15	0.045	-0.048	1.25	0.09
	Orange	-0.20	-0.07	-0.02	-0.04	-0.03	-0.16	-0.17	0.090	0.048	7.83	-0.47
	Brahmaputra	-0.02	-0.02	-0.05	-0.01	-0.05	-0.02	-0.03	-0.012	-0.007	-0.03	-0.03
	Ganges	0.07	0.03	-0.05	0.02	-0.04	0.06	0.05	-0.011	-0.006	0.06	0.05
Australia	Godavari	0.10	0.05				0.10	0.11	-0.018	-0.015	0.13	0.12
	Karnali	0.03	0.01	-0.08	-0.05	-0.07	0.03	0.01	-0.012	-0.008	0.01	0.01
Murray	-0.11	-0.07	-0.13	-0.19	-0.13	-0.10	-0.14	0.035	0.056	-0.11	-0.32	



(continues on next page)

Table 2. Raw linear trends (fraction/decade) for selected catchments (see Figure 4) for 1981-2018 (continued)

Rivers		SWV 0-7		SWV 0-100		T2		OBS	DISm		ROm	
		ERA5	ERA5-Land	ERA5	ERA5-Land	ERA5	ERA5-Land	-	ERA5	ERA5	ERA5-Land	Imp
North Asia	Ob	-0.006	-0.006	-0.007	-0.009	0.22	0.22	0.02	-0.16	-0.12	-0.04	-0.081
	Yenisei	-0.010	-0.010	-0.006	-0.014	0.36	0.35	0.04	-0.15	-0.14	-0.05	-0.097
	Lena	-0.004	-0.001	-0.002	-0.001	0.37	0.36	0.04	-0.17	-0.12	-0.06	-0.065
	Amur	-0.024	-0.019	-0.012	-0.024	0.32	0.31	-0.33	-0.65	-0.52	-0.40	-0.128
North America	Yukon	-0.009	-0.002	-0.014	-0.001	0.40	0.39	0.02	-0.20	-0.18	-0.04	-0.140
	Mackenzie	-0.006	-0.003	-0.007	-0.003	0.34	0.33	0.04	-0.12	-0.10	-0.01	-0.095
	Churchill	0.003	0.006	0.007	0.008	0.13	0.12	0.20	0.22	0.11	0.07	0.045
	Columbia	-0.027	-0.015	-0.027	-0.011	0.34	0.32	-0.02	-0.13	-0.11	-0.03	-0.078
	Colorado	-0.110	-0.094	-0.104	-0.057	0.43	0.44	-0.27	-0.43	-0.37	-0.45	-0.073
Europe	Mississippi	-0.014	-0.016	-0.007	-0.019	0.24	0.25	0.03	-0.11	-0.10	-0.09	-0.006
	Kalixaelven	0.004	-0.001	0.004	0.000	0.48	0.51	0.03	-0.13	-0.10	0.03	-0.125
	Thames	-0.009	-0.001	-0.022	-0.004	0.26	0.25	0.00	-0.20	-0.18	-0.06	-0.118
	Rhine	-0.011	-0.010	-0.015	-0.012	0.40	0.39	-0.06	-0.14	-0.12	-0.07	-0.043
	Danube	-0.010	-0.006	-0.017	-0.006	0.45	0.45	0.08	-0.09	-0.07	0.00	-0.063
South America	Volga	-0.015	-0.016	-0.015	-0.019	0.35	0.35	-0.02	-0.32	-0.27	-0.14	-0.133
	Ebro	-0.032	-0.011	-0.075	-0.013	0.29	0.28	-0.07	-0.36	-0.29	-0.13	-0.167
	Magdalena	-0.012	-0.008	-0.016	-0.008	0.21	0.21	0.01	0.00	-0.01	0.00	-0.003
	Amazon	-0.010	-0.009	-0.011	-0.009	0.19	0.19	0.02	-0.04	-0.04	-0.04	-0.005
	Araguaia	-0.021	-0.022	-0.013	-0.016	0.30	0.32	-0.03	-0.11	-0.11	-0.11	0.002
Africa	Parana	-0.023	-0.027	-0.021	-0.031	0.27	0.29	-0.05	-0.19	-0.16	-0.18	0.015
	Nile	-0.029	-0.033	-0.023	-0.035	0.38	0.39	0.12	-0.13	-0.12	-0.14	0.026
	Niger	0.000	-0.008	0.015	-0.006	0.27	0.28	0.28	-0.09	-0.06	-0.10	0.032
	Shabelle	-0.017	-0.018	-0.006	-0.006	0.23	0.24	0.27	-0.12	-0.12	-0.11	-0.013
	White-Volta	-0.005	-0.019	0.044	0.000	0.17	0.21	-0.08	-0.12	-0.10	-0.14	0.033
	Congo	-0.023	-0.024	-0.028	-0.034	0.28	0.28	0.05	-0.29	-0.24	-0.27	0.032
	Zambesi	-0.013	-0.007	-0.009	0.001	0.20	0.19	0.22	0.10	0.08	0.10	-0.020
	Cunene	-0.007	0.018	-0.028	0.036	0.22	0.18	0.21	0.16	0.14	0.19	-0.057
Orange	-0.036	-0.030	-0.064	-0.025	0.27	0.26	0.13	-0.54	-0.40	-0.36	-0.034	
South Asia	Brahmaputra	0.007	0.003	0.007	0.003	0.21	0.21	-0.05	-0.02	-0.02	-0.03	-0.010
	Ganges	0.009	0.004	0.011	0.007	0.14	0.14	-0.01	0.08	0.07	0.07	-0.002
	Godavari	0.009	0.006	0.014	0.014	0.05	0.05	-0.02	0.16	0.15	0.16	0.015
	Karnali	0.000	-0.007	0.003	-0.004	0.30	0.28	-0.01	0.04	0.04	0.02	-0.018
Australia	Murray	-0.020	-0.032	-0.032	-0.029	0.30	0.31	-0.28	-0.18	-0.14	-0.16	-0.028



3.3 Trends in precipitation, snowfall and 2m temperature

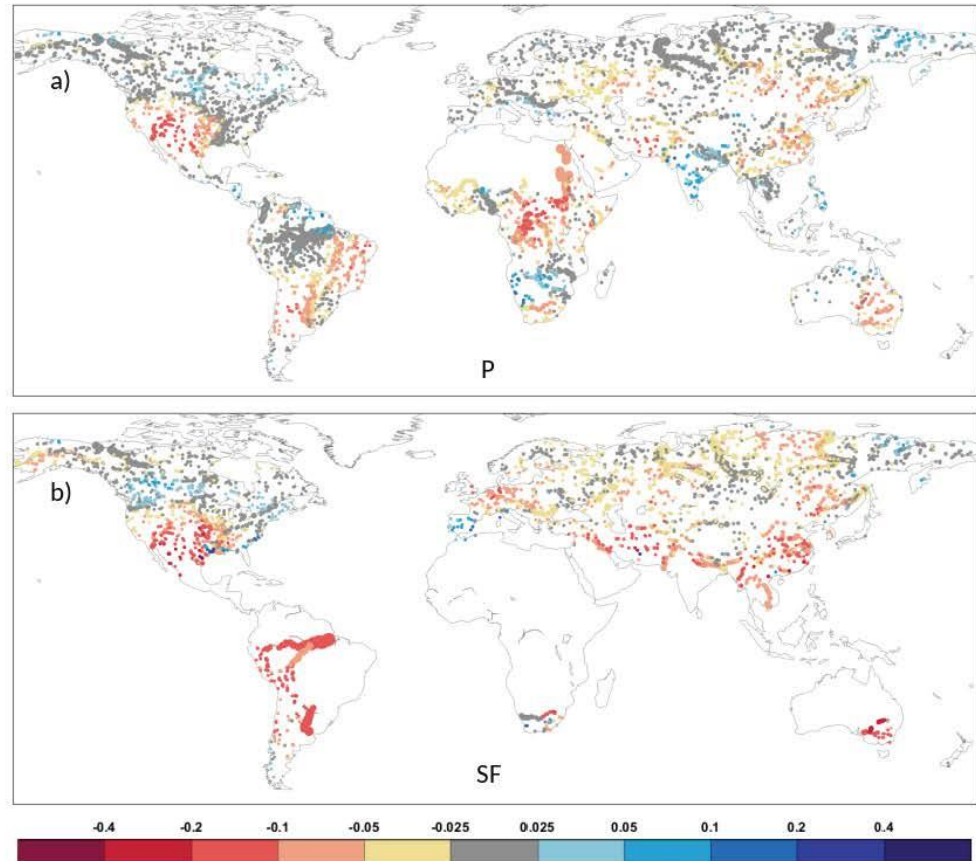


Figure 7. Raw linear trends (fraction/decade) at global river catchments for ERA5 a) Precipitation (P) and b) Snowfall (SF), based on the 1981-2018 period. The circles represent the catchment outlets, while their size the catchment area.

Precipitation shows the same trend signs to river discharge over much of the world (Figure 7a). The raw linear trend magnitudes are generally smaller than in river discharge, but a lot of the differences are in fact related to the larger M-term in Eq.1, as precipitation is generally much larger than river discharge. The standardised linear trends highlight (see Table A1 for the selected global catchments in Appendix B) that in tropical and subtropical areas of the world, including Africa, Central and South America, South Asia and Australia, precipitation and river discharge trends are very similar, while over the Northern Extratropics river discharge trends tend to be more pronounced than precipitation.

The largest linear trends (both raw and standardised) in precipitation are for the Colorado and Rio Grande basins in the United States and also the Congo and upper Nile basins in Central Africa, where the negative raw trend is between -0.1 and -0.2 (10-20% decrease in just 10 years, see also the P column in Table 2; or over 0.5 standard deviation of the annual mean time series sample, see Table A1 with the standardised trends).

Although some of the tropical and subtropical basins, mainly the rivers downstream of the Andes in South America and areas in South Asia, show clearly more negative trends, the snowfall contribution to precipitation is usually very little in those areas (Figure 7b). On the other hand, in the more snow dominant higher latitude areas of the Northern Hemisphere, some areas such as Alaska, central parts of Europe or northern parts of Russia, there is a moderate tendency to more negative trends.

The main reason for the decreasing snowfall is very likely to be the generally increasing temperature. Figure A1 in the Appendix A highlights that the 2m temperature trend is positive almost everywhere. The largest positive trends, above 0.4 degree change in 10 years, are in the Nile basin in Africa, the southwest part of the USA, the eastern parts of Europe, the Middle-East and also many of the smaller rivers in the northern-most latitudes. Trends in ERA5 and ERA5-Land are generally similar, if anything, ERA5-Land tends to be slightly warmer (see the annual mean time series for the selected global river basins in Appendix C), but the trends are almost the same everywhere. These differences can originate from the snow cover and evaporation processes which can differ in coupled (ERA5) vs offline (ERA5-Land) systems.

These large global temperature increases in ERA5 are consistent with the scientific literature (e.g. Hansen et al., 2006; Parmesan and Yohe, 2003, IPCC, 2007 and 2014).

It is difficult to determine how realistic the identified ERA5 precipitation and snowfall trends are, due to the sparse precipitation observing network and the highly variable quality of the satellite derived precipitation data (Sun et al., 2018).

The AR4 (IPCC, 2007) and AR5 (IPCC, 2014) IPCC reports disagree on precipitation trends in many parts of the world, the worst being West Africa where significant positive (AR4) trends turn to significant negative (AR5). Part of the reason is the different periods (1979-2005 in AR4 - Figure 3.13 vs. 1951-2010 in AR5 - Figure 1.1), however, the shortness of the first period (through capturing more of the natural climate variability as trends) and the differences between the used data sets must have also contributed. The estimation uncertainties in precipitation changes were acknowledged in the AR5 report when it concluded that 'Confidence in precipitation change averaged over global land areas since 1901 is low prior to 1951 and medium afterwards'. In fact, neither of the two IPCC reports are similar to the ERA5 precipitation linear trends presented here. In particular, there is no clear sign of either of the two most prominent trend areas in ERA5. While the southwestern USA is represented in AR4, although with smaller negative values in AR4 than in ERA5 (3-10% vs. 10-20%), the large negative ERA5 trends in central Africa are not there in any of these two sources.

Nguyen et al. (2018) analysed precipitation trends with satellite derived data in the 1983-2015 period, which is directly comparable with our period. They concluded that although only few percent of the land mass show pixel-by-pixel significant trends, this increases by regional- or catchment-based analysis, but even on the large catchment-scale (over 200 large rivers), only a smaller fraction has significantly large trends. The catchment-scale precipitation trends in Nguyen et al. (2018) show a lot of similarities, at least in sign, to the ERA5 precipitation linear trends in Figure 7a. For example, the trend patterns are similar in most parts of Australia, Central and South Africa, South America, but also in Europe and large parts of Asia. Major differences are present mainly in Central America, around Canada and also Central Asia where the trends show mainly opposite signs between ERA5 (Figure 7a) and the satellite-derived precipitation in Nguyen et al. (2018).

The dominantly decreasing trend in ERA5 is not supported by either of GPCC or GPCP, two of the available global precipitation estimate data sets, as described in Hersbach et al. (2020; see Figure 23). ERA5 seems to produce more precipitation than either GPCC or GPCP, and the difference gets smaller from 2000. It seems, after the large decline around 2000-2002 (see also Figure 5), the ERA5 precipitation is more realistic in the 21st century, which could potentially come from some changes in the used satellite data in ERA5. Nevertheless, further analysis is going to be needed in the future to better understand the behaviour of the ERA5 precipitation changes.

3.4 Trends in runoff and snowmelt

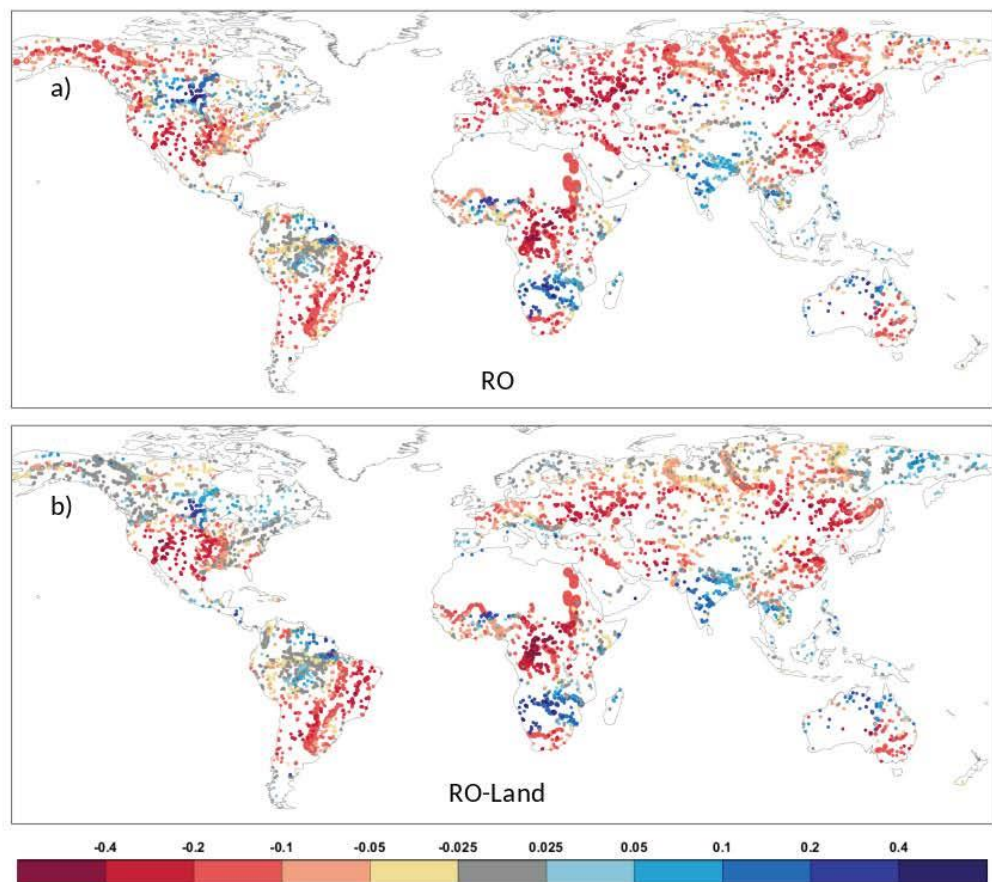


Figure 8. Raw linear trends (fraction/decade) at global river catchments for runoff in a) ERA5 (RO) and b) ERA5-Land (RO-Land), based on the 1981-2018 period. The circles represent the catchment outlets, while their size the catchment area.

The linear trends for runoff in ERA5 (Figure 8a) are almost identical to the GloFAS-ERA5 river discharge trends (Figure 6). This is expected, as the annual mean values of these two variables can usually differ only little, the river routing's time delay is averaged out by computing the mean over the whole year. However, the trends for runoff in ERA5-Land (figure 8b) are different from ERA5 in the

Northern Extratropics, namely the ERA5 trends are more negative in Alaska, western Canada and most of northern Eurasia. As the atmospheric forcing are the same in ERA5 and ERA5-Land, and the land-surface model is also mainly the same, differences in runoff and other surface variables will come from the missing coupling and land data assimilation and the much higher resolution and lapse-rate correction in ERA5-Land. The land data assimilation impact on the hydrological cycle can be substantial, considering both snow and soil moisture, as shown in Zsoter et al (2019), and the resolution change, through the different temperature conditions with the lapse rate correction, is also expected to have a potentially large impact.

As the ERA5 and ERA5-Land runoff trends are very similar in the tropics and subtropics (see Figure 8 and also the bottom half of Table 2 and A1), the likely culprit for the differences in the higher latitudes is the handling of the snow with the possible differences in snowmelt.

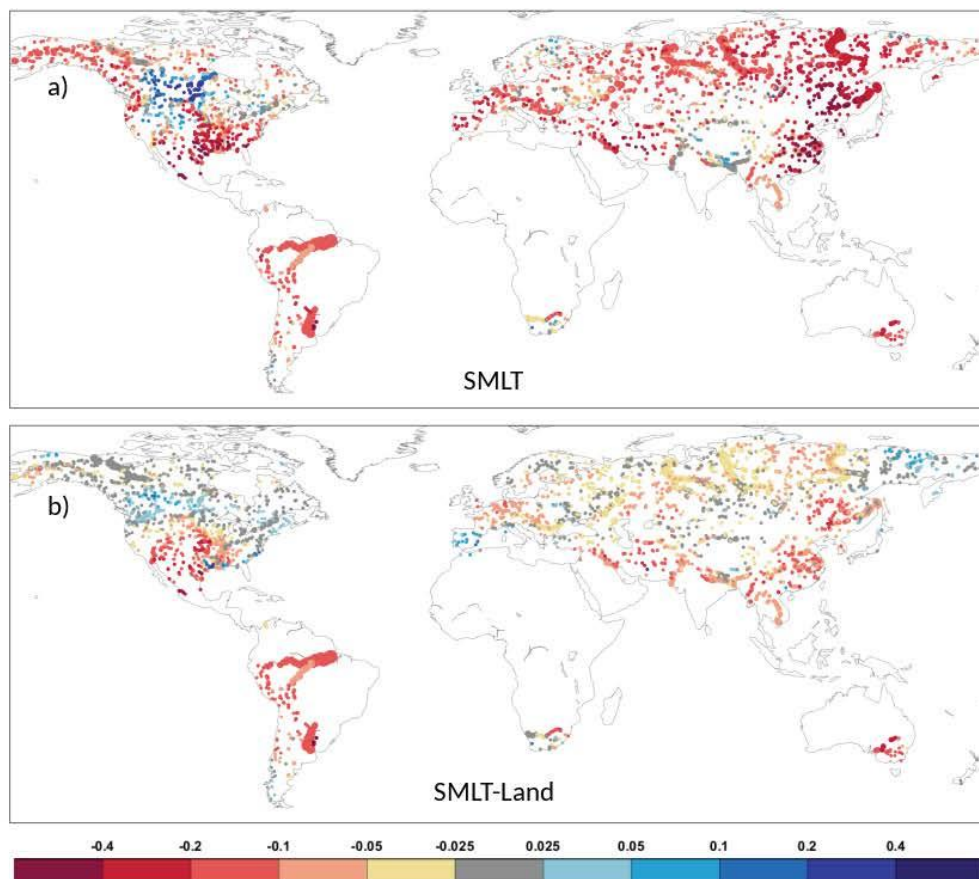


Figure 9. Raw linear trends (fraction/decade) at global river catchments for snowmelt in a) ERA5 (SMLT) and b) ERA5-Land (SMLT-Land), based on the 1981-2018 period. The circles represent the catchment outlets, while their size the catchment area

The runoff raw linear trend differences seem to come from snowmelt which are very different in ERA5 (Figure 9a) and ERA5-Land (Figure 9b). The snowmelt trends in ERA5 are much more negatively oriented in the Northern Extratropics such as the Lena, Amur (in Russia) or the Yukon (in Alaska) rivers. The exceptions are the central part of North America and some catchments in Northern Europe and near the Himalaya, where ERA5 has more positive trends (see the Churchill and Thames rivers in Table 2 and A1; SMLT columns), where snowmelt appears to have increased in the last ~40 years in ERA5. These ERA5 snowmelt linear trends can be partially explained by the changes in snowfall, i.e. the amount of snow available to melt, as the sign of the snowfall and snowmelt trends is the same over large parts of the world. However, the magnitude is clearly different in the Northern Hemisphere, with the snowmelt trends being lot more pronounced.

The scientific literature has documented trends in the snowpack related variables (snowfall, snow depth, snow cover) extensively. Studies are either based on available in situ observations, or satellite derived measurements. However, only snow cover extent (e.g. the IMS snow cover product used at ECMWF) seems to be reliable enough to be used quantitatively from satellites and snow water equivalent (SWE) and snowdepth (SD) have higher uncertainty and thus offer limited help (Hancock et al. 2013).

The 5th IPCC report states that there is very high confidence that the extent of Northern Hemisphere snow cover has decreased since the mid-20th century (IPCC, 2013, see Figure SPM.3). The declining snow cover is strongly related to the increasing global temperatures.

For example, Kunkel et al. (2016) provided a summary of several snow climatology studies and also found a strong decrease in maximum seasonal snow depth in the studied North America and Europe. Connolly et al. (2019) evaluated the snow cover extent trends in the Northern Hemisphere based on the Rutgers University snow cover dataset and found that the trend in the satellite-derived observations are poorly explained by the CMIP5 climate models, the models exhibiting a steadier decline during the 1967-2018 period. In our study no snow cover extent is analysed, however, both the snow water equivalent and snowmelt similarly show a steady decline over the global land areas in Figure 5, at least until 2004. Similarly, Knowles (2015) evaluated the trends in the GHCND climate observations for snow related variables in the 1950-2010 period for the United States. It was found that both snowfall and snow depth showed more negative than positive trends, which is in agreement with our analysis, even though the geographical distribution of their trends is different to the linear trends documented for snowfall and also for snowmelt in this study (Figure 7b and 9).

It can be concluded that the ERA5-Land runoff linear trends (the signs at least) are generally similar to the precipitation trends everywhere in the world. In ERA5, however, the runoff (and thus river discharge) trends seem to be dominated by precipitation in the tropics and subtropics, while the trends in higher latitudes resemble the snowmelt trends, as a consequence of the additional land data assimilation (see the SMLT and RO columns in Table 2 and A1). Note that the attribution of the ERA5/ERA5-Land trend differences to the observed and simulated trends in existing studies, and also the interpretation of these global and regional differences, were beyond the scope of this study, but would be a very important further piece of analysis.

3.5 Trends in evaporation, precipitation minus evaporation and soil moisture

The evaporation (E) and the two soil moisture variables (top 0-7 cm, SWV7 and top 0-100 cm, SWV100 layers) all show smaller raw linear trends for both ERA5 and ERA5-Land (Figure A2 and A4-5). The

smaller trend magnitudes mainly come from the much larger volume in these variables, thus a larger M term in Eq. 1 and thus a smaller trend magnitude (T) in relative terms. This is supported by Table A1 in Appendix B which shows that the standardised trends, defined by using the standardised variables, are generally in the same magnitude range as the other variables. More noticeable trends (both raw and standardised) are present e.g. in the Nile basin and Middle East regions or the southwestern United States area for evaporation and the soil water content variables.

The scientific literature agrees that during the last few decades the global land evaporation has generally increased (e.g. Jung et al. 2010; Zhang et al. 2016; Anabalon and Sharma et al. 2017). However, there seems to be evidence that this stopped in the 1998-2008 period (Jung et al. 2010), and then evaporation is relatively stable since (Javadian et al. 2020).

In contrast, the ERA5 and also ERA5-Land global land average evaporation trend (in Figure 5) is more positive than negative, somewhat contrary to the literature (showing slightly decreasing absolute values). However, the first period shows a small increase until 1998, followed by a marked decrease and then mainly no change in the last 10 years which is broadly similar to what is reported in the literature.

ERA5/ERA5-Land agrees with the large negative evaporation trends shown by Zhang et al. (2016) in the Middle East, western United States and the generally positive trends in the northern latitudes and also in Southeast Asia. However other areas show marked differences, especially parts of Africa and much of Australia.

Soil moisture has been shown to be generally decreasing in the last few decades by several studies (e.g. Feng and Zhang et al. 2016; Albergel et al 2013; Pan et al. 2019; Dorigo et al. 2012), agreeing with the dominantly negative trends documented in this study. About 30% of land is shown to have significant trend in these studies, a majority of being negative. However, there are large differences in the actual pattern depending on the data set used. For example, as shown in Albergel (2013), the ERAI-Land, MERRA-Land and SM-MW (a microwave-based multisatellite surface soil moisture dataset) all show marked differences in the trend patterns. Similarly, the ERA5 and ERA5-Land soil moisture trends show notable similarities to the SM-MW trends only in Central Asia, the Middle East, central South America, otherwise the match is poor.

The precipitation minus evaporation (P-E), i.e. the water source to the land-surface, and the two analysed soil moisture variables show roughly similar trend signs (see Figure A3-5). Moreover, ERA5 and ERA5-Land are broadly similar for all these four variables and only exhibit a few regional variations, most notably in Africa where they change noticeably in ERA5-Land. This happens over areas like the Niger, White-Volta and Cunene rivers, where evaporation even changes trend sign from negative to positive (Niger, White-Volta), which actually means decreasing amount of water leaving the land-surface through evaporation, as evaporation is dominantly negative over the world, or the Cunene river where evaporation changes from positive to negative, or the Nile river where the positive trend gets more pronounced, all these changes coinciding with also large swings in soil water content (see Table 2 and A1; the E, TP-E, SVW7 and SWV100 columns).

3.6 Trends in observed river discharge

The obvious question about the large trends in GloFAS-ERA5 river discharge is whether they are also present in the observations. Figure 10 highlights the match between the simulated and observed raw linear river discharge trends.

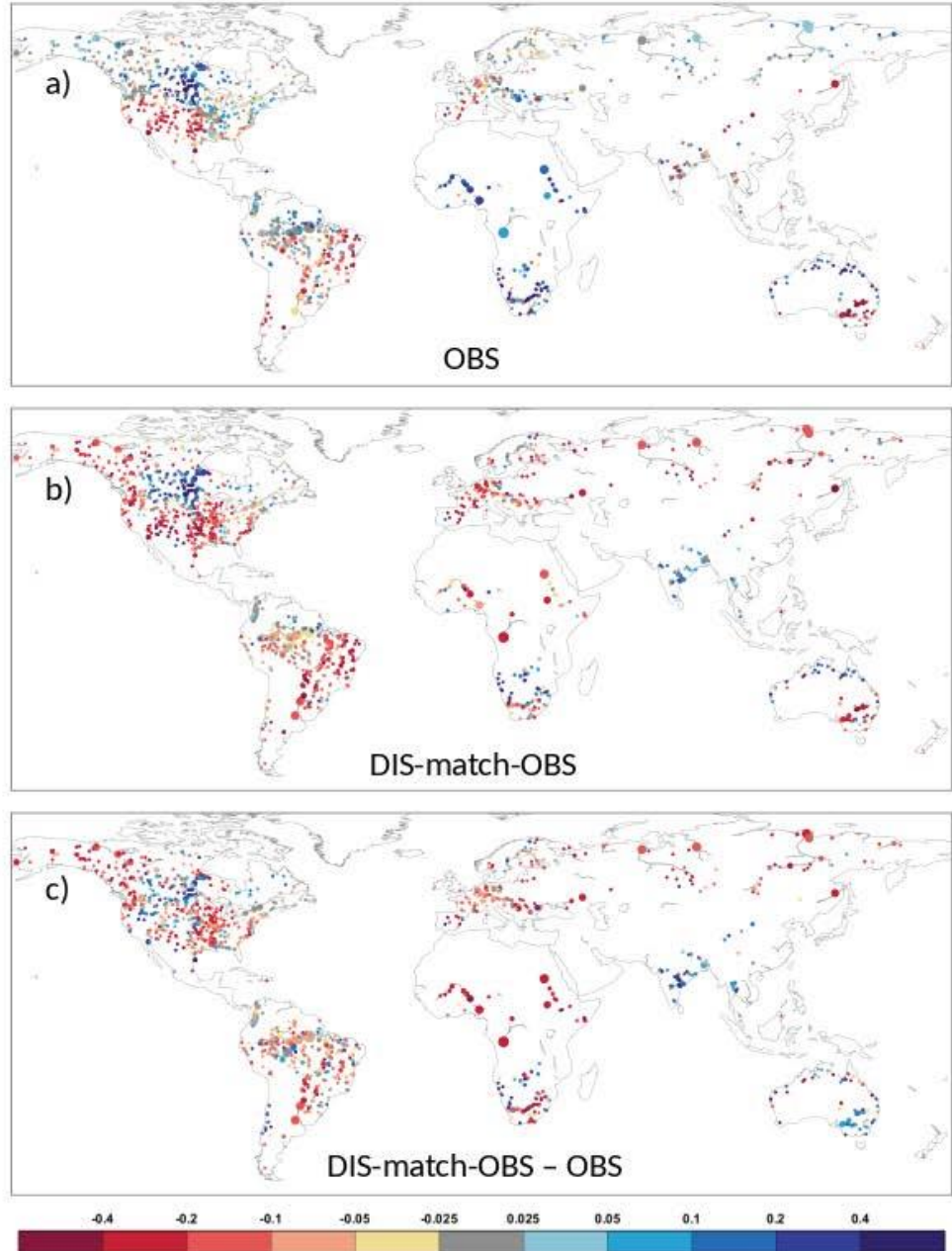


Figure 10. Raw linear trends (fraction/decade) at global river catchments for a) river discharge observations (OBS), b) GloFAS-ERA5 river discharge matching the available observations (DIS-match-OBS) and c) the difference (DIS-match-OBS minus OBS) catchments that have at least 16 years of available observations based on the 1981-2018 period. The circles represent the catchment outlets, while their size the catchment area.

The observed trends (Figure 10a) show a mixed picture with few more positive than negative changes. Almost all catchments show positive trends in Africa (Congo, Nile, Niger, Orange rivers), but also many in Russia, Canada, northern Australia and some in Amazonia (see for trend details in Table 2; OBS).

Other scientific studies, for example Su et al. (2018) or Dai et al. (2009) found similar results after analysing hundreds of the world's largest ocean-reaching rivers with mixed positive and negative trends. Although the main emphasis of their trends is shifted to generally more rivers showing negative than positive trends, this could likely be related to the different period (1948-2004) or the different geographical distribution of the analysed catchments.

In contrast, the GloFAS-ERA5 linear trends are dominantly negative, even though these trends are calculated over the exact same periods as for the observations. The observation-matched-period-based GloFAS-ERA5 raw linear trend (Figure 10b) can be quite different to the 38-year-based version in Figure 6, but the overall pattern is the same in both. The GloFAS-ERA5 trends are dominantly negative, while the observation trends are more positive than negative, therefore the difference between them is overwhelmingly negative (Figure 10c). This can be demonstrated by the rather different colours in Table 2 and A1 (columns of OBS and DIS-m as abbreviated from DIS-match-OBS), the simulation trend being dominantly orange (negative), while the observation trend is blue (positive). Clusters of positive differences (i.e. observations have a stronger tendency to increase) can mainly be seen in South-Asia, southern Australia and parts of central North America.

The trends in the GloFAS-ERA5 river discharge are thus only a poor match for the trends of the available observations. Apart from the likely reason of the unrealistic trends in the ERA5 forcing, some of this can be explained by the inadequate handling of the human influence in GloFAS-ERA5, which in some areas can have very large impact on river discharge, even though this is not necessarily will impact the sign of the trends. For example, see the Nile river in Appendix C, which has observed river discharge that is only a fraction of the GloFAS-ERA5 value. A large part of this comes from the fact that the river is highly regulated with also irrigation being important in the area.

3.7 Trend error comparison ERA5 vs ERA5-Land

It was shown earlier that the ERA5 and ERA5-Land snowmelt trends are markedly different in the Northern Hemisphere higher latitudes, which then directly influences the runoff trends. The error of ERA5 and ERA5-Land runoff raw linear trends, computed against the observed river discharge raw linear trends, are compared in Figure 11 (the raw trend values are shown in Table 2, while the standardised trends in A1 for selected catchments, with green and purple colours). It shows the difference of the absolute trend errors, with blue (green in Table 2 and A1) catchments showing where the ERA5 runoff trends are closer to the observed trends, while red (purple in Table 2 and A1) showing where the ERA5-Land trends are closer. It is clear that, due to the large difference in snowmelt, the ERA5-Land runoff linear trend is clearly closer to the observed trends in the higher latitudes (see the

dominantly green cells in Table 2 and A1, R0m/Imp column, over North Asia, North America and Europe). However, in the tropical and subtropical areas, and also in the central part of the United States, ERA5 is closer to the observations or the two have similar trends (see in Table 2 and A1 the slightly more purple colours over South America, Africa, South Asia and Australia).

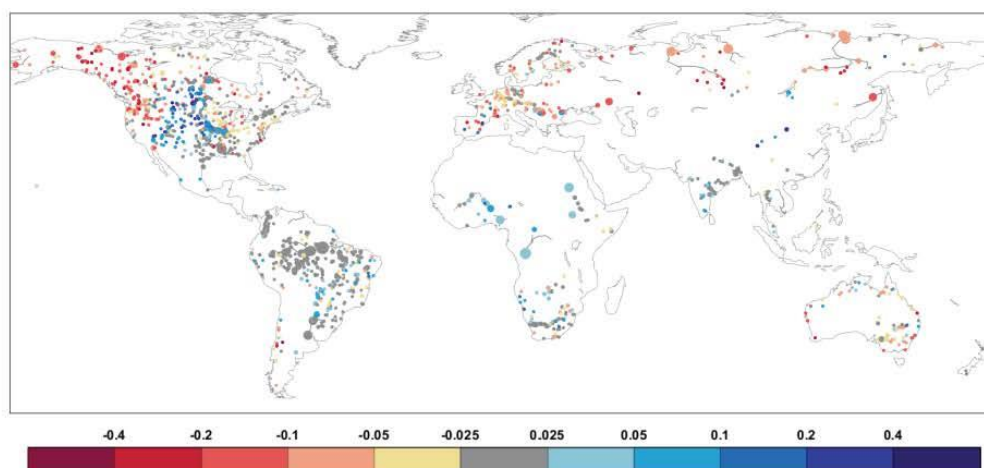


Figure 11. Difference between ERA5-Land and ERA5 absolute raw linear trend errors (simulated trend minus observed trend; fraction/decade) at catchments that have at least 16 years of available river discharge observations based on the 1981-2018 period. The circles represent the catchment outlets, while their size the catchment area.

3.8 ERA5 and ERA5-Land trends in 1981-2003 and 2004-2018

In this section the trend is compared in the two subperiods, 1981-2003 (period1, 23 years) and 2004-2018 (period2, 15 years), after removing the impact of the IMS satellite snow cover caused discontinuity in the time series.

The linear trends are generally different in the two periods, and usually larger for period2, although this shorter period would be expected to show larger random variability anyway. Figure 12 shows the raw linear trends for a few variables, while Table 3 highlights them for several variables for the selected global rivers, similarly to Table 2 and A1 (for location see Figure 4).

For precipitation, the rivers in southwestern United States, eastern South America, central Africa, the Middle East and eastern Australia stand out with their large negative trends in period1 (Figure 12a). In fact, these trends are very similar to the original 38-year trends in Figure 8a, for the majority of the world, especially for the stand-out negative areas. The second part of the 38-year period, on the other hand, does seem to show less stand-out geographical areas, the picture more mixed, even though the trend values are quite large, likely related to the shortness of the period. For snowmelt (Figure 12c-d) the same is valid.

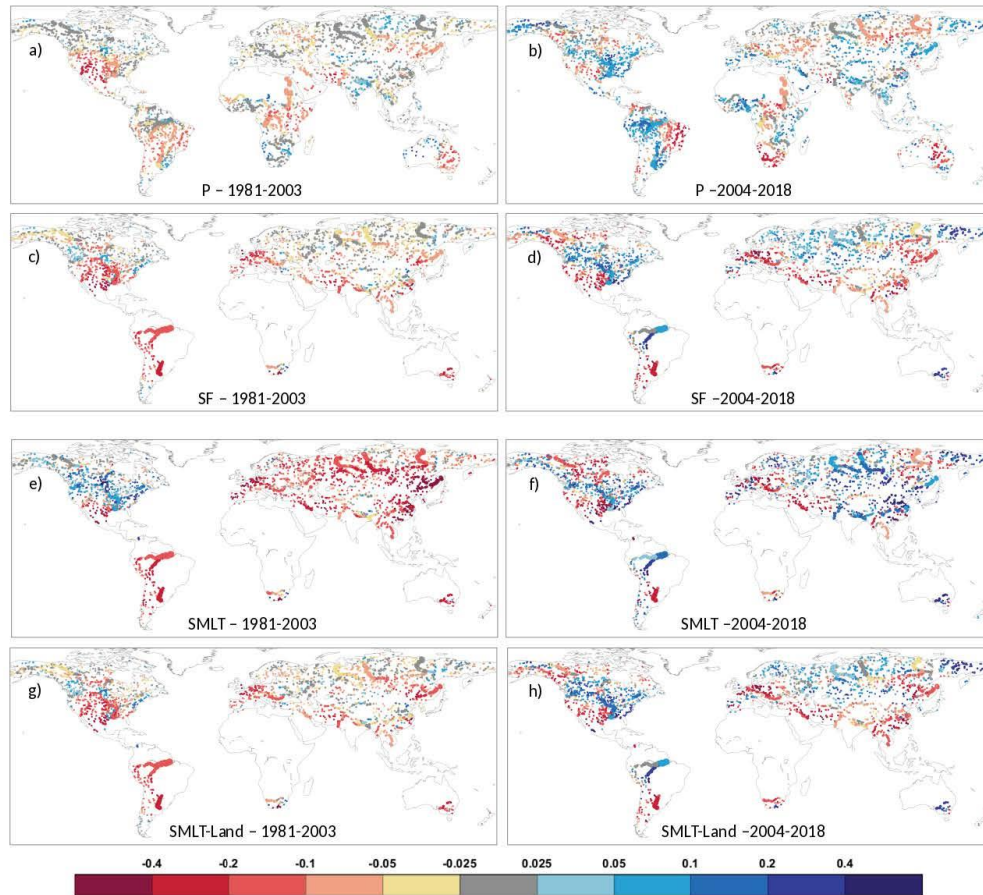
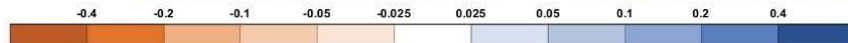


Figure 12. Raw linear trends (fraction/decade) at global river catchments for 1981-2003 (left column) and 2004-2018 (right column) for ERA5 precipitation (a-b), snowfall (c-d) and for ERA5 (e-f) and ERA5-Land (g-h) snowmelt. The circles represent the catchment outlets, while their size the catchment area.

Regarding snowmelt, the behaviour of ERA5 (Figure 12e-f) and ERA5-Land (Figure 12g-h) clearly differ, even after splitting the 38-year period in two. While period 2 behaves similarly for both (compare Figure 12f and 12h), the magnitude of the linear trends is in the same range, period1 shows larger differences between ERA5 and ERA5-Land (compare Figure 12e and 12g), Eurasia is more negative, while North America is somewhat more positive. This is also visible amongst the selected catchments in Table 3 (SMLT and SMLT-Land columns), where the 1981-2003 column for ERA5 is more orange than for ERA5-Land over Europe and North Asia, and generally more blue areas for North America. The pronounced differences between the ERA5 and ERA5-Land snowmelt for period1 suggests that the snow assimilation likely plays a role in producing the negative trends even before the introduction of the IMS snow product in 2004.

Table 3. Raw linear trends (fraction/decade) for selected catchments (see Figure 4) for two periods, 1981-2003 and 2004-2018, for GloFAS-ERA5 river discharge (DIS), runoff for ERA5 (RO) and ERA5-Land (RO-Land), precipitation (P), snowfall (SF) and snowmelt for ERA5 (SMLT) and ERA5-Land (SMLT-Land). Empty cells correspond to cases for which trend computation was not possible. Coloured cells indicate negative (orange) and positive (blue) trends. Where there is no raw linear trend, defined for absolute values less than 0.025, cells are not coloured. Darkening shades show increasing trend magnitudes. For location of the catchments see Figure 4.

Rivers		DIS		RO		RO-Land		P		SF		SMLT		SMLT-Land	
		1981-2003	2004-2018	1981-2003	2004-2018	1981-2003	2004-2018	1981-2003	2004-2018	1981-2003	2004-2018	1981-2003	2004-2018	1981-2003	2004-2018
North Asia	Ob	-0.13	0.01	-0.10	0.00	0.00	0.01	0.01	0.01	-0.03	0.04	-0.20	0.06	-0.03	0.03
	Yenisei	-0.13	-0.16	-0.12	-0.12	-0.06	-0.14	-0.02	-0.07	-0.04	0.00	-0.18	0.10	-0.05	0.02
	Lena	-0.18	-0.19	-0.13	-0.17	-0.07	-0.17	-0.01	-0.07	-0.01	-0.04	-0.14	-0.06	-0.02	-0.04
North America	Amur	-0.28	0.21	-0.23	0.15	-0.20	0.09	-0.06	0.06	-0.08	-0.16	-0.47	0.09	-0.14	-0.15
	Yukon	-0.03	0.07	-0.03	0.10	-0.04	0.12	-0.02	0.08	-0.03	-0.06	0.02	0.01	-0.03	-0.07
	Mackenzie	0.00	-0.27	0.00	-0.19	0.00	-0.08	0.00	-0.01	-0.04	-0.08	0.01	-0.15	-0.04	-0.05
	Churchill	0.65	-1.14	0.29	-0.53	0.04	-0.10	0.01	-0.02	-0.14	0.03	0.03	-0.17	-0.17	0.06
Europe	Columbia	-0.02	0.14	-0.02	0.12	-0.07	0.17	-0.07	0.04	0.01	0.06	0.10	0.09	0.01	0.09
	Colorado	-0.62	-0.64	-0.56	-0.52	-0.68	-0.47	-0.27	-0.10	-0.22	-0.29	-0.20	-0.28	-0.22	-0.28
	Mississippi	0.01	0.13	-0.01	0.15	-0.12	0.22	-0.06	0.09	-0.12	0.07	0.08	0.03	-0.12	0.09
	Kalixaelven	-0.27	0.61	-0.20	0.45	-0.03	0.12	-0.01	0.03	0.01	0.09	-0.13	0.55	0.01	0.19
South America	Thames	-0.07	0.39	-0.07	0.36	0.03	0.37	0.02	0.02	-0.67	0.21	-0.94	-0.09	-0.65	0.17
	Rhine	-0.11	0.16	-0.09	0.13	-0.07	0.04	-0.02	-0.03	-0.19	-0.18	-0.37	0.02	-0.18	-0.13
	Danube	-0.07	-0.04	-0.07	-0.08	-0.05	-0.11	0.01	-0.03	-0.09	-0.12	-0.15	-0.12	-0.10	-0.13
	Volga	-0.32	-0.11	-0.27	-0.10	-0.11	-0.11	-0.03	-0.06	0.01	0.05	-0.14	0.04	0.00	0.03
	Ebro	-0.21	0.52	-0.16	0.42	-0.14	0.33	-0.01	0.17	-0.06	0.12	-0.19	0.60	-0.08	0.14
Africa	Magdalena	-0.03	-0.17	-0.04	-0.17	-0.05	-0.17	-0.03	-0.11						
	Amazon	-0.03	0.10	-0.03	0.10	-0.03	0.11	-0.02	0.06	-0.15	0.07	-0.17	0.10	-0.16	0.07
	Araguaia	-0.16	-0.33	-0.15	-0.29	-0.17	-0.28	-0.08	-0.09						
South Asia	Parana	-0.12	0.11	-0.11	0.07	-0.12	0.09	-0.05	0.05	-0.21	-0.26	-0.23	-0.26	-0.21	-0.26
	Nile	-0.12	-0.01	-0.12	-0.04	-0.14	-0.05	-0.07	-0.07						
	Niger	-0.01	0.09	0.00	0.08	-0.02	0.08	0.00	0.08						
	Shabelle	-0.13	-0.04	-0.13	-0.04	-0.13	-0.05	-0.09	0.03						
	White-Volta	0.26	0.54	0.22	0.46	0.17	0.51	0.00	0.17						
	Congo	-0.24	-0.16	-0.20	-0.16	-0.23	-0.17	-0.08	-0.05						
	Zambesi	-0.05	0.30	-0.03	0.20	0.01	0.21	-0.02	0.01						
Australia	Cunene	0.22	-0.31	0.20	-0.29	0.26	-0.16	0.10	-0.05						
	Orange	0.17	-0.53	0.08	-0.42	0.08	-0.43	0.00	-0.22	-0.07	-0.15	-0.09	-0.09	-0.07	-0.14
	Brahmaputra	0.01	0.03	0.01	0.03	0.01	0.01	0.00	0.01	-0.02	-0.06	-0.04	0.06	-0.02	-0.09
	Ganges	0.12	0.12	0.10	0.11	0.12	0.10	0.05	0.04	-0.03	-0.04	-0.04	0.18	-0.01	-0.03
South Asia	Godavari	0.07	0.10	0.07	0.09	0.10	0.11	0.06	0.01						
	Karnali	0.07	0.11	0.07	0.11	0.06	0.10	0.04	0.05	-0.04	-0.09	-0.04	0.03	-0.03	-0.06
	Murray	-0.22	0.38	-0.17	0.19	-0.23	0.12	-0.10	-0.05	-0.19	0.21	-0.27	0.23	-0.19	0.22



4 Conclusions

This study has analysed the GloFAS-ERA5 river discharge reanalysis data set, the related ERA5 and ERA5-Land surface variables, and the available river discharge observations, for noticeable changes in the time series characterised by linear trends, in the 38-year period of 1981-2018, also including the 1981-2003 and 2004-2018 subperiods. It was found that the river discharge simulation shows a dominantly negative change across the world during 1981-2018, with some major world rivers having quite substantial decrease (Yenisei, Volga, Congo, Amur, Colorado, Yukon, Nile, Lena, Yellow, see Table 2 and A1).

The river discharge observations do not support such dominantly negative linear trends, and although varied, observations show overall more positive than negative changes in the 38-year period. The scientific literature generally documents a similar behaviour to the observational trend analysis

presented here, with mixed trends globally, but with slightly more major rivers showing negative than positive changes in river discharge during the period of 1948-2004 (Su et al., 2018; Dai et al., 2009).

The linear trends in GloFAS-ERA5 seem to be driven by changes in precipitation over the tropical and subtropical areas of the world, with the snowmelt changes showing a very strong influence in determining the river discharge trends in the northern latitudes of the Northern Hemisphere. The reason for this atypical behaviour in the northern latitudes is likely to be related to changes in the snowmelt producing processes, including the snow assimilation.

The snowmelt exhibits a pronounced negative linear trend in large parts of the world in ERA5, while this is not present in ERA5-Land. This suggests that the negative trends are, at least partially, related to the snow assimilation tendency to remove water from the water cycle as a consequence of the suboptimal snow scheme in HTESSEL (Zsoter et al. 2019). Some of the issues stem from the use of the IMS snow product from 2004 in the snow assimilation, which creates a discontinuity in the ERA5 time series (see Figure 5).

It has to be acknowledged that such discontinuity can make any trend analysis unreliable. However, as Figure 5 suggests, the snow evolution in ERA5 and ERA5-Land seems to be more complex than a single discontinuity in 2004, so linear trends can still deliver valuable information even on the whole 1981-2018 period.

After splitting the period into 1981-2003 and 2004-2018 and removing the impact of this discontinuity on the linear trends, it could still be shown, that even in the first subperiod of 1981-2003 there is a large area globally with significant negative snowmelt trends in ERA5, mainly in Asia and Europe, which is not present in ERA5-Land. This highlights that there should be other contributing factors in generating such negative trends in the ERA5 snowmelt, other than just the introduction of the IMS satellite product. Potentially the generally lower temperatures in mountains due to the higher orography in ERA5-Land could also contribute to this by decreasing the snowmelt amount compared with ERA5.

Two particularly interesting areas with the largest linear trends, that were highlighted by this study, are the central region of Africa (e.g. Congo and Nile river basins), and the southwestern part of the United States (e.g. Colorado river basin). Both these areas show very dominant and large negative trends both in precipitation and river discharge, but also in runoff by not just ERA5 but also equally by ERA5-Land. However, based on the limited analysis of the scientific literature, there was no indication of such strong precipitation trends in the explored studies, and similarly no such large trends were shown either in the river discharge observations available in this study.

It will require more work in the future to better identify the underlying reasons for these very dominant negative trends. Moreover, it would also be beneficial to repeat the analysis including an improved precipitation observation data set, preferably one that is high quality and merges several of the available gauge- and/or satellite-based data sets, such as the MSWEP (Beck et al., 2019), or the latest bias corrected ERA5 data set, WFDE5 (Cucchi et al. 2020). This would allow us to directly evaluate the quality of the ERA5 trends against the best available observation estimates.

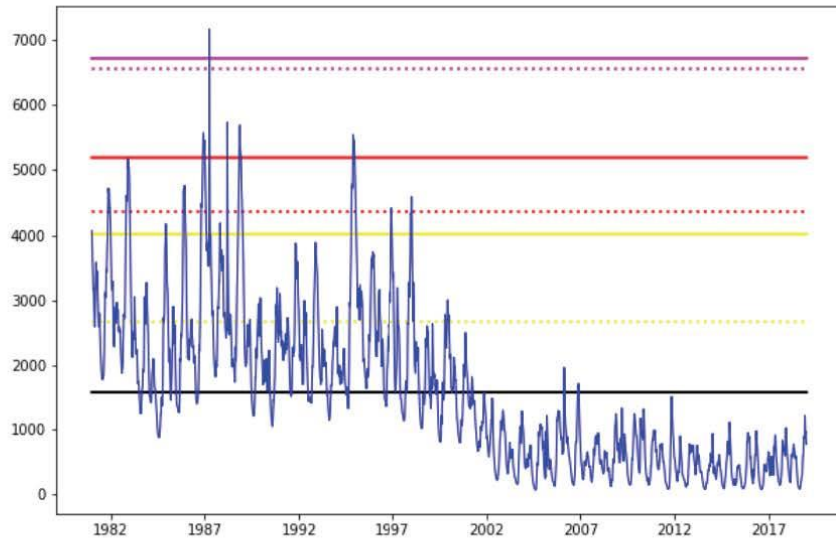


Figure 13. Example of the daily GloFAS-ERA5 river discharge (m^3/s) time series for an upstream catchment in the Congo river basin and the 2- (yellow lines), 5- (red lines) and 20-year (magenta lines) flood thresholds based on two different fitting methods. The dashed lines represent the flood thresholds fitted using all annual maxima, while the solid lines were produced using only annual maxima over the sample mean discharge shown by the black line.

A potential driver of the exposed ERA5 trends is the method of production: ERA5 is produced in several streams that were later merged into one consolidated data set. These streams have a year overlap to allow for long enough spin-up. According to Hersbach et al. (2020), in the deep soil, where spin-up can take several years, discontinuities could be observed. The deep soil can certainly impact on the runoff through the sub-surface runoff, which can impact on river discharge, however, it is not expected to have any noticeable impact on variables such as precipitation or evaporation which vary on a significantly shorter timescale.

The GloFAS-ERA5 linear trends, presented here, have a direct impact on the quality of the GloFAS flood warnings through the use of the flood thresholds. The presence of significant trends, or a very substantial regime change, such as in the example provided in Figure 13, makes it difficult to produce flood thresholds that correctly represent the extreme event behaviour in the forecasts. In the provided example, the river discharge level collapses to less than a third after 2000. Thus, the thresholds will be much too high and represent only the first half of the period. In this case, the real time GloFAS forecast will likely be similar to the latter part of the reanalysis period and will hardly ever exceed these flood thresholds, making the flood warnings very unreliable.

In other catchments, where the river discharge is significantly increasing, the situation is the opposite. In such catchments the real time forecasts will be mostly similar to the latter part of the reanalysis and the flood thresholds will be biased towards the lower earlier years. In this case, the thresholds will be

little too low and the real time GloFAS forecasts are expected to show too frequent flood events, making the forecasts unreliable again.

The example in Figure 13 is a very extreme one from the upstream part of the Congo basin. However, it is not an isolated case causing problems, as areas where the linear trend shows at least ~20% change in the ~40-year period, extreme value fitting to compute the flood thresholds is likely to provide us with difficulties. There is no scientific basis for this 20% minimum value (equivalent to ~0.05 raw linear trend magnitude), further tests would be needed to identify the expected forecast reliability loss due to such trends in the reanalysis time series. Based on the current reanalysis period, as used in GloFAS, at least 50% of land areas show raw linear trends higher (lower) than 0.05 (-0.05) (see Figure 6 and also Table 2) causing a significant issue in the operational running of GloFAS.

The extension of ERA5 back to 1950 (Hersbach et al., 2020) will provide an increase in the period used in the flood thresholds computation. However, it is strongly recommended that hydrological trends, with any potential discontinuities or regime changes due to merging different streams in the ERA5 production, should be analysed in that extended data set before making any change in GloFAS.

In addition, it will be important to explore ways to better derive the flood thresholds for catchments that are impacted by large trends in the GloFAS-ERA5 river discharge. An option could be to limit the period for the flood threshold computation to most recent decades, on a case by case basis (for single catchments, or maybe whole regions) selecting the period length that provides historical time series with low enough trend magnitudes. Based on this study, and the current ERA5 reanalysis, this could be as short as the last 15-17 years from ~2004, which although not ideal for estimating longer return periods, may provide improvements to make the GloFAS flood warnings more reliable in the future.

Zsoter et al. 2020 recommends that river discharge ensemble reforecasts should be used to compute flood thresholds instead of reanalysis, which would help to create more reliable flood warnings especially at longer lead times. The use of the reforecasts would be beneficial, as although they are initialised from GloFAS-ERA5 and are thus bound to inherit any trend problems that are present in this reanalysis, they are only generated on the most recent 20-year period (currently 1999-2018), which would undoubtedly lessen the trend impact documented in this study.

Finally, the documented river discharge linear trends in the GloFAS-ERA5 dataset are large enough to warrant further investigation of the underlying causes to the general behaviour of the water cycle variables in ERA5. This is crucial in order to provide improvements in hydrological variables such as river discharge, especially in the context of any future version of ECMWF reanalysis data sets, such as ERA6.

References

- Anabalón A., and A. Sharma, 2017: On the divergence of potential and actual evapotranspiration trends: An assessment across alternate global datasets. *Earth's Future*, **5**, 905–917.
- Albergel C., and Coauthors, 2013: Skill and Global Trend Analysis of Soil Moisture from Reanalyses and Microwave Remote Sensing. *J. Hydrometeor.*, **14**, 1259–1277, <https://doi.org/10.1175/JHM-D-12-0161.1>.

- 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. <http://dx.doi.org/10.1175/2008JHM1068.1>.
- Beck H. E., E. F. Wood, M. Pan, and C. K. Fisher, 2019: MSWEP V2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment. *Bull. Amer. Meteor. Soc.* **100**, 473–500, <https://doi.org/10.1175/BAMS-D-17-0138.1>.
- Connolly R., M. Connolly, W. Soon, D. R. Legates, R. G. Cionco, and H. V. M. Velasco Herrera, 2019: Northern Hemisphere Snow-Cover Trends (1967–2018): A Comparison between Climate Models and Observations. *Geosciences*, **9**, 135. <https://doi.org/10.3390/geosciences9030135>.
- Cucchi M., G. P. Weedon, A. Amici, N. Bellouin, S. Lange, H. M. Schmied, H. Hersbach, and C. Buontempo, 2020: WFDE5: bias adjusted ERA5 reanalysis data for impact studies, *Earth Syst. Sci. Data Discuss.*, <https://doi.org/10.5194/essd-2020-28>, in review, 2020.
- Dai A., T. Qian, K. E. Trenberth, and J. D. Milliman, 2009: Changes in continental freshwater discharge from 1948 to 2004. *Journal of Climate*, **22**(10), 2773–2792. <https://doi.org/10.1175/2008JCLI2592.1>.
- Dorigo W., R. de Jeu, D. Chung, R. Parinussa, Y. Liu, W. Wagner, and D. Fernández-Prieto, 2012: Evaluating global trends (1988–2010) in harmonized multi-satellite surface soil moisture. *Geophys. Res. Lett.*, **39** (18), 10.1029/2012GL052988.
- Drusch M., D. Vasiljevic, and P. Viterbo, 2004: ECMWF's Global Snow Analysis: Assessment and Revision Based on Satellite Observations. *J. Appl. Meteor.*, **43**, 1282–1294, [https://doi.org/10.1175/1520-0450\(2004\)043<1282:EGSAAA>2.0.CO;2](https://doi.org/10.1175/1520-0450(2004)043<1282:EGSAAA>2.0.CO;2).
- Dutra E., P. Viterbo, P. M. A. Miranda, and G. Balsamo, 2012: Complexity of Snow Schemes in a Climate Model and Its Impact on Surface Energy and Hydrology. *J. Hydrometeor.*, **13**, 521–538, <https://doi.org/10.1175/JHM-D-11-072.1>.
- Feng H., and M. Zhang, 2016: Global land moisture trends: drier in dry and wetter in wet over land. *Sci Rep*, **5**, 18018. <https://doi.org/10.1038/srep18018>.
- Hansen J., S. Makiko, R. Ruedy, K. Lo, D. W. Lea, and M. Medina-Elizade. 2006: Global temperature change. *PNAS*, **6**, 14288–14293; <https://doi.org/10.1073/pnas.0606291103>.
- Hancock S., R. Baxter, J. Evans, and B. Huntley, 2013: Evaluating global snow water equivalent products for testing land surface models. *Remote Sens. Environ.*, **128**, 107–117, <https://doi.org/10.1016/j.rse.2012.10.004>.
- Harrigan S., E. Zsoter, L. Alfieri, C. Prudhomme, P. Salamon, F. Wetterhall, H. Cloke, and F. Pappenberger, 2020: GloFAS-ERA5 operational global river discharge reanalysis 1979–present, *Earth Syst. Sci. Data*, <https://doi.org/10.5194/essd-12-2043-2020>.
- Hersbach H., & Coauthors, 2020: The ERA5 Global Reanalysis. *Q. J. R. Meteorol. Soc.*, <https://doi.org/10.1002/qj.3803>.
- IPCC AR4 report, 2007, https://www.ipcc.ch/site/assets/uploads/2018/05/ar4_wg1_full_report-1.pdf.

- IPCC AR5 report WG1 contribution, 2013, https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_all_final.pdf.
- IPCC AR5 report, 2014, https://www.ipcc.ch/site/assets/uploads/2018/05/SYR_AR5_FINAL_full_wcover.pdf.
- Javadian M., A. Behrangi, W. K. Smith, and J. B. Fisher, 2020: Global Trends in Evapotranspiration Dominated by Increases across Large Cropland Regions. *Remote Sens.*, **12**(7), <https://doi.org/10.3390/rs12071221>.
- Jung M., M. Reichstein, and P. Ciais, 2020: Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, **467**, 951–954, <https://doi.org/10.1038/nature09396>.
- Knowles N., 2015: Trends in Snow Cover and Related Quantities at Weather Stations in the Conterminous United States. *J. Climate*, **28**, 7518–7528, <https://doi.org/10.1175/JCLI-D-15-0051.1>.
- Kunkel K. E., D. A. Robinson, and S. Champion, 2016: Trends and Extremes in Northern Hemisphere Snow Characteristics. *Curr Clim Change Rep*, **2**, 65–73, <https://doi.org/10.1007/s40641-016-0036-8>.
- 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.*, doi:10.1088/1748-9326/ab52b2.
- Nguyen P., A. Thorstensen, S. Sorooshian, K. Hsu, A. Aghakouchak, H. Ashouri, H. Tran, and D. Braithwaite, 2018: Global precipitation trends across spatial scales using satellite observations. *Bull. Amer. Meteor. Soc.*, **99**, 689–697, <https://doi.org/10.1175/BAMS-D-17-0065.1>.
- Pan N., S. Wand, Y. Liu, W. Zhao, and B. Fu, 2019: Global Surface Soil Moisture Dynamics in 1979–2016 Observed from ESA CCI SM Dataset. *Water*, **11**(5), 883; <https://doi.org/10.3390/w11050883>.
- Parmesan C., and G. Yohe, 2003: Globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37–42, <https://doi.org/10.1038/nature01286>.
- Pavelsky T.M., M. T. Durand, K. M. Andreadis, R. E. Beighley, R. C. D. Paiva, G. H. Allen, and Z. F. Miller, 2014: Assessing the potential global extent of SWOT river discharge observations. *Journal of Hydrology*, **519** (PB), 1516–1525. <https://doi.org/10.1016/j.jhydrol.2014.08.044>.
- Rodda J. C., S. A. Pieyns, N. S. Sehmi, and G. Matthews, 1993: Towards a world hydrological cycle observing system *Hydrol. Sci. J.* **38**, 373–378. <https://doi.org/10.1080/026266693099492687>.
- de Rosnay P., L. Isaksen, and M., Dahoui, 2015: Snow data assimilation at ECMWF. *ECMWF newsletter*, **143**, <https://www.ecmwf.int/sites/default/files/elibrary/2015/17328-snow-data-assimilation-ecmwf.pdf>.
- de Rosnay P., G. Balsamo, and C. Albergel, 2014: Initialisation of Land Surface Variables for Numerical Weather Prediction. *Surv Geophys*, **35**, 607–621, <https://doi.org/10.1007/s10712-012-9207-x>.

- Stahl K., L. M. Tallaksen, J. Hannaford, and H. A. J. van Lanen, 2012: Filling the white space on maps of European runoff trends: estimates from a multi-model ensemble. *Hydrology Earth System Sciences*, **16** (7), 2035–2047.
- Su L., C. Miao, D. Kong, Q. Duan, X. Lei, Q. Hou, and H. Li, 2018: Long-term trends in global river flow and the causal relationships between river flow and ocean signals. *Journal of Hydrology*. <https://doi.org/10.1016/j.jhydrol.2018.06.058>.
- Sun Q., C. Miao, Q. Duan, H. Ashouri, S. Sorooshian, and K. L. Hsu, 2018: A review of global precipitation datasets: Data sources, estimation, and intercomparisons. *Rev. Geophys.*, **56**, 79–107, <https://doi.org/10.1002/2017RG000574>.
- 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. Inf. Sci.*, **24**, 189–212, <https://doi.org/10.1080/13658810802549154>.
- Zhang Y., J. Peña-Arancibia, and T. McVicar, 2016: Multi-decadal trends in global terrestrial evapotranspiration and its components. *Sci Rep*, **6**, 19124, <https://doi.org/10.1038/srep19124>.
- Zsoter E., H. Cloke, E. Stephens, P. de Rosnay, J. Muñoz-Sabater, C. Prudhomme, and F. Pappenberger, 2019: How well do operational Numerical Weather Prediction setups represent hydrology? *J. Hydrometeorol.*, **14**, 1533–1552, <https://doi.org/10.1175/JHM-D-18-0086.1>.
- Zsoter E., C. Prudhomme, E. Stephens, F. Pappenberger and H. Cloke, 2020: Using ensemble reforecasts to generate flood thresholds for improved global flood forecasting. *J. Flood Risk Management*, <https://doi.org/10.1111/jfr3.12658>.

Appendix A: ERA5 and ERA5-Land trends for some additional surface variables

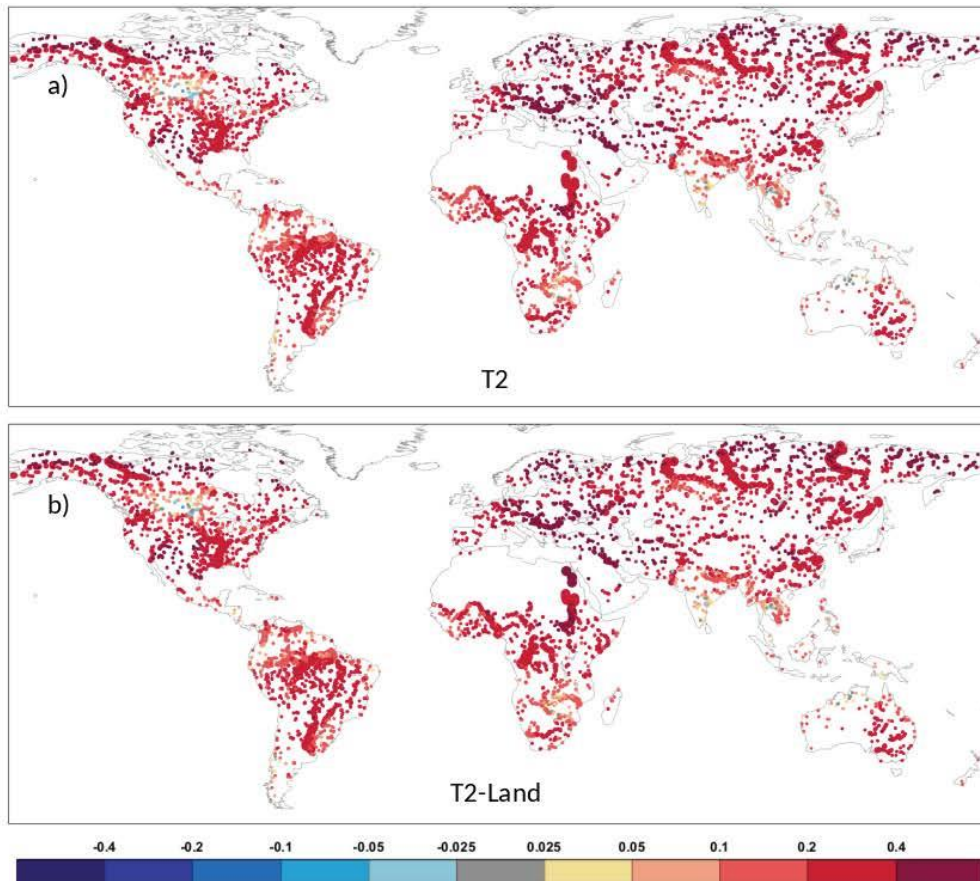


Figure A1. Raw linear trend (K) for global river catchments for ERA5 (a) and ERA5-Land (b) 2m temperature, based on the 1981-2018 period. The circles represent the catchment outlets, while their size the catchment area.

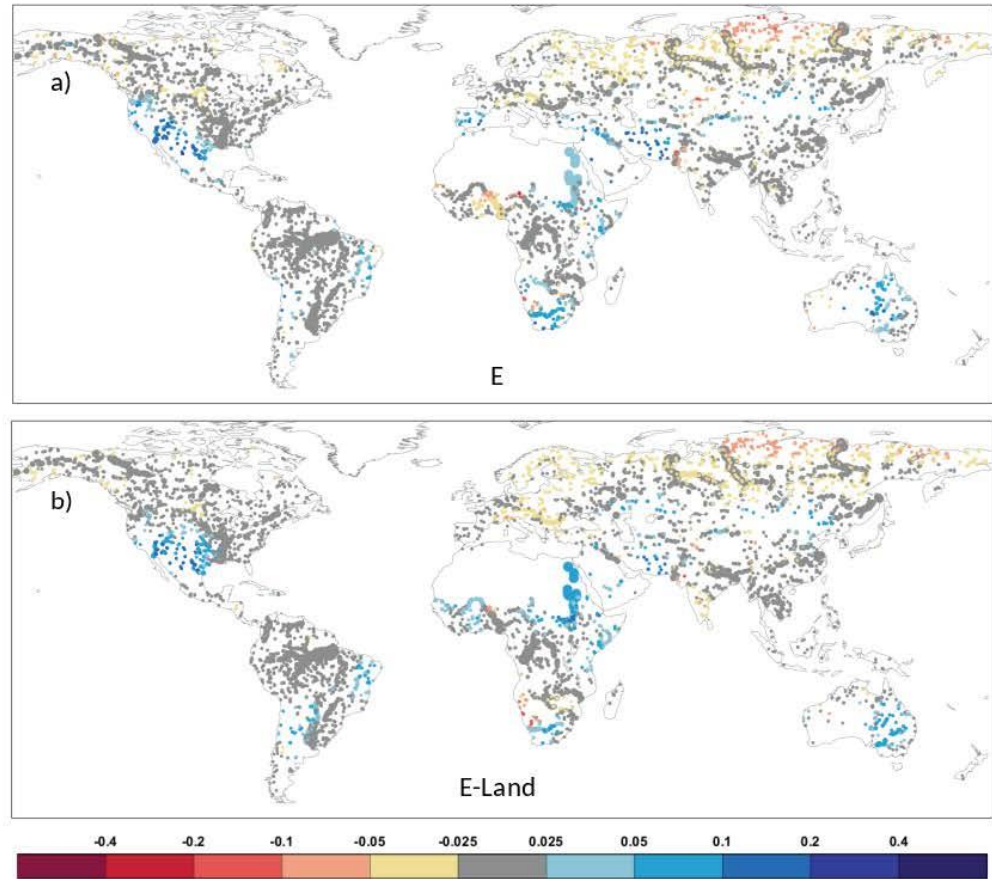


Figure A2. Raw linear trend (fraction/decade) for global river catchments for ERA5 (a) and ERA5-Land (b) evaporation, based on the 1981-2018 period. The circles represent the catchment outlets, while their size the catchment area.

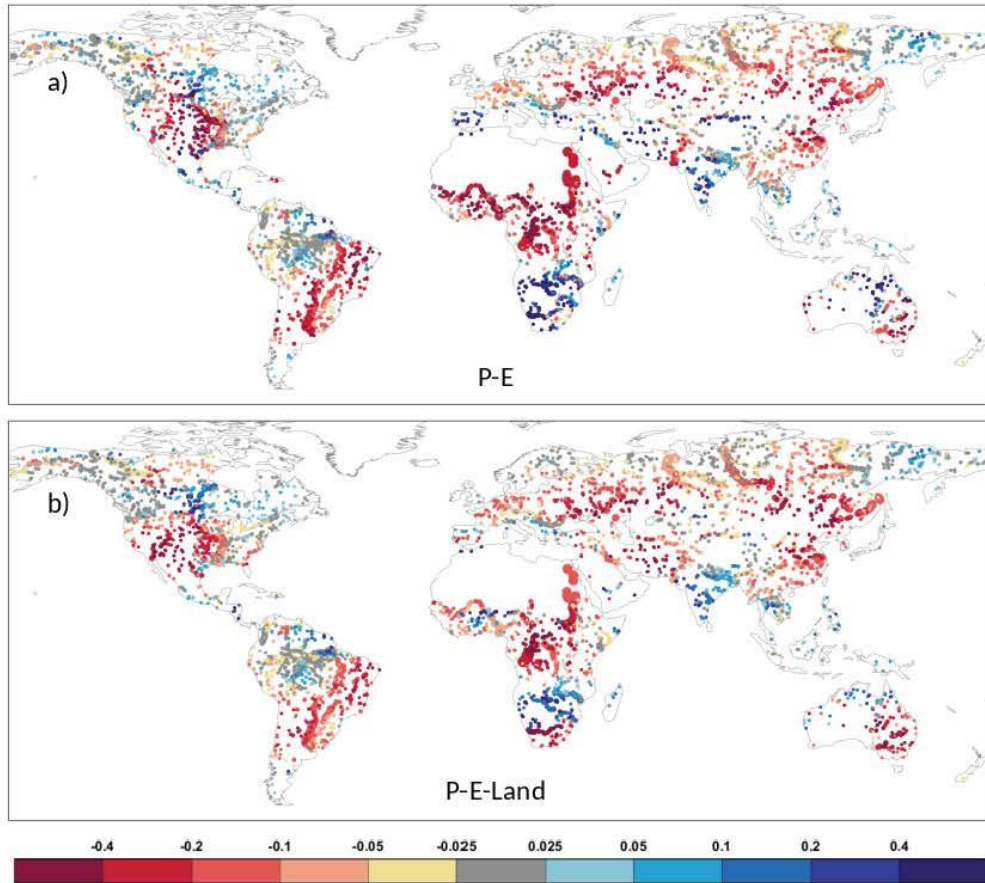


Figure A3. Raw linear trend (fraction/decade) for global river catchments for ERA5 (a) and ERA5-Land (b) evaporation, based on the 1981-2018 period. The circles represent the catchment outlets, while their size the catchment area.

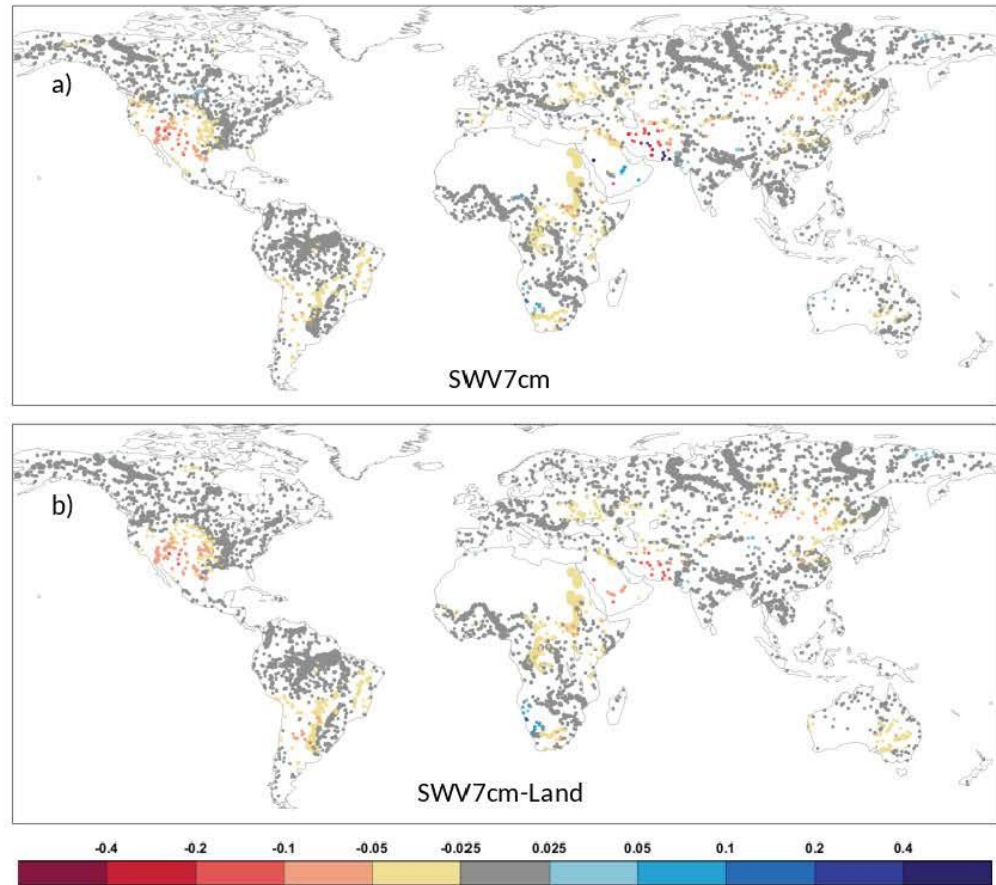


Figure A4. Raw linear trend (fraction/decade) for global river catchments for ERA5 (a) and ERA5-Land (b) soil moisture in the top 0-7 cm layer of the soil, based on the 1981-2018 period. The circles represent the catchment outlets, while their size the catchment area.

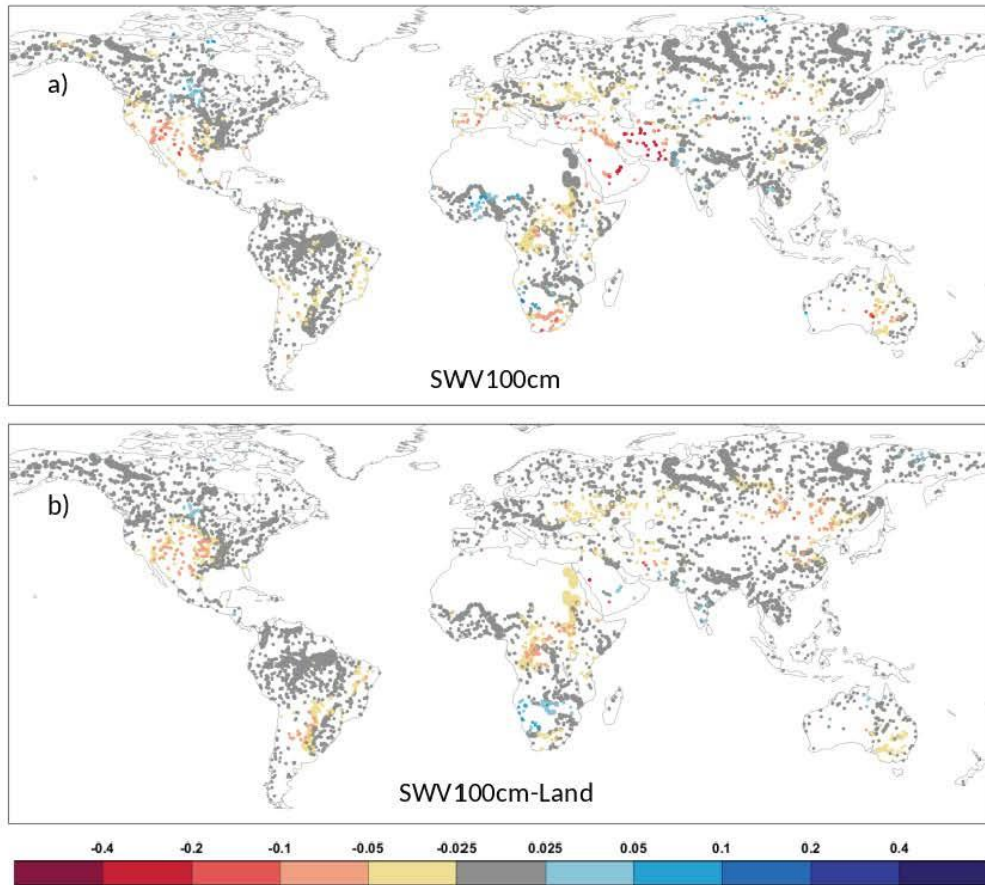


Figure A5. Raw linear trend (fraction/decade) for global river catchments for ERA5 (a) and ERA5-Land (b) soil moisture in the top 0-100 cm layer of the soil, based on the 1981-2018 period. The circles represent the catchment outlets, while their size the catchment area.

Appendix B: Standardised trend magnitudes for selected global catchments

Table A1. Standardised linear trends (fraction of standard deviation / decade) for selected catchments (see Figure 4) for 1981-2018, for GloFAS-ERA5 river discharge (DIS), ERA5 precipitation (P) and snowfall (SF) and both ERA5 and ERA5-Land snowmelt (SMLT), runoff (RO), evaporation (E), precipitation minus evaporation (P-E), soil moisture in the top 0-7 cm (SWV7) and 0-100 cm (SWV100) and 2m temperature (T2). Raw linear trends are also provided for observed river discharge (OBS) and the observation availability matched GloFAS-ERA5 river discharge (DISm) and ERA5 and ERA5-Land runoff (ROm). Differences in the absolute raw linear trend errors between ERA5-Land and ERA5 are also indicated in the last column (Imp). Empty cells correspond to cases for which trend computation was not possible. Coloured cells indicate negative (orange), positive (blue) trends and also decreasing (green) and increasing (purple) trend errors in ERA5 (Imp column). Where there is no raw linear trend, defined for absolute values less than 0.025, cells are not coloured. Darkening shades show increasing trend magnitudes.

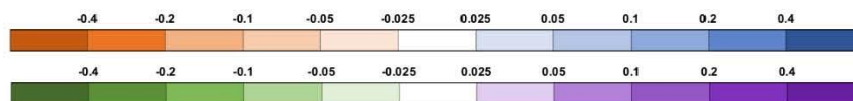
Rivers		DIS	P	SF	SMLT		RO		E		P-E	
		ERA5	ERA5	ERA5	ERA5	ERA5-Land	ERA5	ERA5-Land	ERA5	ERA5-Land	ERA5	ERA5-Land
North Asia	Ob	-0.50	-0.07	-0.30	-0.70	-0.25	-0.58	-0.40	-0.58	-0.45	-0.28	-0.23
	Yenisei	-0.81	-0.30	-0.29	-0.99	-0.22	-0.82	-0.49	-0.47	-0.49	-0.44	-0.44
	Lena	-0.57	-0.07	-0.36	-1.11	-0.21	-0.69	-0.27	-0.49	-0.63	-0.21	-0.26
	Amur	-0.63	-0.44	-0.33	-0.98	-0.26	-0.70	-0.63	-0.24	-0.14	-0.48	-0.46
North America	Yukon	-0.60	-0.10	-0.21	-0.65	-0.18	-0.73	-0.22	-0.03	-0.27	-0.10	-0.21
	Mackenzie	-0.54	-0.03	-0.06	-0.43	-0.04	-0.75	-0.13	-0.22	-0.23	-0.11	-0.12
	Churchill	0.12	0.27	0.15	0.30	0.10	0.19	0.37	-0.19	-0.20	0.22	0.22
	Columbia	-0.38	-0.15	0.04	-0.30	0.06	-0.39	-0.06	0.65	0.34	0.04	-0.06
	Colorado	-0.62	-0.58	-0.44	-0.39	-0.44	-0.69	-0.66	0.66	0.61	-0.06	-0.18
Mississippi	-0.28	-0.21	-0.34	-0.32	-0.32	-0.31	-0.31	-0.17	0.19	-0.28	-0.17	
Europe	Kalixaelven	0.04	0.24	0.10	-0.26	0.12	0.04	0.22	-0.44	-0.48	0.17	0.17
	Thames	-0.25	0.06	-0.14	0.11	-0.13	-0.24	0.00	0.06	-0.30	0.09	-0.02
	Rhine	-0.40	-0.23	-0.26	-0.50	-0.32	-0.41	-0.40	-0.48	-0.59	-0.33	-0.35
	Danube	-0.41	0.12	-0.18	-0.46	-0.24	-0.44	-0.21	-0.51	-0.72	0.00	-0.06
	Volga	-0.68	-0.37	-0.22	-0.59	-0.18	-0.69	-0.53	-0.65	-0.36	-0.53	-0.46
Ebro	-0.44	0.02	0.13	-0.49	0.11	-0.53	-0.06	0.66	-0.03	0.42	0.01	
South America	Magdalena	-0.25	-0.23				-0.28	-0.25	0.09	-0.56	-0.22	-0.26
	Amazon	-0.18	-0.07	-0.47	-0.50	-0.42	-0.18	-0.11	-0.11	-0.43	-0.08	-0.12
	Araguaia	-0.53	-0.52				-0.56	-0.54	0.12	0.43	-0.48	-0.42
	Parana	-0.45	-0.55	-0.57	-0.62	-0.54	-0.50	-0.59	0.17	0.70	-0.59	-0.44
Africa	Nile	-0.59	-0.70				-0.67	-0.78	0.49	0.88	-0.76	-0.51
	Niger	-0.28	-0.20				-0.32	-0.45	-0.62	0.15	-0.61	-0.18
	Shabelle	-0.21	-0.17				-0.23	-0.17	0.19	0.22	-0.12	-0.08
	White-Volta	0.24	-0.20				0.27	0.17	-0.62	0.46	-0.69	0.09
	Congo	-0.71	-0.81				-0.87	-0.91	0.38	0.61	-0.83	-0.80
	Zambesi	0.17	0.09				0.28	0.33	0.70	0.31	0.31	0.19
	Cunene	0.27	0.33				0.24	0.40	0.50	-0.51	0.79	0.09
Orange	-0.22	-0.32	-0.03	-0.05	-0.04	-0.30	-0.37	0.63	0.33	0.24	-0.18	
South Asia	Brahmaputra	-0.12	-0.19	-0.50	-0.05	-0.50	-0.12	-0.20	-0.45	-0.25	-0.21	-0.20
	Ganges	0.30	0.25	-0.34	0.12	-0.32	0.32	0.29	-0.39	-0.18	0.17	0.21
	Godavari	0.24	0.27				0.26	0.29	-0.36	-0.26	0.20	0.22
	Karnali	0.23	0.11	-0.38	-0.25	-0.38	0.23	0.09	-0.40	-0.25	0.06	0.07
Australia	Murray	-0.13	-0.30	-0.29	-0.42	-0.21	-0.26	-0.33	0.28	0.35	-0.25	-0.15



(continues on next page)

Table A1. Standardised linear trends (fraction of standard deviation / decade) for selected catchments (see Figure 4) for 1981-2018 (continued)

	Rivers	SWV 0-7		SWV 0-100		T2		OBS	DISm	ROm		
		ERA5	ERA5-Land	ERA5	ERA5-Land	ERA5	ERA5-Land	-	ERA5	ERA5	ERA5-Land	Imp
North Asia	Ob	-0.25	-0.28	-0.27	-0.39	0.26	0.25	0.16	-0.70	-0.83	-0.34	-0.49
	Yenisei	-0.51	-0.55	-0.31	-0.72	0.38	0.37	0.43	-1.14	-1.20	-0.49	-0.72
	Lena	-0.27	-0.06	-0.16	-0.04	0.43	0.42	0.30	-0.82	-0.88	-0.43	-0.46
	Amur	-0.70	-0.56	-0.34	-0.70	0.46	0.44	-1.36	-2.05	-2.27	-1.93	-0.34
North America	Yukon	-0.40	-0.11	-0.61	-0.06	0.41	0.40	0.20	-1.05	-1.28	-0.40	-0.88
	Mackenzie	-0.34	-0.14	-0.39	-0.15	0.35	0.34	0.39	-0.67	-0.95	-0.11	-0.84
	Churchill	0.08	0.18	0.18	0.22	0.12	0.11	0.63	0.26	0.36	0.37	-0.01
	Columbia	-0.62	-0.35	-0.66	-0.28	0.45	0.43	-0.10	-0.52	-0.53	-0.21	-0.32
	Colorado	-0.59	-0.58	-0.77	-0.48	0.56	0.56	-0.21	-0.95	-1.04	-1.01	-0.03
	Mississippi	-0.34	-0.40	-0.16	-0.46	0.31	0.32	0.15	-0.38	-0.44	-0.51	0.07
Europe	Kalixaelven	0.10	-0.01	0.09	0.01	0.51	0.53	0.24	-0.23	-0.24	0.11	-0.35
	Thames	-0.17	-0.02	-0.43	-0.07	0.42	0.42	-0.01	-0.43	-0.40	-0.14	-0.26
	Rhine	-0.28	-0.26	-0.38	-0.32	0.53	0.52	-0.30	-0.57	-0.57	-0.46	-0.12
	Danube	-0.25	-0.15	-0.40	-0.14	0.62	0.62	0.56	-0.44	-0.41	-0.03	-0.37
	Volga	-0.47	-0.50	-0.46	-0.58	0.41	0.41	-0.08	-1.22	-1.25	-0.87	-0.38
	Ebro	-0.48	-0.17	-1.02	-0.20	0.53	0.50	-0.26	-0.86	-1.06	-0.48	-0.59
South America	Magdalena	-0.48	-0.34	-0.63	-0.35	0.51	0.50	0.03	-0.03	-0.06	-0.03	-0.03
	Amazon	-0.67	-0.62	-0.73	-0.62	0.60	0.60	0.25	-0.46	-0.46	-0.40	-0.07
	Araguaia	-0.68	-0.68	-0.46	-0.52	0.64	0.69	-0.14	-0.62	-0.70	-0.67	-0.03
	Parana	-0.58	-0.69	-0.52	-0.80	0.67	0.70	-0.20	-0.82	-0.94	-1.10	0.16
Africa	Nile	-0.66	-0.78	-0.54	-0.87	0.75	0.78	0.19	-1.46	-1.59	-1.88	0.29
	Niger	-0.01	-0.25	0.49	-0.19	0.64	0.66	1.35	-0.52	-0.54	-0.79	0.24
	Shabelle	-0.23	-0.24	-0.12	-0.11	0.63	0.64	0.32	-0.97	-1.06	-0.85	-0.21
	White-Volta	-0.09	-0.35	0.87	0.01	0.48	0.59	-0.28	-0.13	-0.14	-0.20	-0.06
	Congo	-0.71	-0.73	-0.85	-1.04	0.75	0.76	0.27	-1.59	-1.89	-1.99	0.10
	Zambesi	-0.35	-0.20	-0.27	0.04	0.53	0.50	0.48	0.33	0.53	0.61	0.08
	Cunene	-0.09	0.25	-0.42	0.55	0.52	0.41	0.40	0.40	0.39	0.54	0.14
	Orange	-0.34	-0.29	-0.61	-0.25	0.45	0.43	0.17	-0.78	-0.95	-0.97	0.02
South Asia	Brahmaputra	0.44	0.22	0.46	0.23	0.45	0.45	-0.32	-0.12	-0.12	-0.20	-0.08
	Ganges	0.29	0.12	0.37	0.23	0.36	0.35	-0.03	0.45	0.48	0.45	-0.02
	Godavari	0.23	0.15	0.38	0.37	0.15	0.16	-0.04	0.41	0.44	0.48	0.04
	Karnali	0.01	-0.24	0.09	-0.15	0.50	0.46	-0.05	0.29	0.29	0.14	-0.14
Australia	Murray	-0.22	-0.34	-0.40	-0.33	0.49	0.52	-0.25	-0.39	-0.64	-0.74	0.10

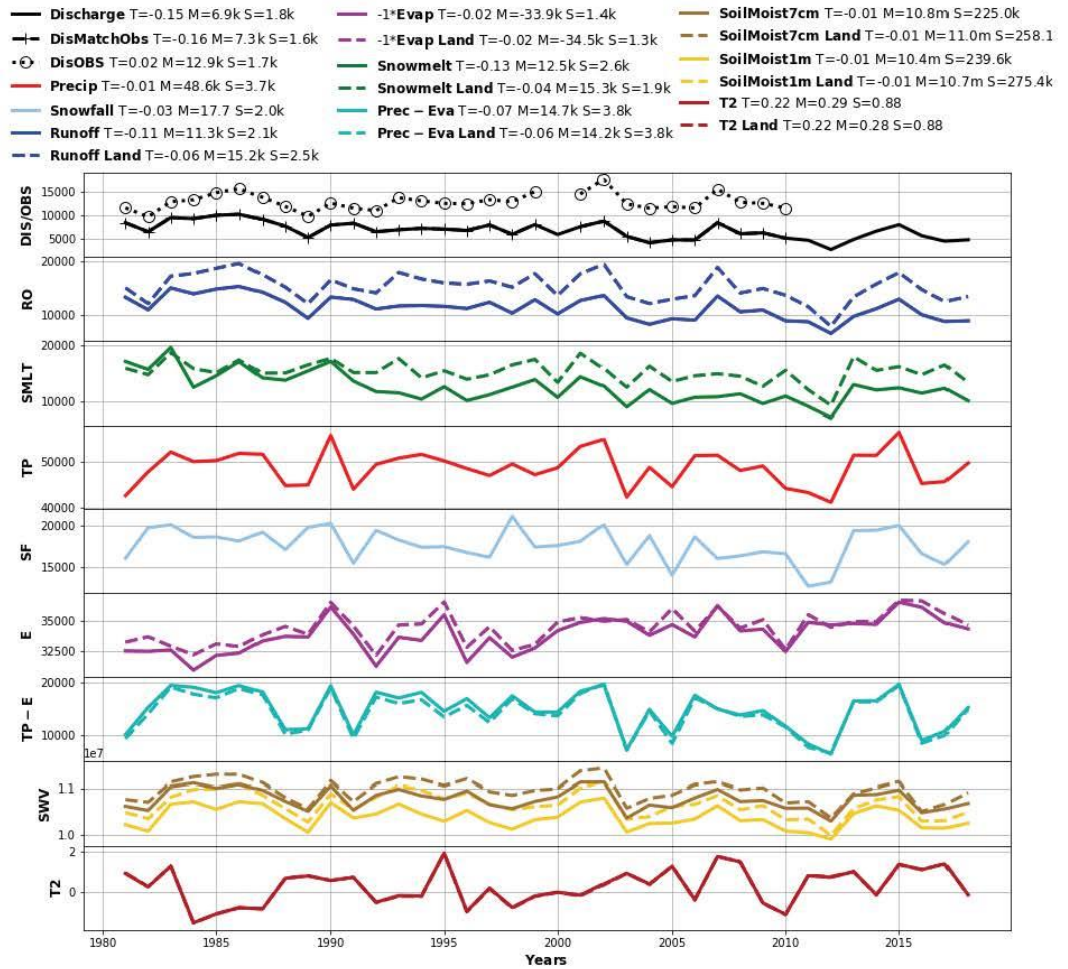


Appendix C: Trend hydrographs for selected global catchments

Annual mean time series diagrams with all the analysed variables (Table 1) are provided for several world rivers (33 in total), grouped into continents (see Figure 4 for the locations). These graphs include in the line legends also the trend magnitudes and the mean and the standard deviation values of the annual mean time series in the 38-year period for each variable (other than the observed river discharge which includes trend values only if at least 16 years were available).

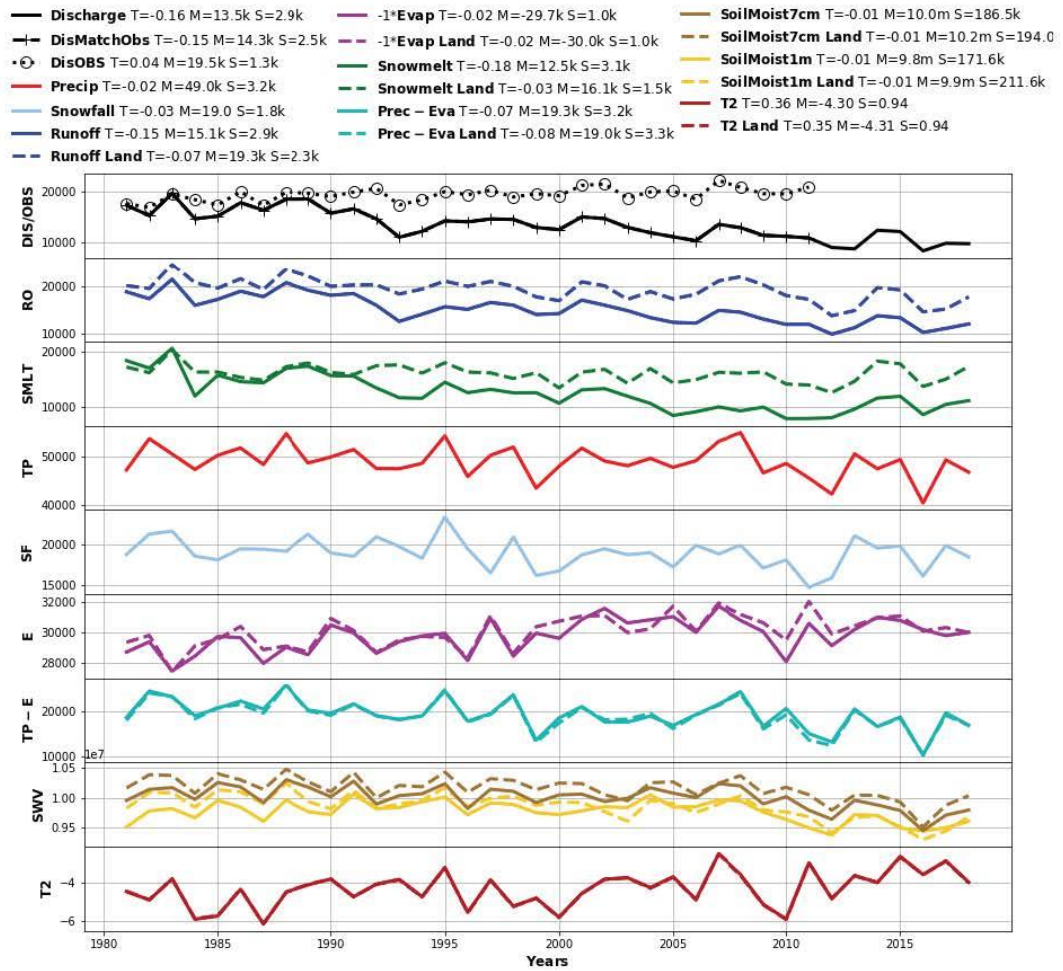
North Asia - No.1

River: Ob, Outlet: [66.55N; 66.45E], Area: 2541170 km2



North Asia - No.2

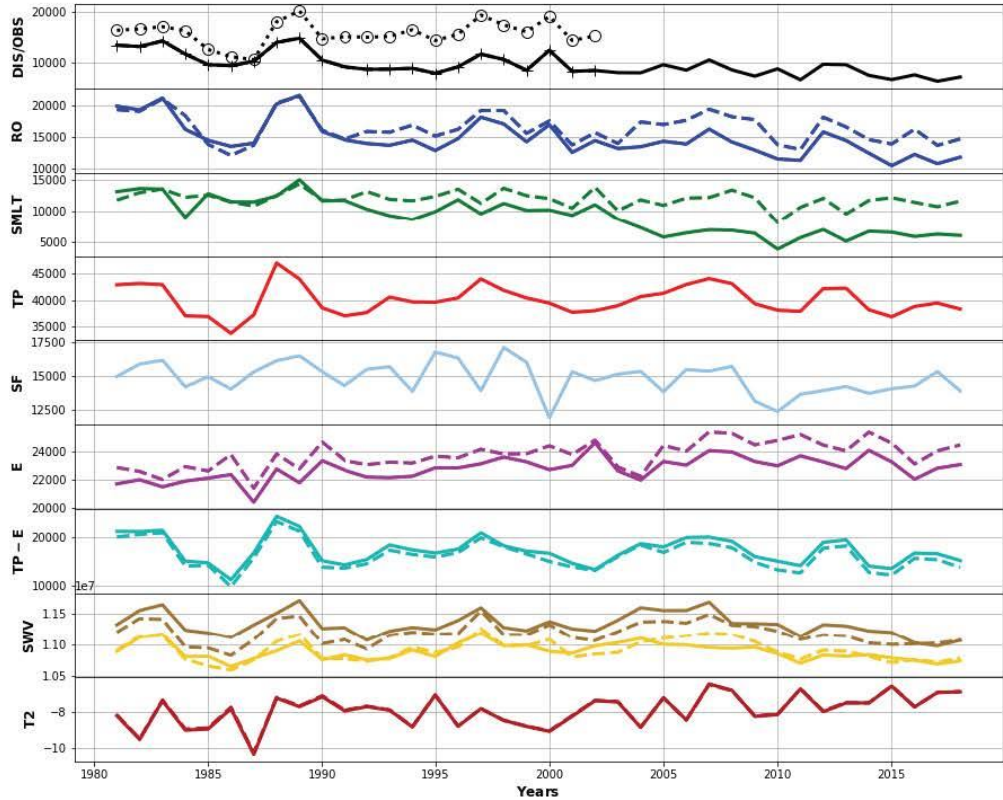
River: Yenisei, Outlet: [67.45N; 86.45E], Area: 2651150 km²



North Asia - No.3

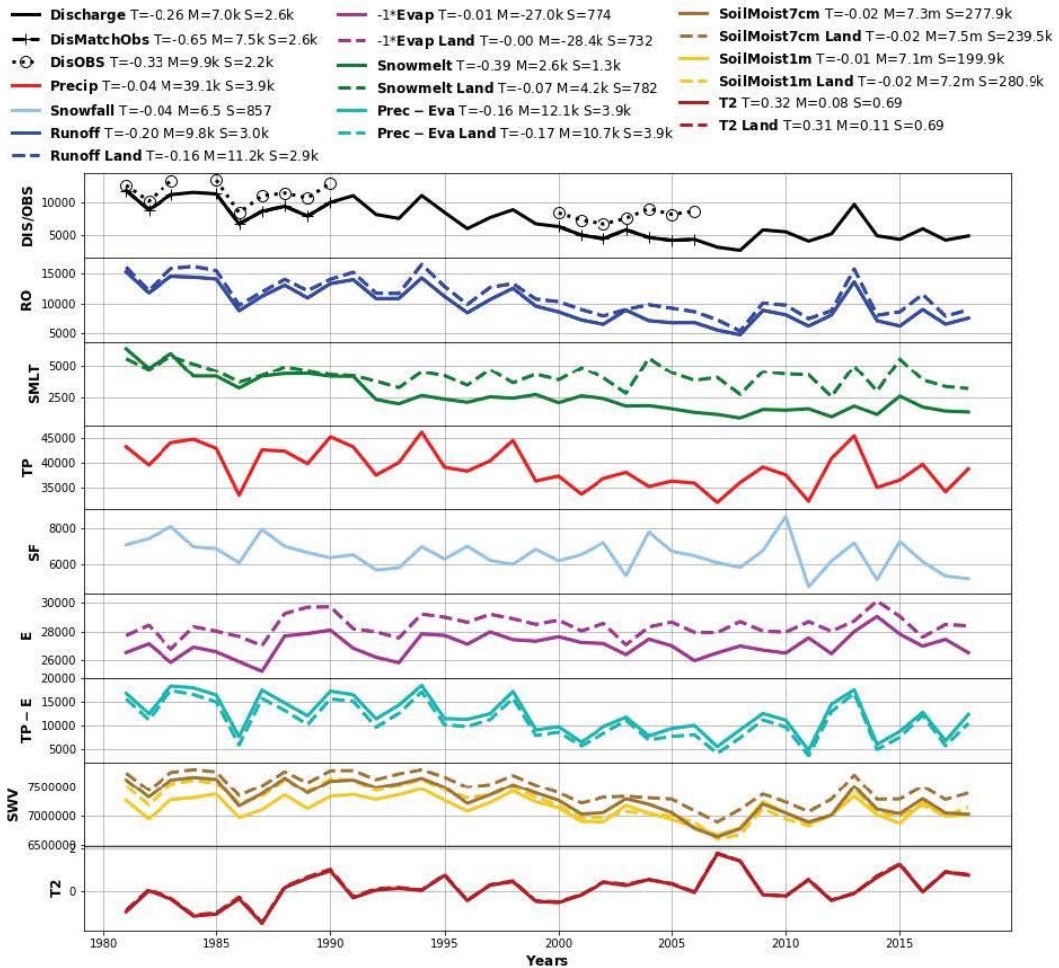
River: Lena, Outlet: [72.25N; 126.75E], Area: 2443690 km2

- Discharge T=-0.14 M=9.7k S=2.2k
- DisMatchObs T=-0.17 M=10.7k S=2.1k
- DisOBS T=0.04 M=15.8k S=2.3k
- Precip T=-0.00 M=40.0k S=2.7k
- Snowfall T=-0.03 M=14.9 S=1.2k
- Runoff T=-0.11 M=14.9k S=2.8k
- Runoff Land T=-0.04 M=16.6k S=2.3k
- -1*Evap T=-0.02 M=-22.8k S=812
- -1*Evap Land T=-0.02 M=-23.8k S=950
- Snowmelt T=-0.24 M=9.2k S=2.8k
- Snowmelt Land T=-0.04 M=11.9k S=1.2k
- Prec - Eva T=-0.04 M=17.3k S=2.8k
- Prec - Eva Land T=-0.05 M=16.3k S=2.9k
- SoilMoist7cm T=-0.00 M=11.3m S=180.7k
- SoilMoist7cm Land T=-0.00 M=11.2m S=171.9
- SoilMoist1m T=-0.00 M=10.9m S=133.7k
- SoilMoist1m Land T=-0.00 M=10.9m S=169.6k
- T2 T=0.37 M=-7.93 S=0.87
- T2 Land T=0.36 M=-7.91 S=0.86



North Asia - No.4

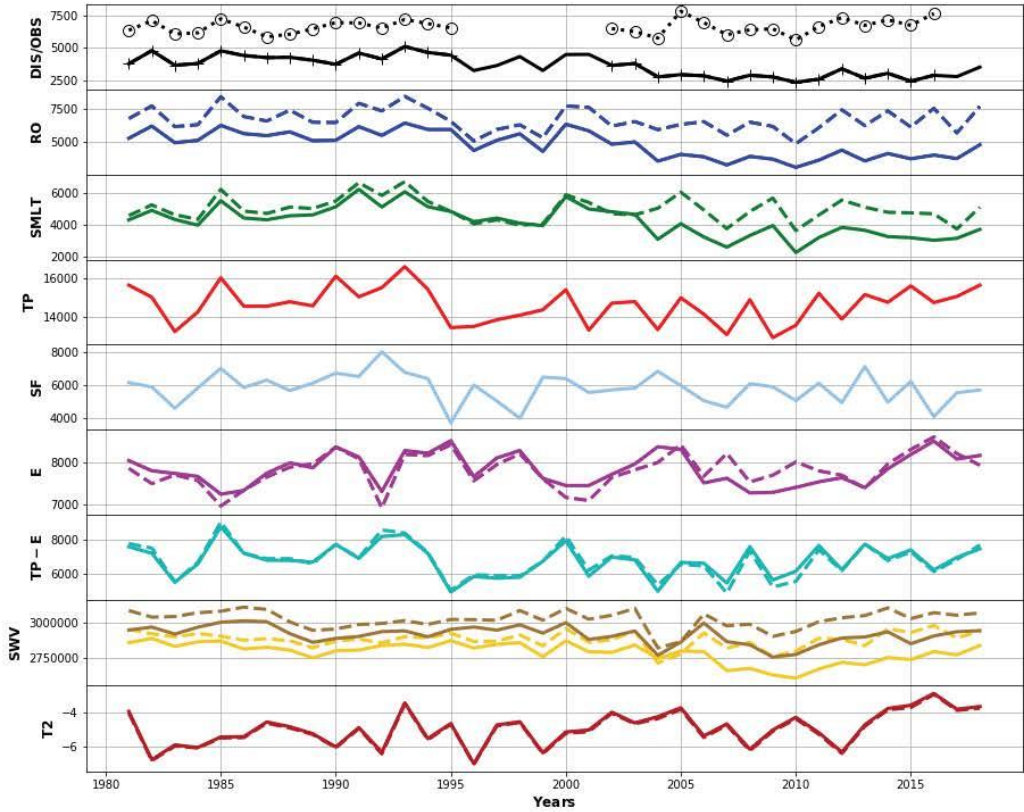
River: Amur, Outlet: [50.55N; 137.15E], Area: 1846250 km2



North America - No.1

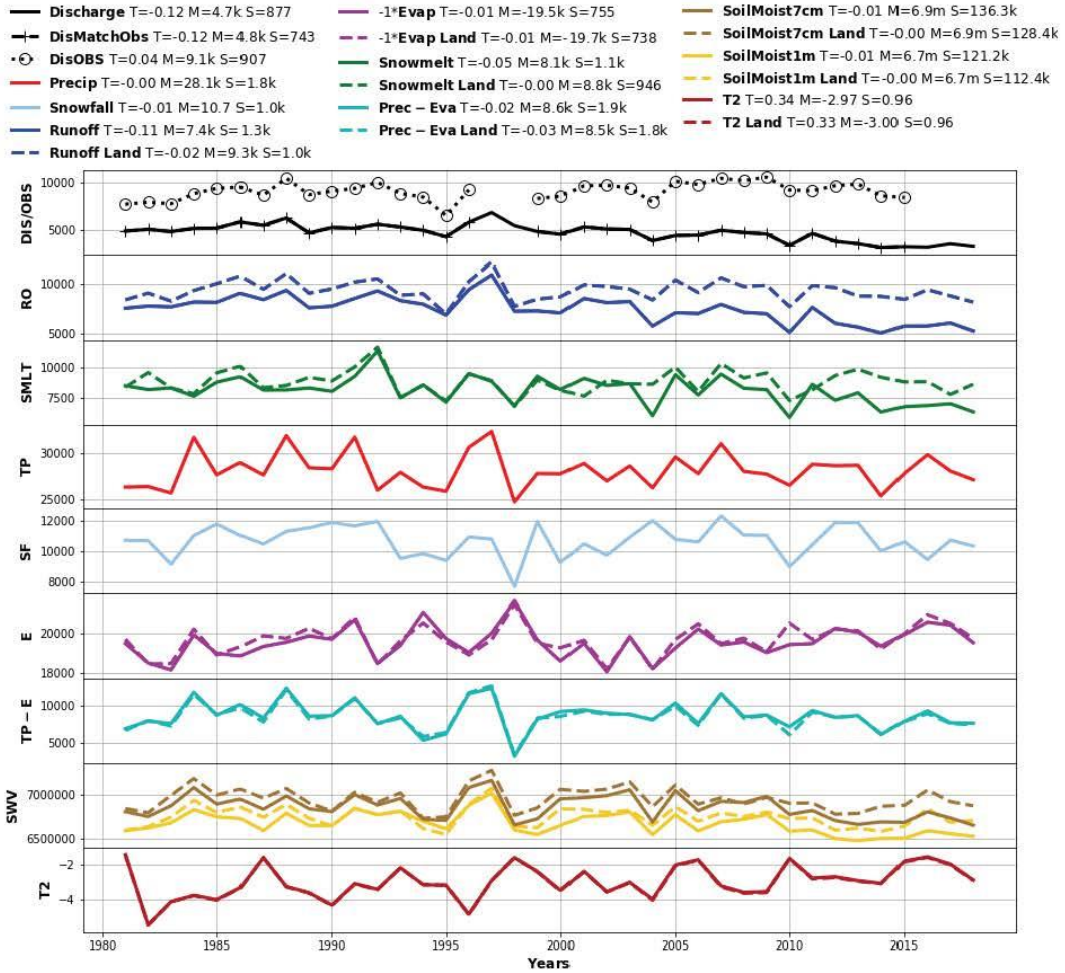
River: Yukon, Outlet: [61.85N; -162.85W], Area: 865451 km2

- Discharge T=-0.14 M=3.6k S=781
- DisMatchObs T=-0.20 M=3.6k S=819
- DisOBS T=0.02 M=6.7k S=534
- Precip T=-0.01 M=14.6k S=894
- Snowfall T=-0.03 M=5.8 S=889
- Runoff T=-0.13 M=4.8k S=956
- Runoff Land T=-0.03 M=6.7k S=857
- -1*Evap T=-0.00 M=-7.8k S=373
- -1*Evap Land T=-0.01 M=-7.8k S=401
- Snowmelt T=-0.13 M=4.2k S=933
- Snowmelt Land T=-0.03 M=5.0k S=750
- Prec - Eva T=-0.01 M=6.8k S=905
- Prec - Eva Land T=-0.03 M=6.8k S=979
- SoilMoist7cm T=-0.01 M=2.9m S=63.4k
- SoilMoist7cm Land T=-0.00 M=3.0m S=64.7k
- SoilMoist1m T=-0.01 M=2.8m S=70.6k
- SoilMoist1m Land T=-0.00 M=2.9m S=56.2k
- T2 T=0.40 M=-4.95 S=0.98
- T2 Land T=0.39 M=-5.02 S=0.97



North America - No.2

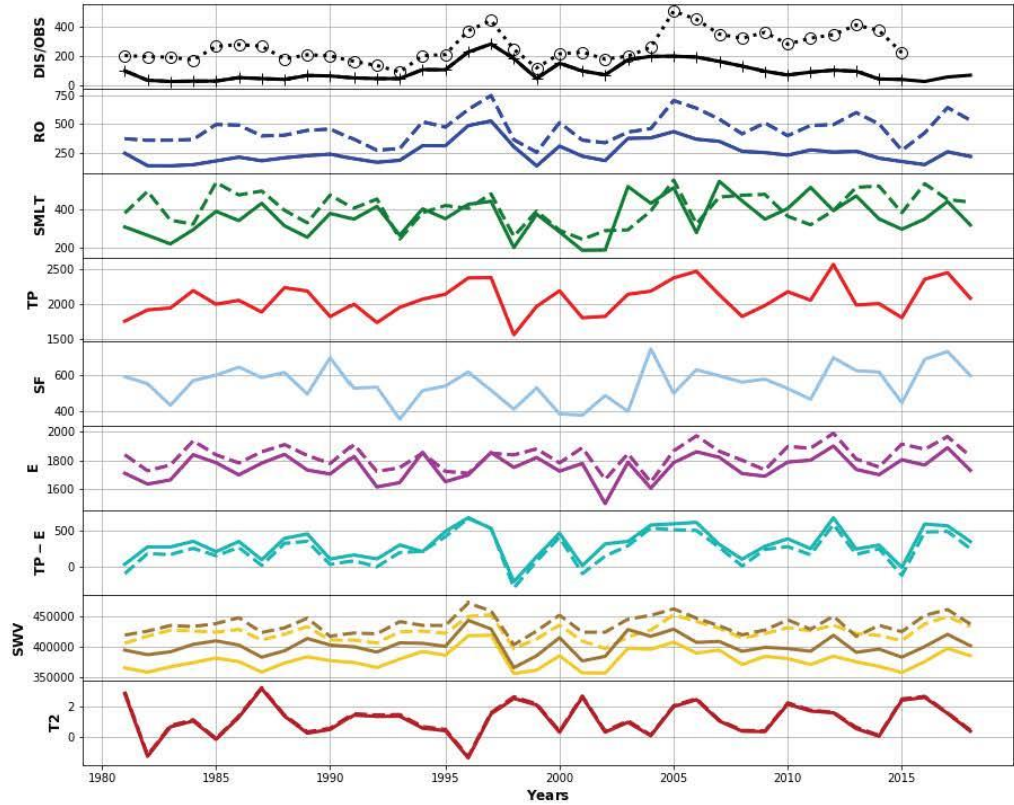
River: Mackenzie, Outlet: [67.45N; -133.65W], Area: 1726490 km2



North America - No.3

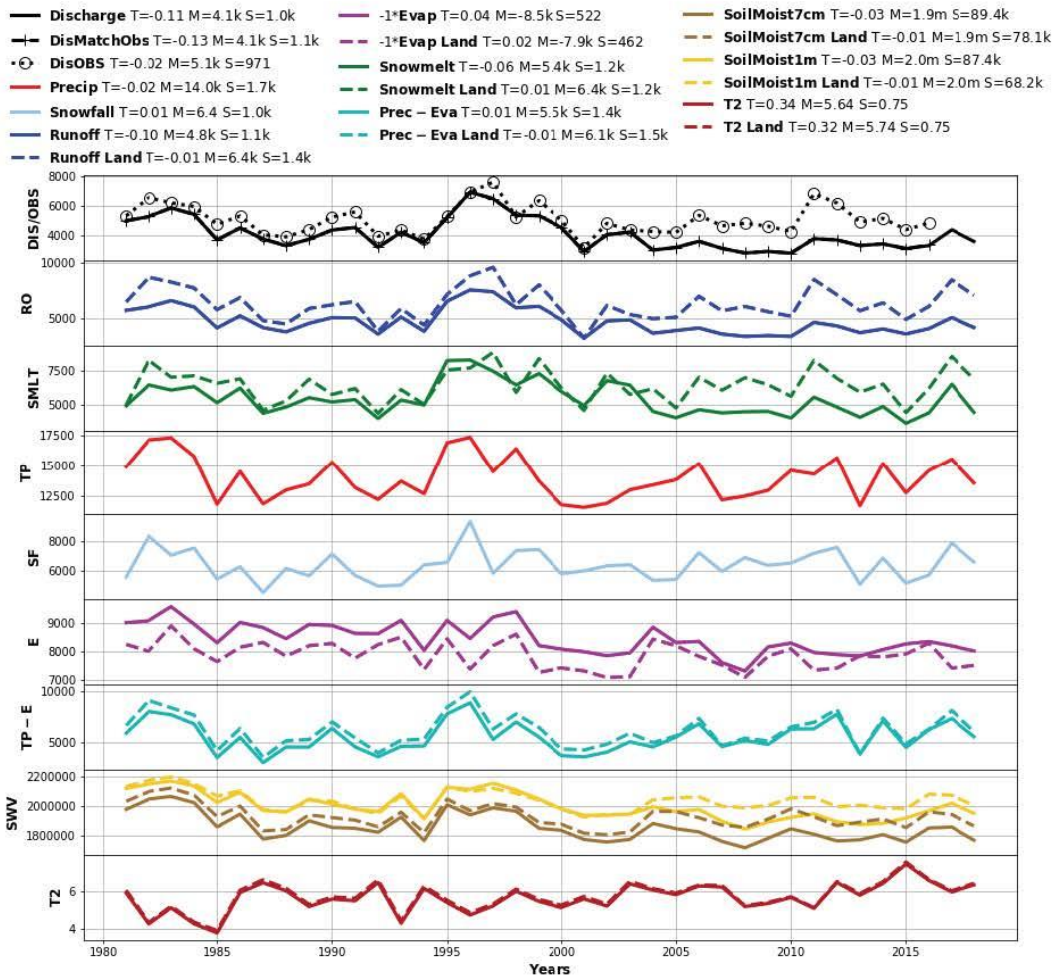
River: Churchill, Outlet: [55.65N; -104.75W], Area: 120999 km2

- Discharge T=0.12 M=95 S=64
- DisMatchObs T=0.22 M=99 S=65
- DisOBS T=0.20 M=262 S=98
- Precip T=0.03 M=2.1k S=227
- Snowfall T=0.03 M=554 S=97
- Runoff T=0.07 M=259 S=93
- Runoff Land T=0.07 M=457 S=115
- -1*Evap T=-0.01 M=-1.7k S=85
- -1*Evap Land T=-0.01 M=-1.8k S=81
- Snowmelt T=0.07 M=360 S=93
- Snowmelt Land T=0.02 M=406 S=86
- Prec - Eva T=0.15 M=315 S=200
- Prec - Eva Land T=0.19 M=237 S=211
- SoilMoist7cm T=0.00 M=401.8k S=15.8k
- SoilMoist7cm Land T=0.01 M=436.1k S=14.7k
- SoilMoist1m T=0.01 M=379.0k S=15.9k
- SoilMoist1m Land T=0.01 M=424.2k S=13.9k
- T2 T=0.13 M=1.14 S=1.09
- T2 Land T=0.12 M=1.22 S=1.09



North America - No.4

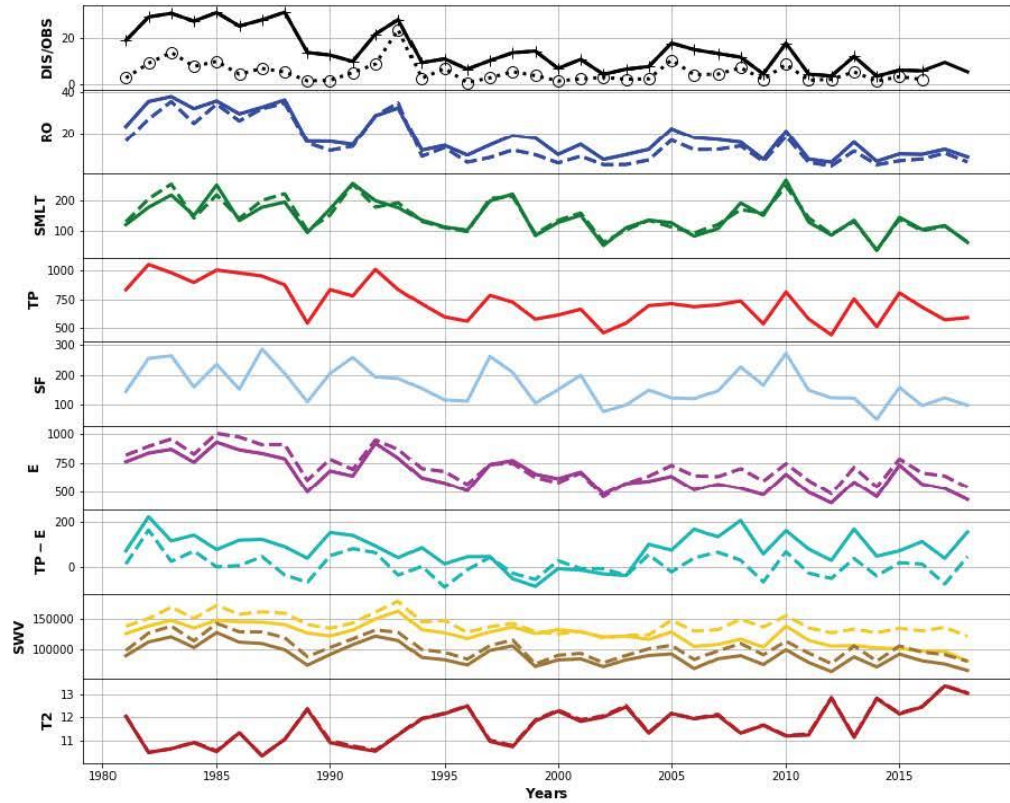
River: Columbia, Outlet: [45.65N; -121.25W], Area: 612077 km2



North America - No.5

River: Colorado, Outlet: [36.15N; -111.75W], Area: 68780 km2

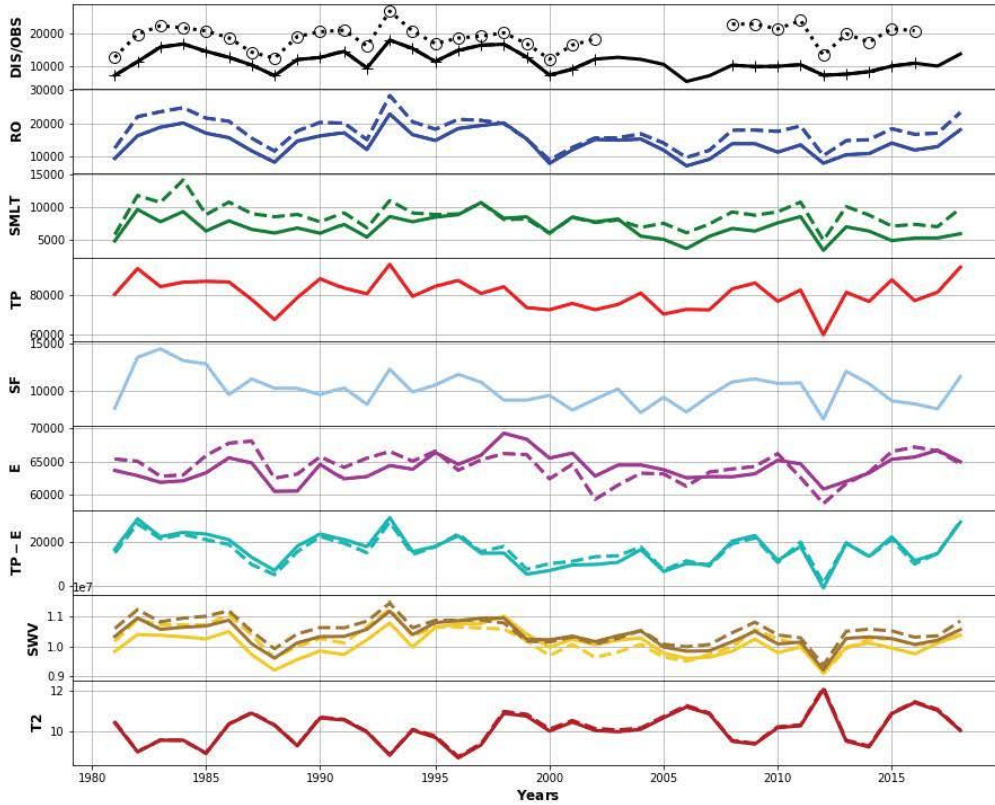
- Discharge T=-0.42 M=14 S= 8
- DisMatchObs T=-0.43 M=14 S= 8
- DisOBS T=-0.27 M= 5 S= 4
- Precip T=-0.13 M=724 S=165
- Snowfall T=-0.16 M=165 S=58
- Runoff T=-0.36 M=17 S= 9
- Runoff Land T=-0.43 M=14 S= 9
- -1*Evap T=0.14 M=-645 S=141
- -1*Evap Land T=0.12 M=-714 S=140
- Snowmelt T=-0.15 M=144 S=54
- Snowmelt Land T=-0.16 M=146 S=53
- Prec - Eva T=-0.05 M=79 S=69
- Prec - Eva Land T=-1.05 M=10 S=50
- SoilMoist7cm T=-0.11 M=89.8k S=16.7k
- SoilMoist7cm Land T=-0.09 M=102.7k S=17.9k
- SoilMoist1m T=-0.10 M=123.5k S=17.7k
- SoilMoist1m Land T=-0.06 M=140.8k S=14.9k
- T2 T=0.43 M=11.64 S=0.78
- T2 Land T=0.44 M=11.67 S=0.78



North America - No.6

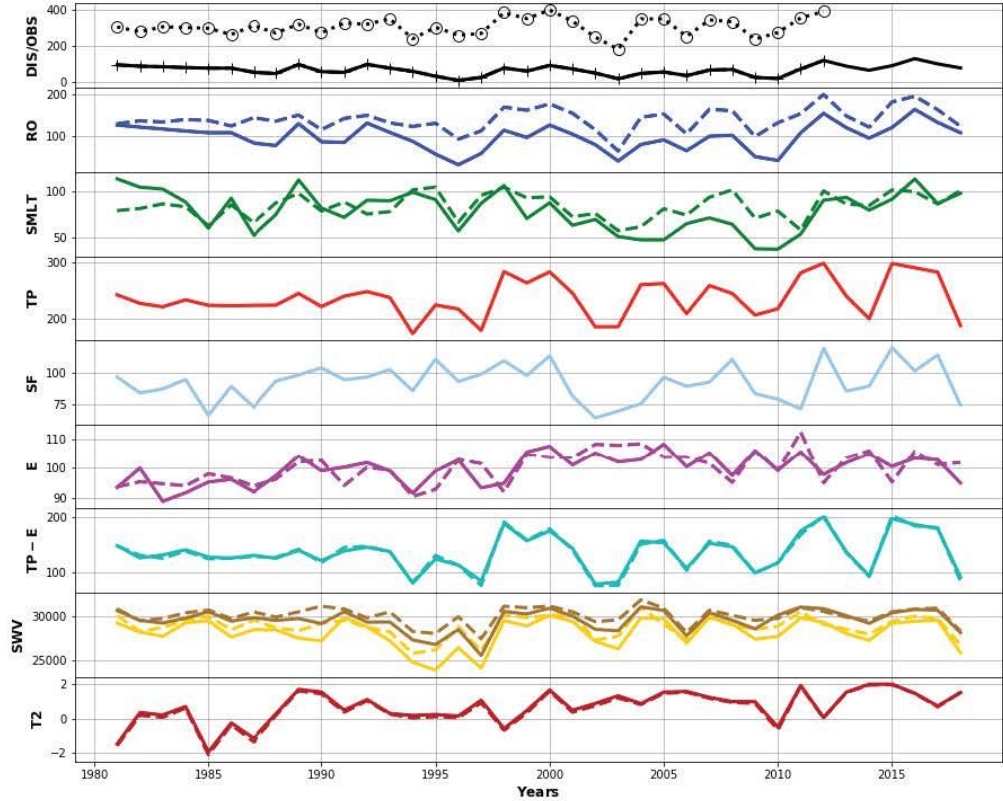
River: Mississippi, Outlet: [32.15N; -91.05W], Area: 2963330 km2

- **Discharge** T=-0.09 M=11.4k S=3.1k
- **DisMatchObs** T=-0.11 M=11.7k S=3.1k
- **DisOBS** T=0.03 M=18.9k S=3.4k
- **Precip** T=-0.02 M=80.6k S=7.4k
- **Snowfall** T=-0.06 M=10.2 S=1.7k
- **Runoff** T=-0.08 M=14.2k S=3.7k
- **Runoff Land** T=-0.07 M=17.5k S=4.2k
- **-1*Evap** T=-0.01 M=-64.1k S=1.9k
- **-1*Evap Land** T=0.01 M=-64.3k S=2.2k
- **Snowmelt** T=-0.08 M=6.8k S=1.7k
- **Snowmelt Land** T=-0.06 M=8.6k S=1.8k
- **Prec - Eva** T=-0.12 M=16.5k S=7.3k
- **Prec - Eva Land** T=-0.07 M=16.3k S=6.5k
- **SoilMoist7cm** T=-0.01 M=10.4m S=395.4k
- **SoilMoist7cm Land** T=-0.02 M=10.5m S=407.4
- **SoilMoist1m** T=-0.01 M=10.1m S=417.2k
- **SoilMoist1m Land** T=-0.02 M=10.2m S=458.5k
- **T2** T=0.24 M=10.11 S=0.77
- **T2 Land** T=0.25 M=10.16 S=0.78



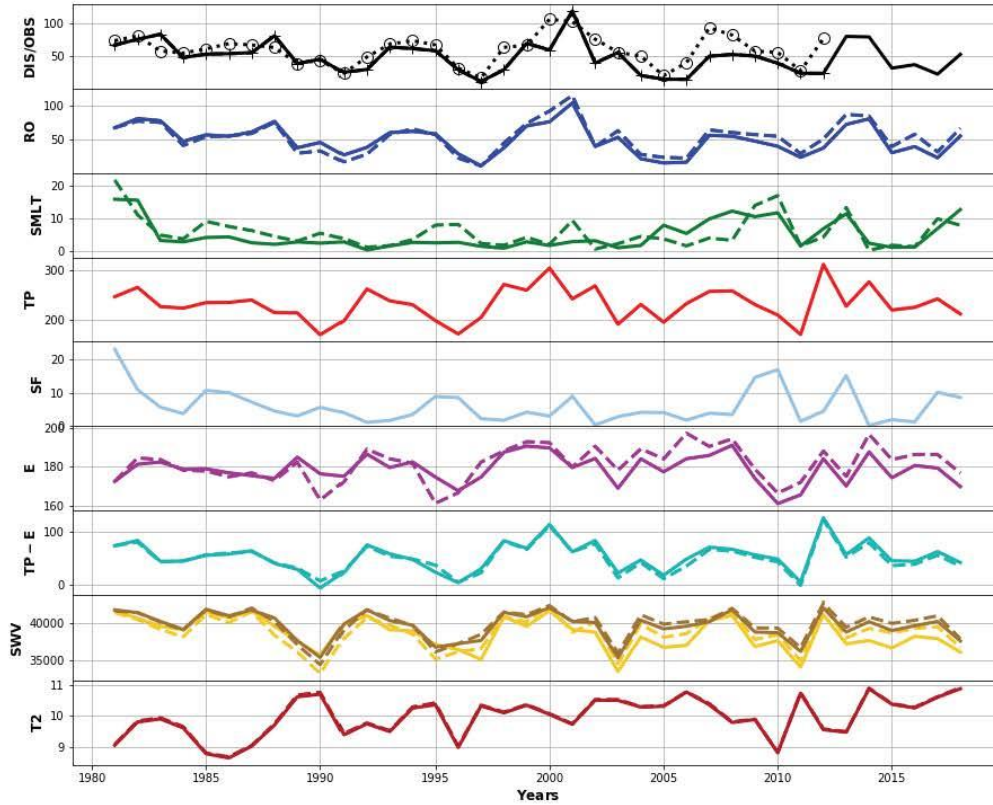
Europe - No.1

River: Kalixaelven, Outlet: [66.15N; 22.85E], Area: 10996 km2



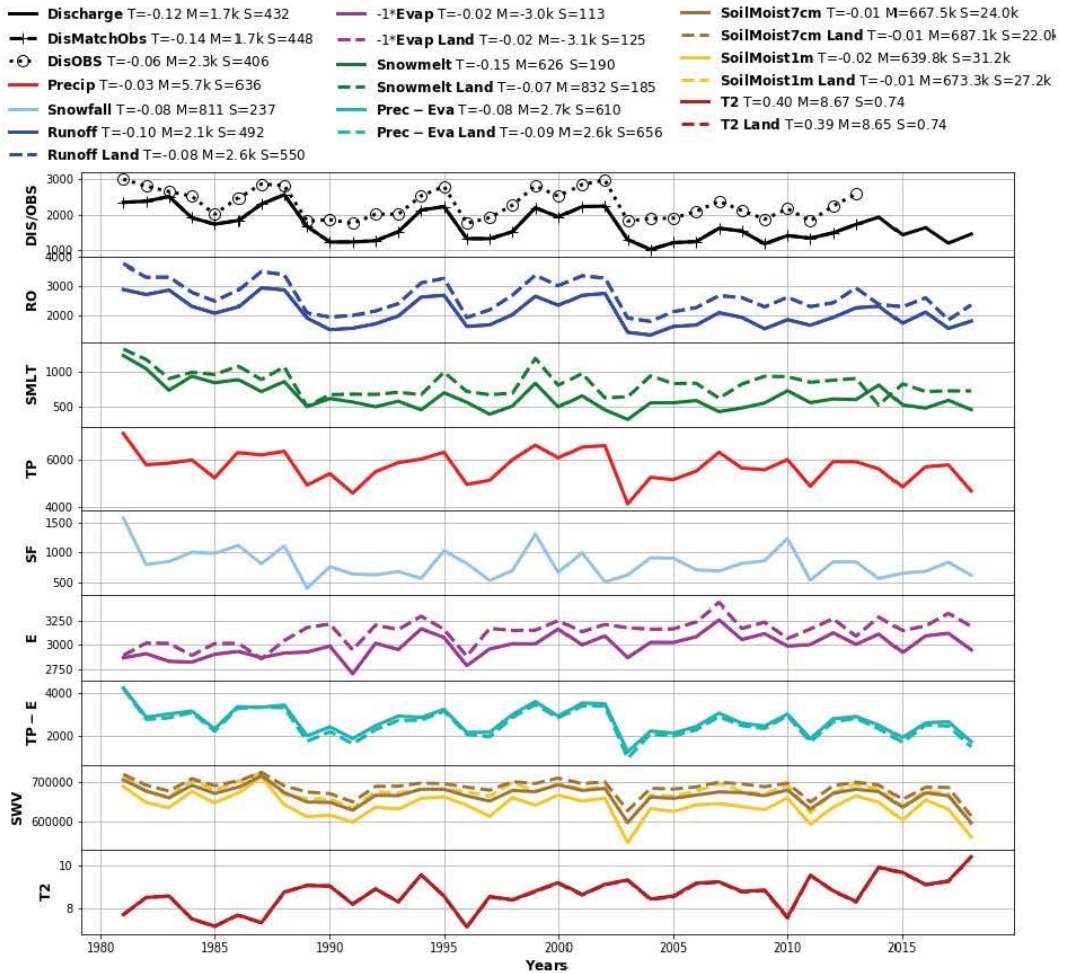
Europe - No.2

River: Thames, Outlet: [51.35N; -0.45W], Area: 9926 km2



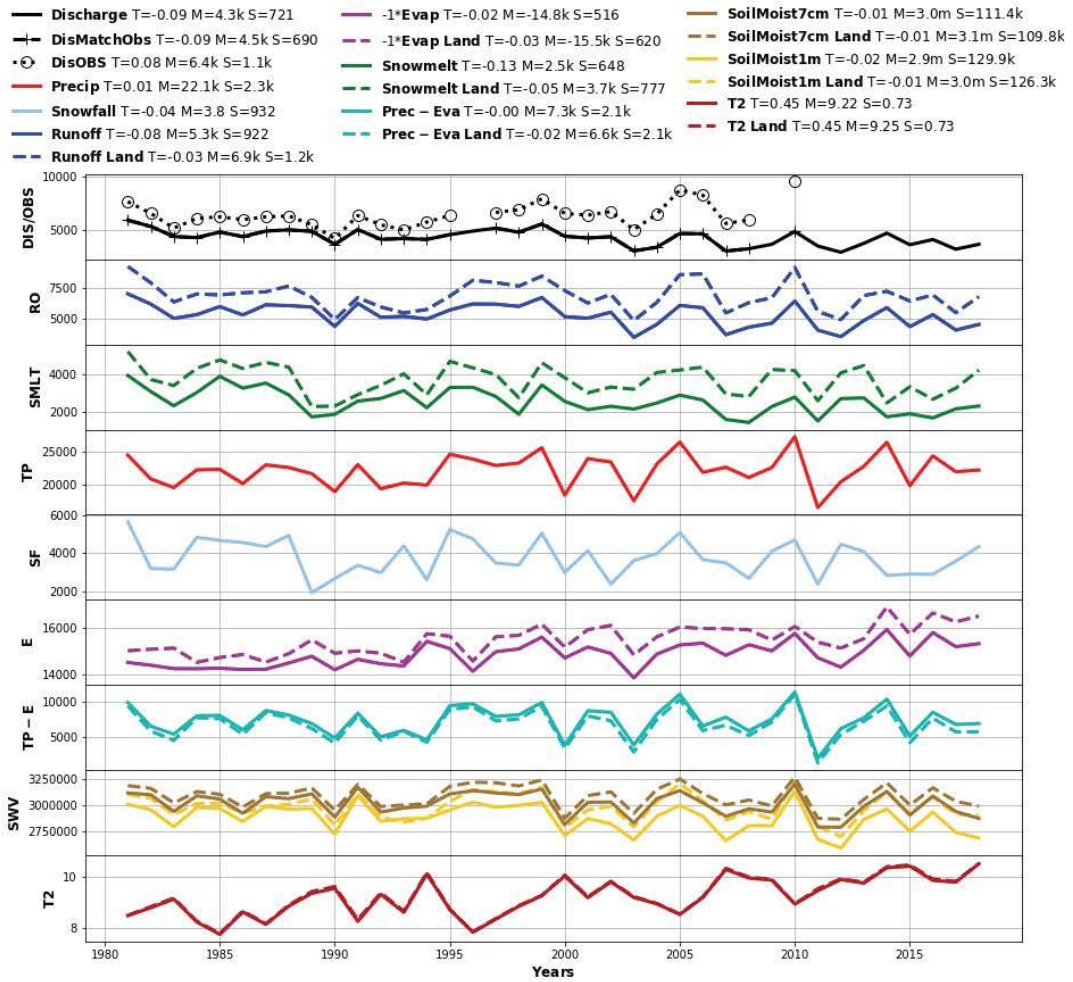
Europe - No.3

River: Rhine, Outlet: [51.85N; 6.15E], Area: 163315 km2



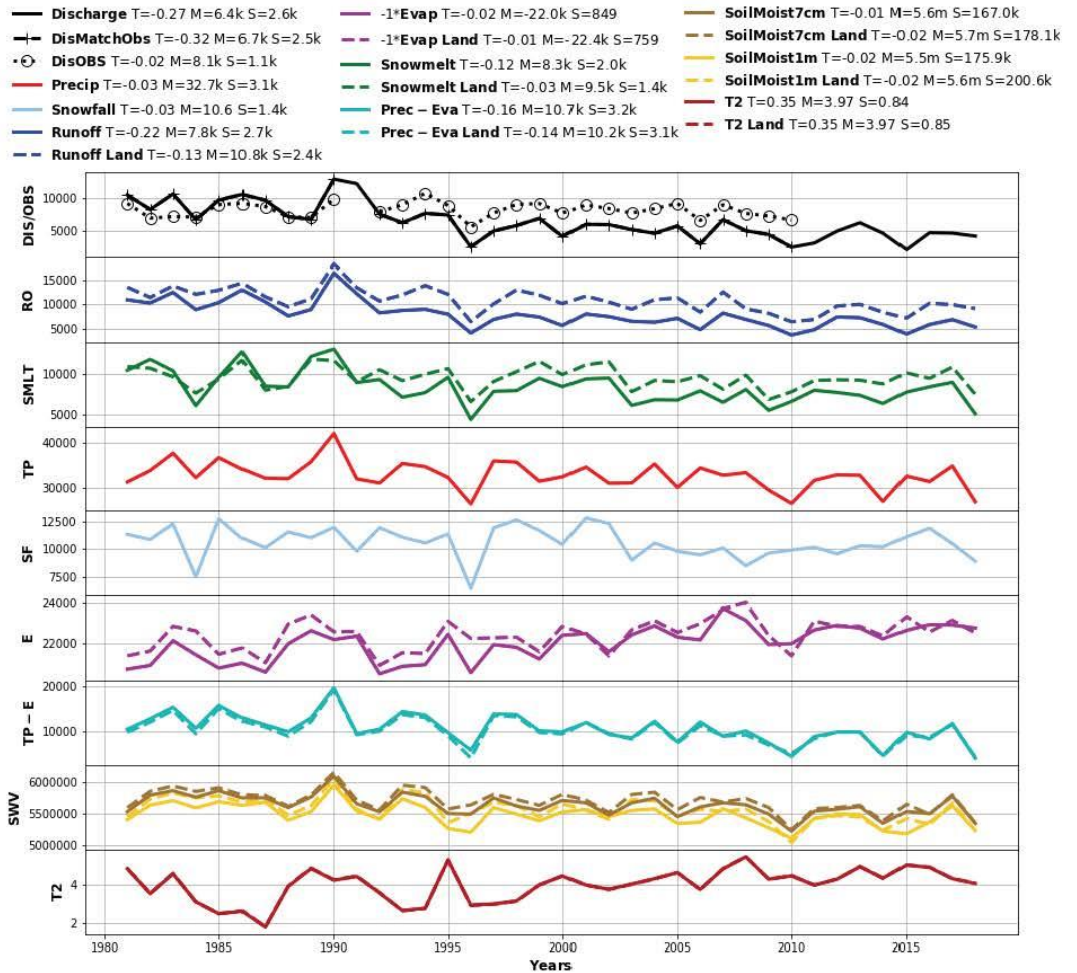
Europe - No.4

River: Danube, Outlet: [45.25N; 28.75E], Area: 788003 km2



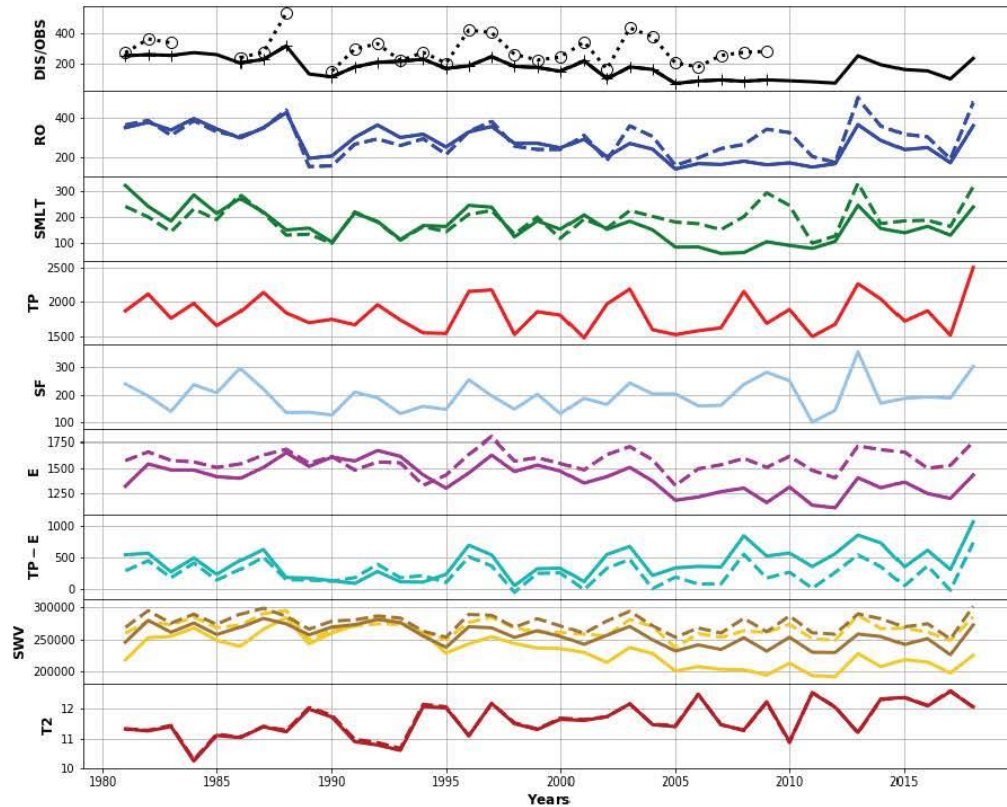
Europe - No.5

River: Volga, Outlet: [48.85N; 44.65E], Area: 1445620 km2



Europe - No.6

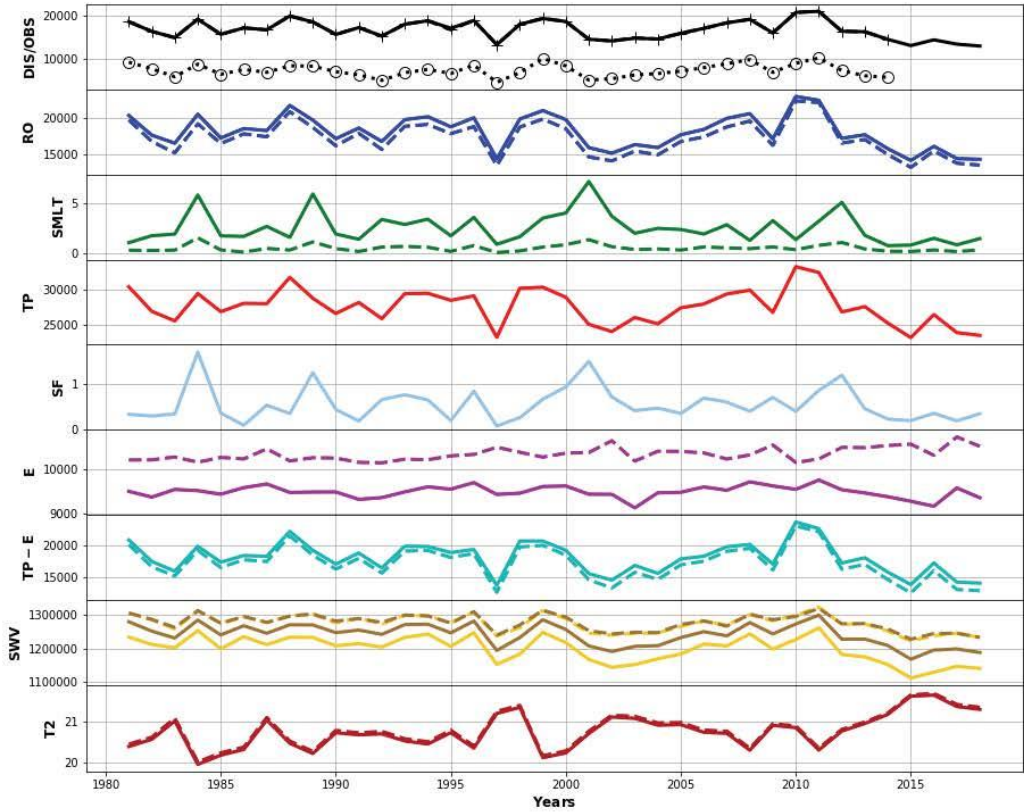
River: Ebro, Outlet: [40.85N; 0.55E], Area: 85737 km2



South America - No.1

River: Magdalena, Outlet: [10.15N; -74.95W], Area: 259487 km2

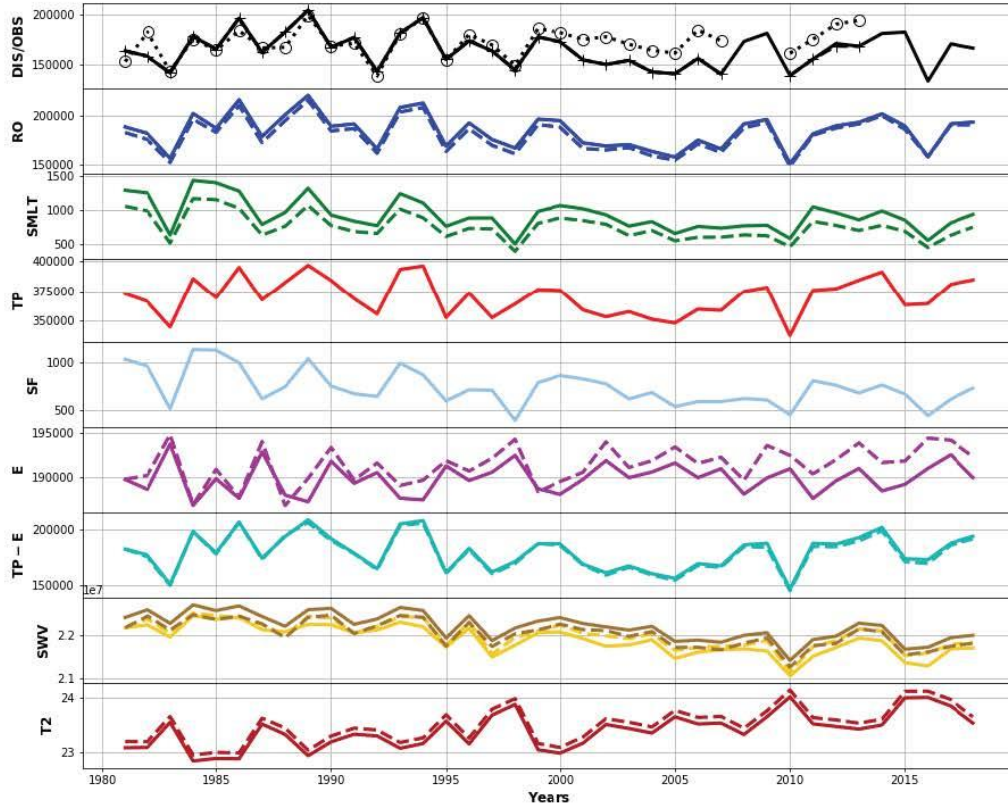
- Discharge T=-0.03 M=16.7k S=2.2k
- + DisMatchObs T=-0.00 M=17.1k S=2.0k
- DisOBS T=0.01 M=7.4k S=1.4k
- Precip T=-0.02 M=27.6k S=2.5k
- Snowfall T=nan M= 0 S= 0
- Runoff T=-0.04 M=18.1k S=2.3k
- Runoff Land T=-0.03 M=17.2k S=2.4k
- 1*Evap T=0.00 M=-9.5k S=134
- 1*Evap Land T=-0.01 M=-10.3k S=147
- Snowmelt T=-0.08 M= 2 S= 1
- Snowmelt Land T=-0.04 M= 0 S= 0
- Prec - Eva T=-0.03 M=18.1k S=2.4k
- Prec - Eva Land T=-0.04 M=17.2k S=2.5k
- SoilMoist7cm T=-0.01 M=1.2m S=32.2k
- SoilMoist7cm Land T=-0.01 M=1.3m S=24.1k
- SoilMoist1m T=-0.02 M=1.2m S=37.9k
- SoilMoist1m Land T=-0.01 M=1.3m S=25.5k
- T2 T=0.21 M=20.76 S=0.42
- T2 Land T=0.21 M=20.81 S=0.41



South America - No.2

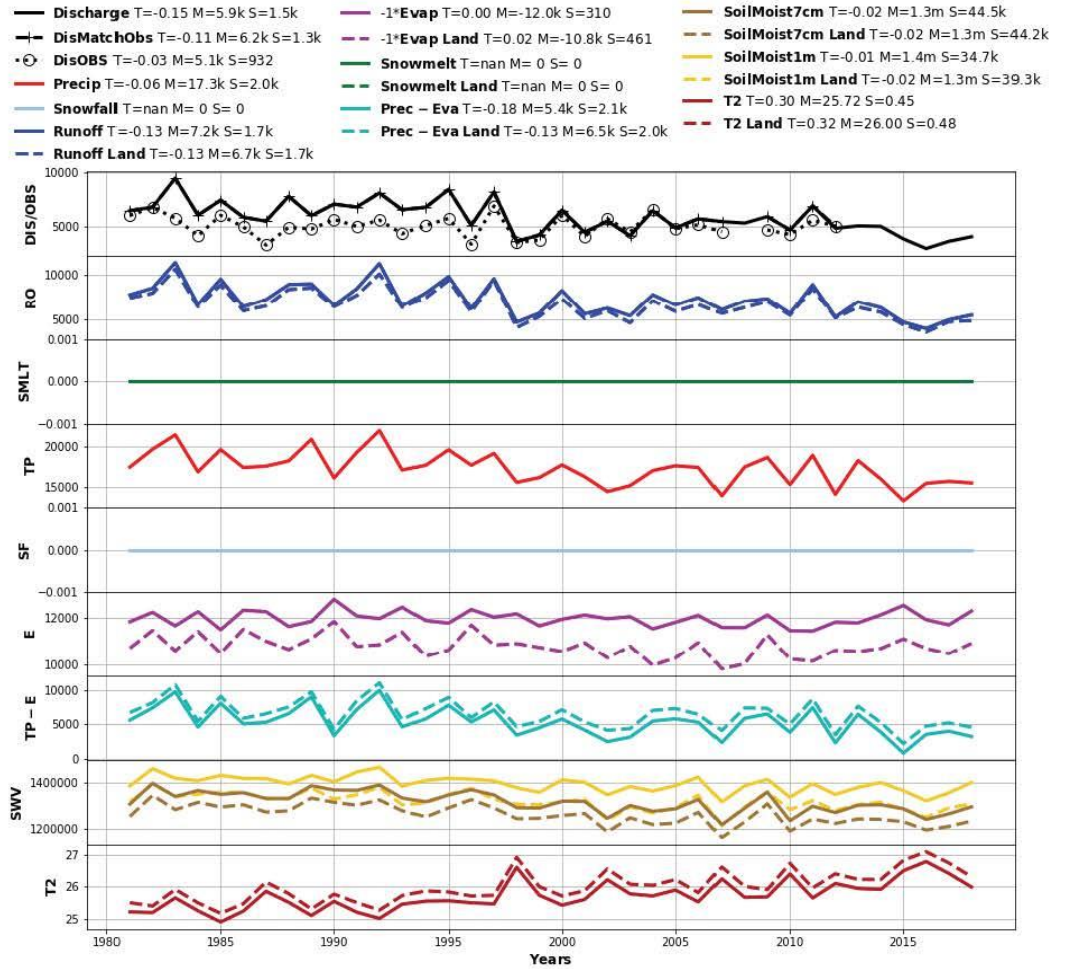
River: Amazon, Outlet: [-1.95S; -55.55W], Area: 4664200 km2

- Discharge T=-0.02 M=165.1k S=17.1k
- DisMatchObs T=-0.04 M=163.9k S=17.2k
- DisOBS T=0.02 M=172.4k S=14.5k
- Precip T=-0.00 M=370.3k S=15.0k
- Snowfall T=-0.12 M=737 S=184
- Runoff T=-0.02 M=184.7k S=17.1k
- Runoff Land T=-0.01 M=180.5k S=16.8k
- -1*Evap T=-0.00 M=-189.9k S=1.7k
- -1*Evap Land T=-0.00 M=-191.4k S=2.1k
- Snowmelt T=-0.12 M=920 S=232
- Snowmelt Land T=-0.12 M=746 S=189
- Prec - Eva T=-0.01 M=180.4k S=16.1k
- Prec - Eva Land T=-0.01 M=178.9k S=16.0k
- SoilMoist7cm T=-0.01 M=22.2m S=317.9k
- SoilMoist7cm Land T=-0.01 M=22.0m S=306.6
- SoilMoist1m T=-0.01 M=21.9m S=336.0k
- SoilMoist1m Land T=-0.01 M=22.0m S=317.8k
- T2 T=0.19 M=23.39 S=0.32
- T2 Land T=0.19 M=23.50 S=0.32



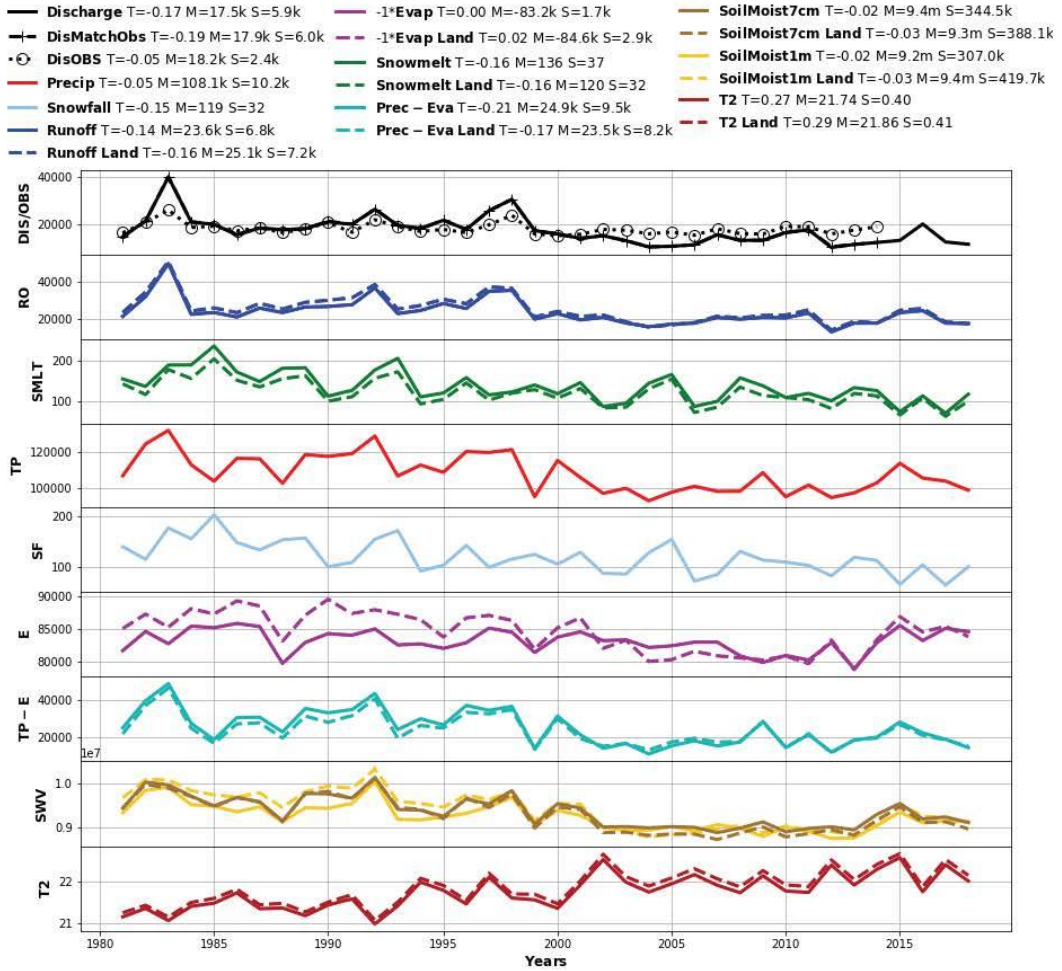
South America - No.3

River: Araguaia, Outlet: [-8.35S; -49.25W], Area: 319643 km2



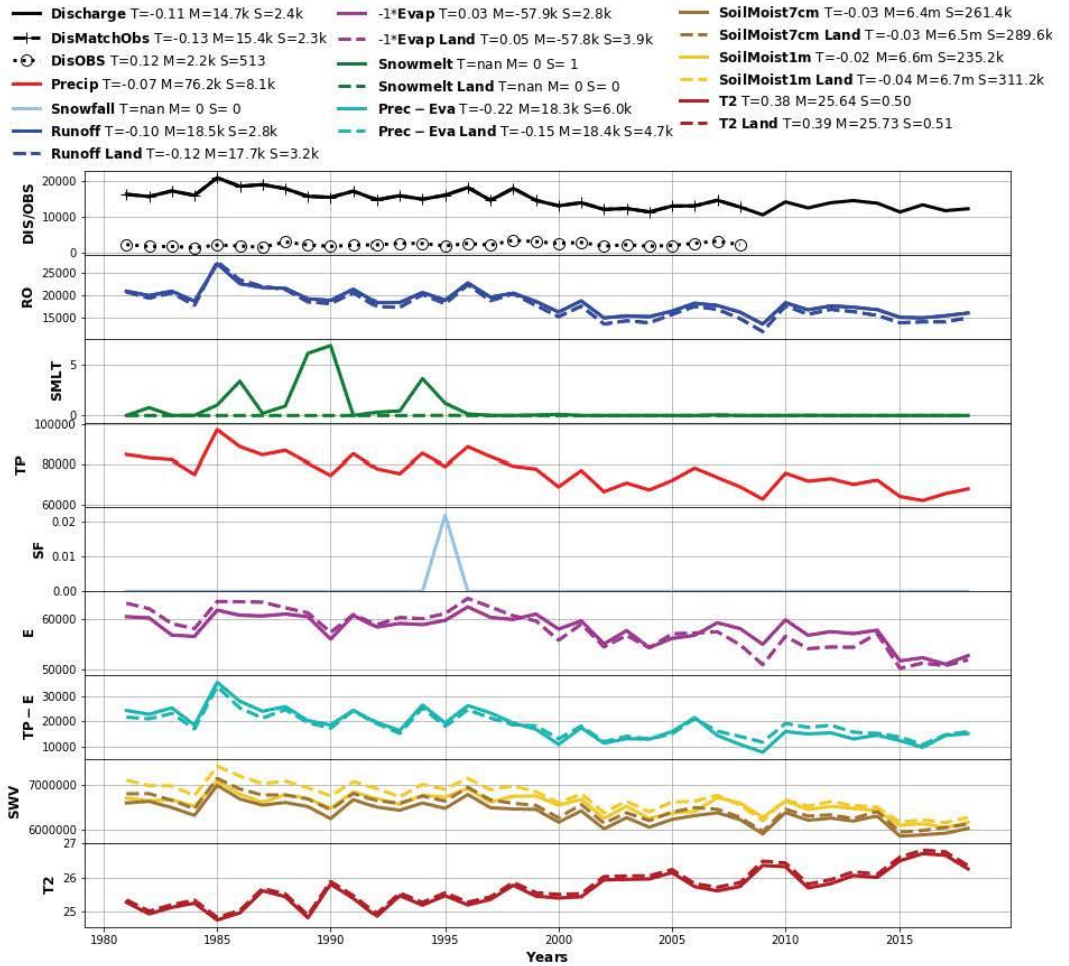
South America - No.4

River: Parana, Outlet: [-32.75S; -60.75W], Area: 2430030 km2



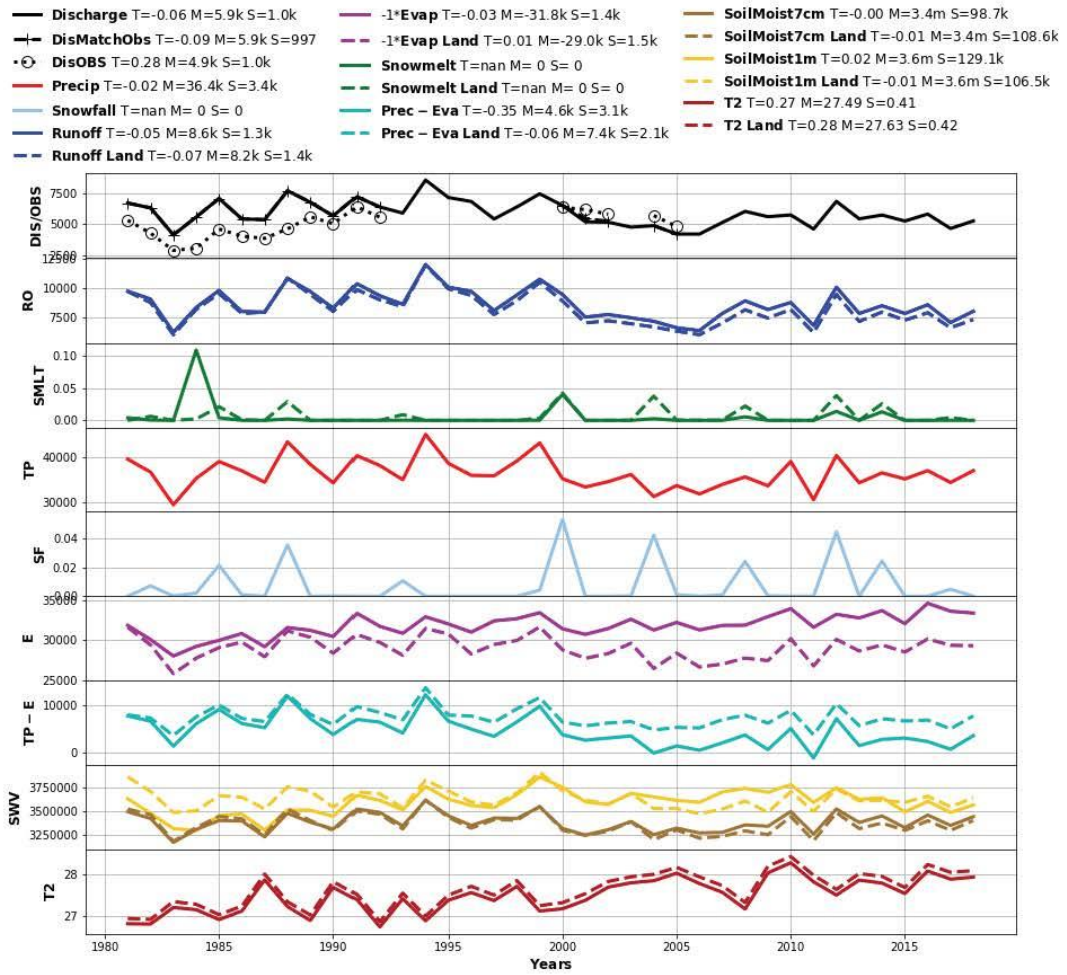
Africa - No.1

River: Nile, Outlet: [19.15N; 30.45E], Area: 2586140 km2



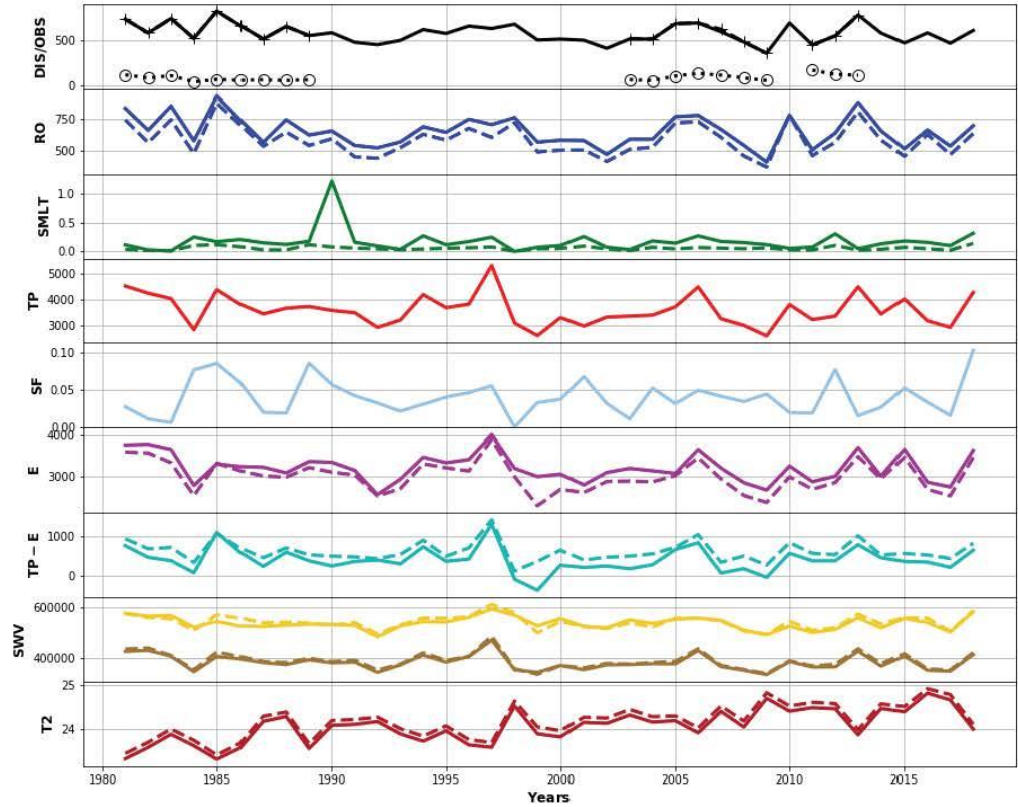
Africa - No.2

River: Niger, Outlet: [7.75N; 6.75E], Area: 2094830 km2



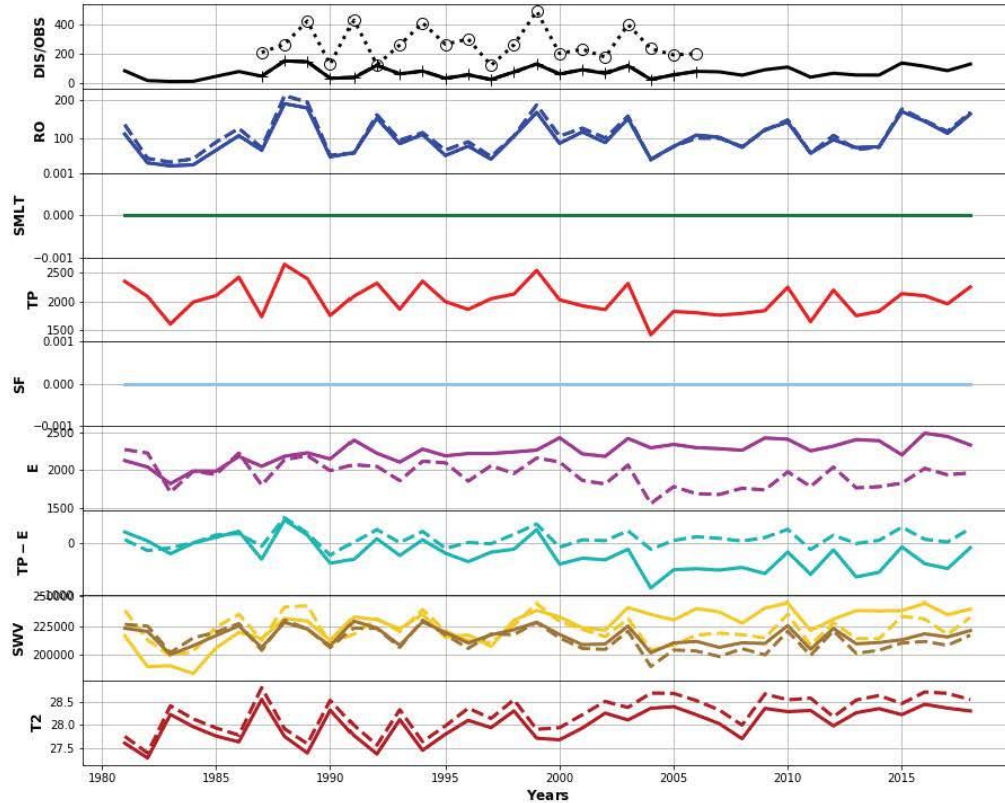
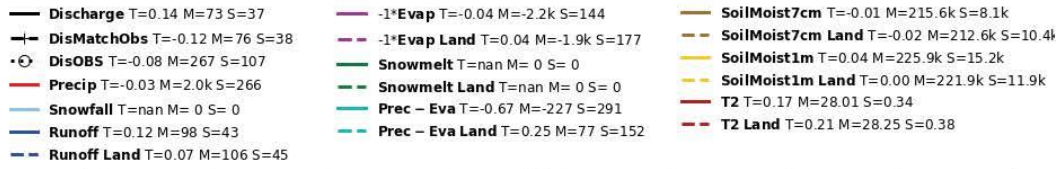
Africa - No.3

River: Shabelle, Outlet: [3.85N; 45.55E], Area: 210122 km2



Africa - No.4

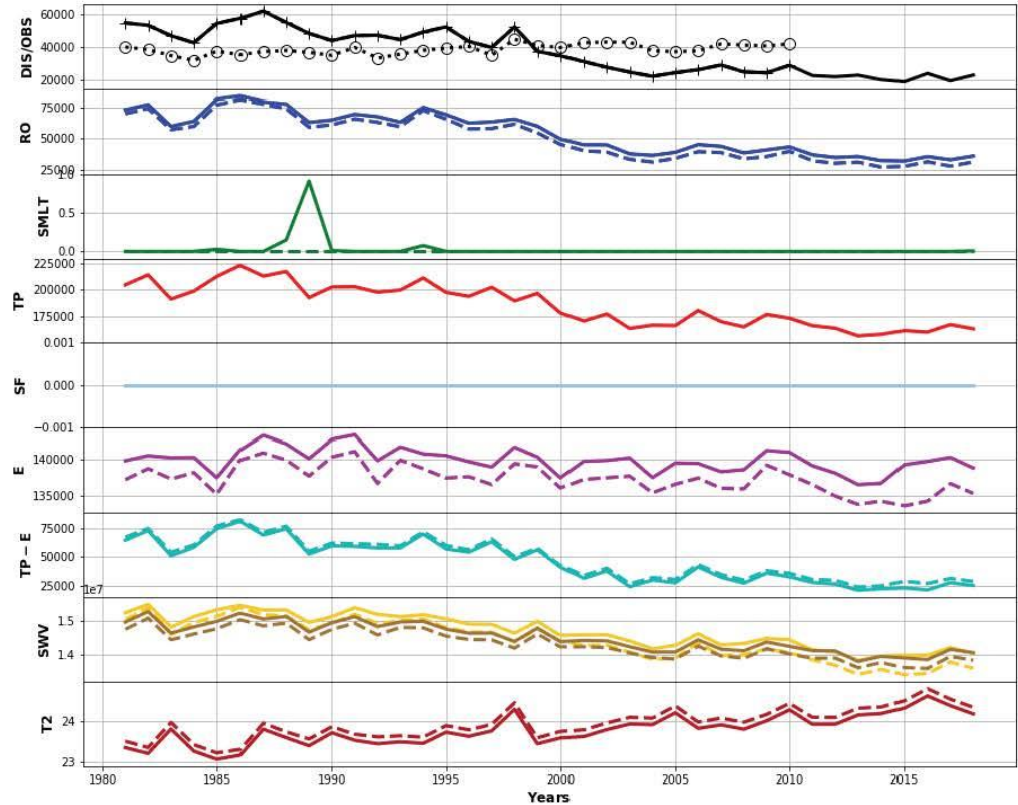
River: White-Volta, Outlet: [9.75N; -0.95W], Area: 93343 km2



Africa - No.5

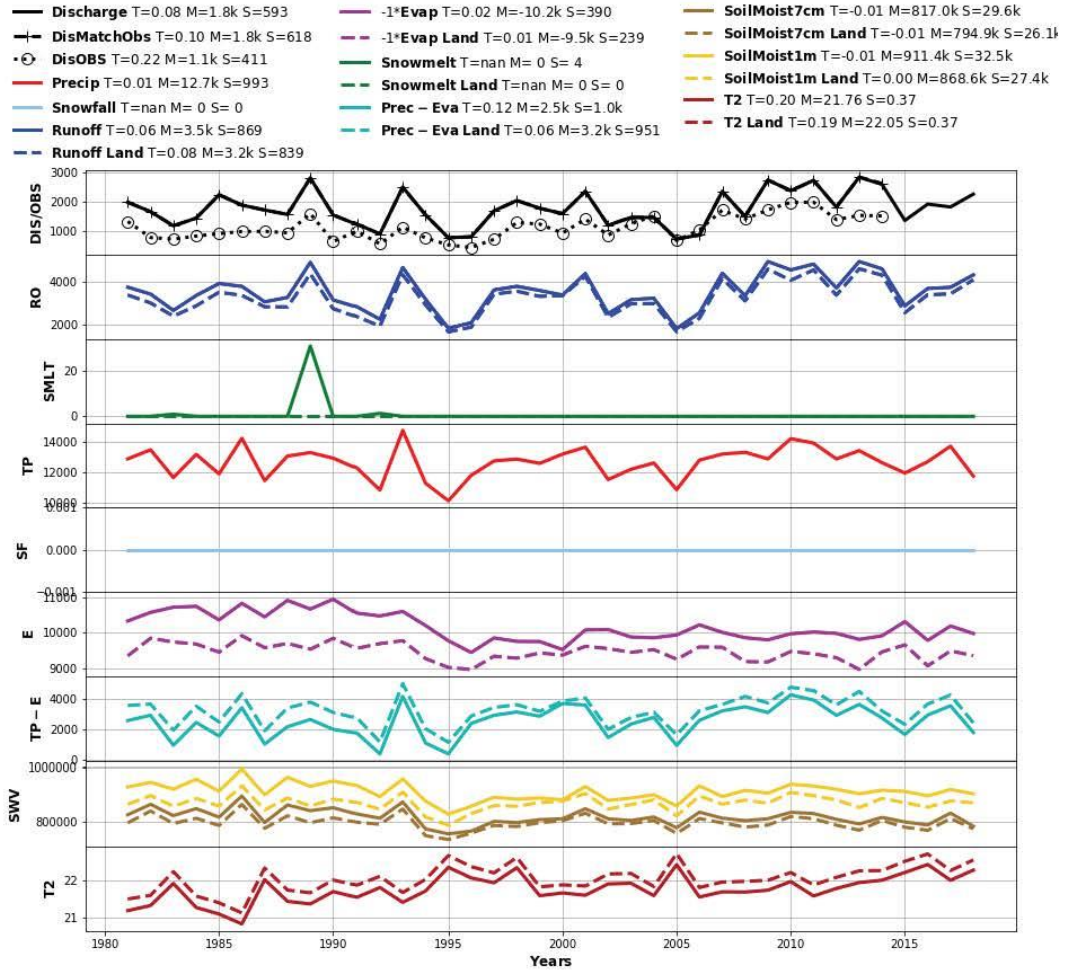
River: Congo, Outlet: [-4.05S; 15.65E], Area: 3618130 km2

- Discharge T=-0.30 M=36.8k S=13.4k
- DisMatchObs T=-0.29 M=41.0k S=12.0k
- DisOBS T=0.05 M=38.6k S=3.1k
- Precip T=-0.08 M=185.7k S=19.3k
- Snowfall T=nan M= 0 S= 0
- Runoff T=-0.25 M=54.1k S=16.4k
- Runoff Land T=-0.28 M=49.8k S=17.0k
- -1*Evap T=0.01 M=-140.0k S=1.7k
- -1*Evap Land T=0.01 M=-137.3k S=2.0k
- Snowmelt T=nan M= 0 S= 0
- Snowmelt Land T=nan M= 0 S= 0
- Prec - Eva T=-0.33 M=45.8k S=18.3k
- Prec - Eva Land T=-0.30 M=48.4k S=17.9k
- SoilMoist7cm T=-0.02 M=14.5m S=411.1k
- SoilMoist7cm Land T=-0.02 M=14.3m S=418.2
- SoilMoist1m T=-0.03 M=14.7m S=481.2k
- SoilMoist1m Land T=-0.03 M=14.4m S=583.0k
- T2 T=0.28 M=23.79 S=0.37
- T2 Land T=0.28 M=23.95 S=0.37



Africa - No.6

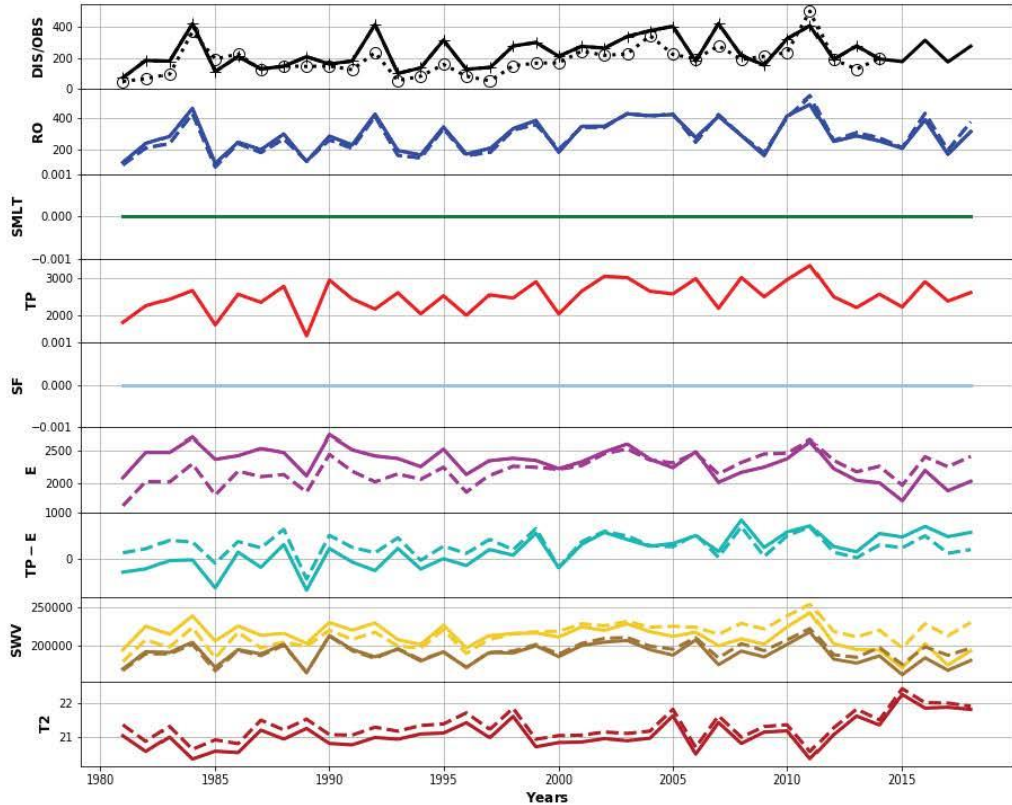
River: Zambesi-upstream, Outlet: [-17.45S; 24.35E], Area: 331113 km2



Africa - No.7

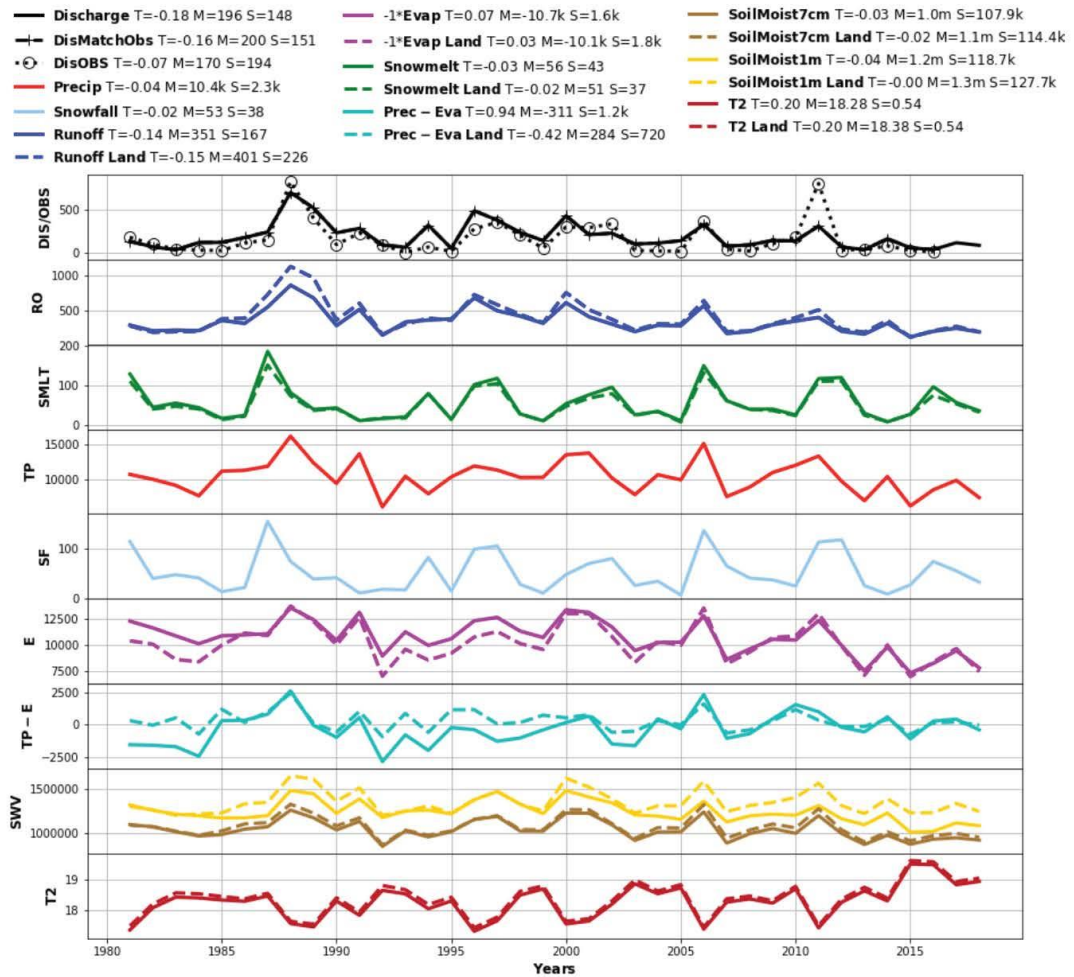
River: Cunene, Outlet: [-17.45S; 14.15E], Area: 85253 km2

- Discharge T=0.12 M=237 S=99
- † DisMatchObs T=0.16 M=237 S=102
- DisOBS T=0.21 M=183 S=93
- Precip T=0.05 M=2.5k S=399
- Snowfall T=nan M= 0 S= 0
- Runoff T=0.09 M=284 S=104
- Runoff Land T=0.15 M=280 S=110
- -1*Evap T=0.05 M=-2.3k S=216
- -1*Evap Land T=-0.05 M=-2.2k S=205
- Snowmelt T=nan M= 0 S= 0
- Snowmelt Land T=nan M= 0 S= 0
- Prec - Eva T=1.25 M=190 S=357
- Prec - Eva Land T=0.09 M=288 S=248
- SoilMoist7cm T=-0.01 M=189.8k S=13.2k
- SoilMoist7cm Land T=0.02 M=193.5k S=12.8k
- SoilMoist1m T=-0.03 M=212.1k S=15.4k
- SoilMoist1m Land T=0.04 M=215.3k S=15.1k
- T2 T=0.22 M=21.09 S=0.44
- T2 Land T=0.18 M=21.32 S=0.42



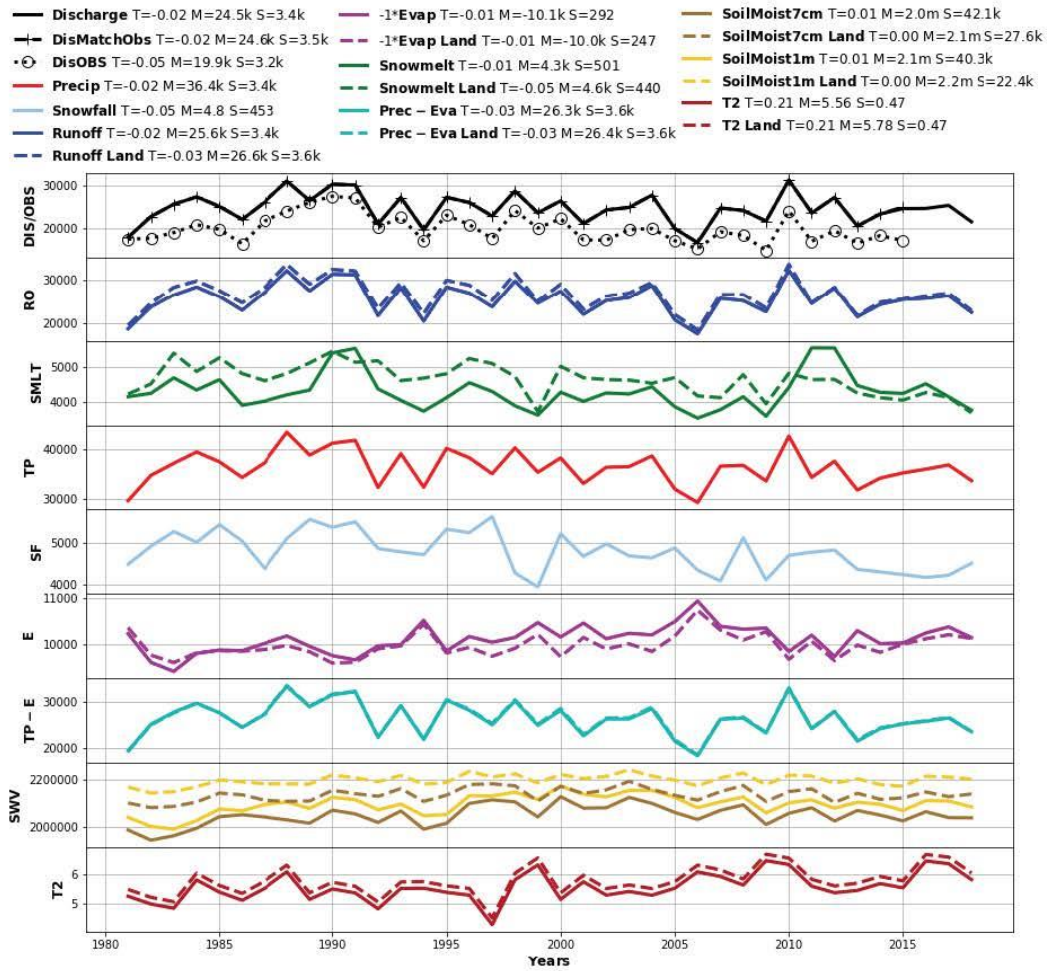
Africa - No.8

River: Orange, Outlet: [-28.75S; 17.65E], Area: 734698 km2



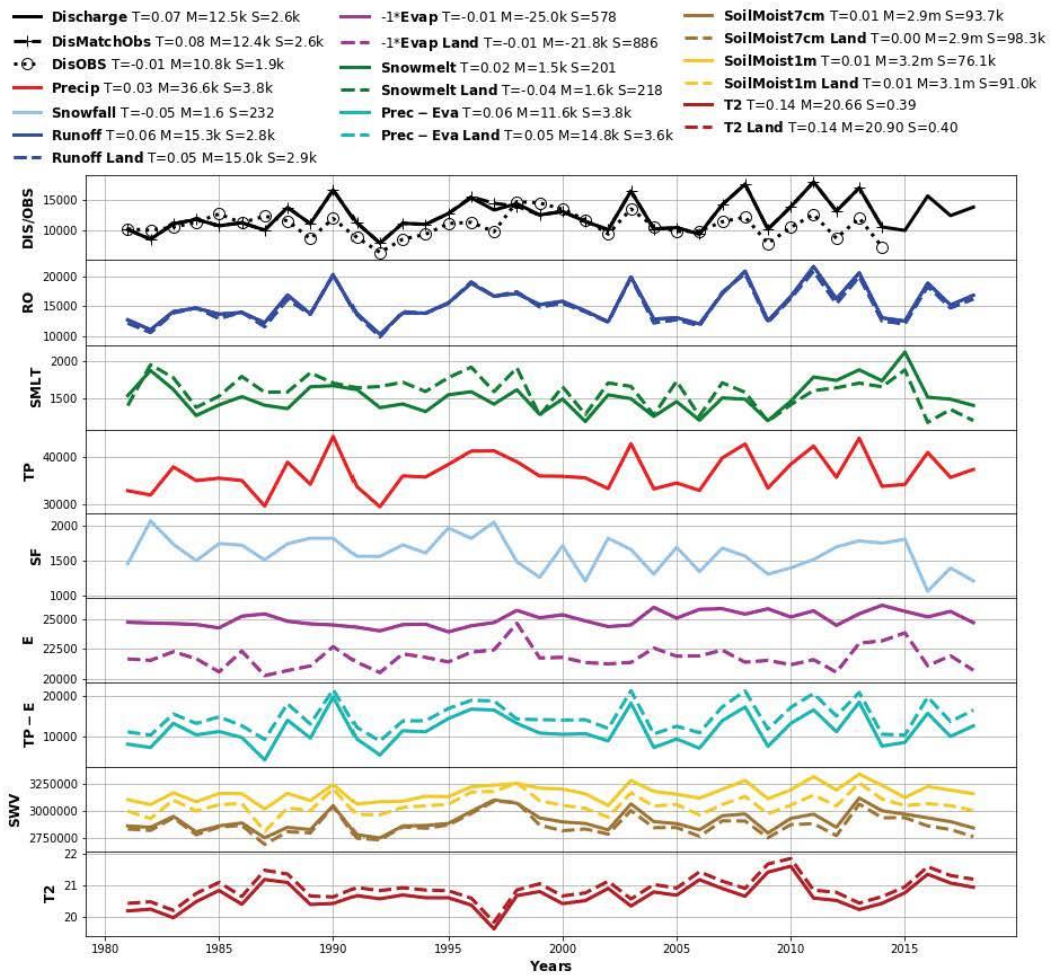
South Asia - No.1

River: Brahmaputra, Outlet: [25.15N; 89.65E], Area: 518898 km2



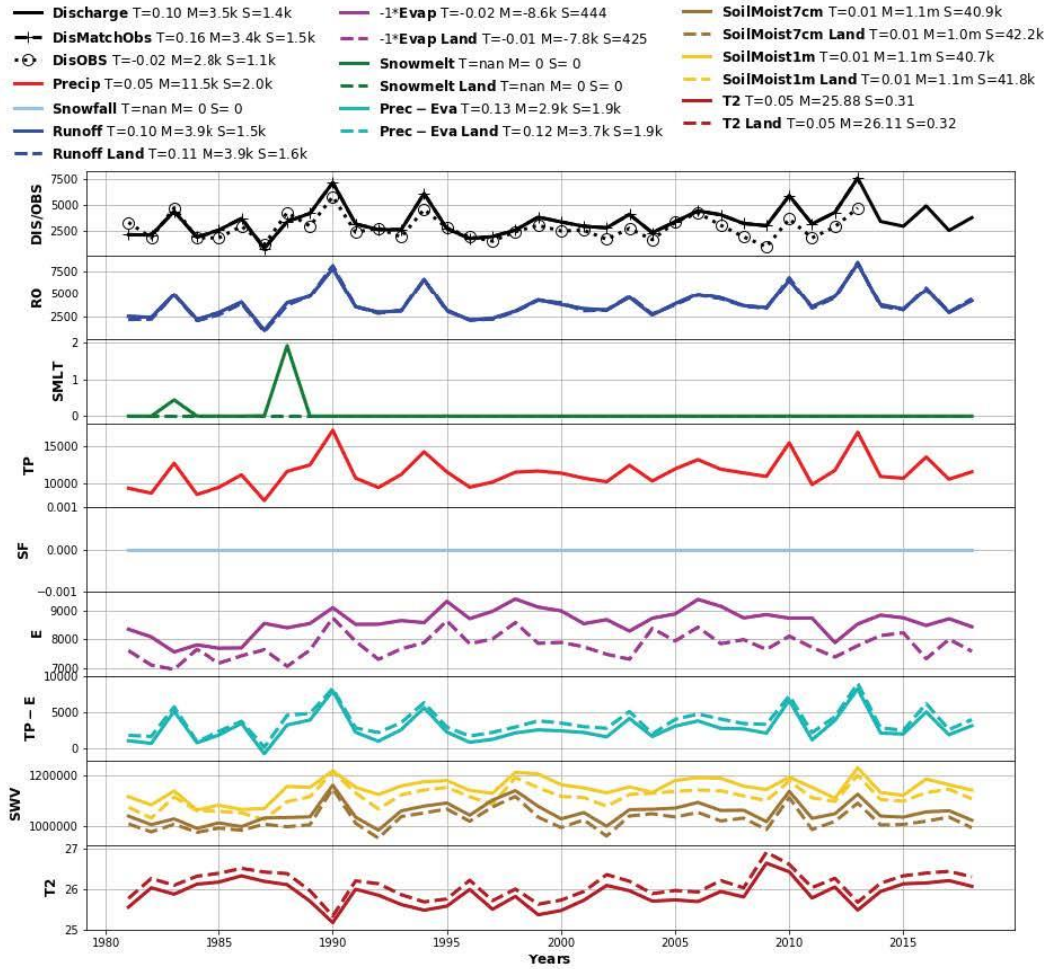
South Asia - No.2

River: Ganges, Outlet: [24.05N; 89.05E], Area: 951786 km2



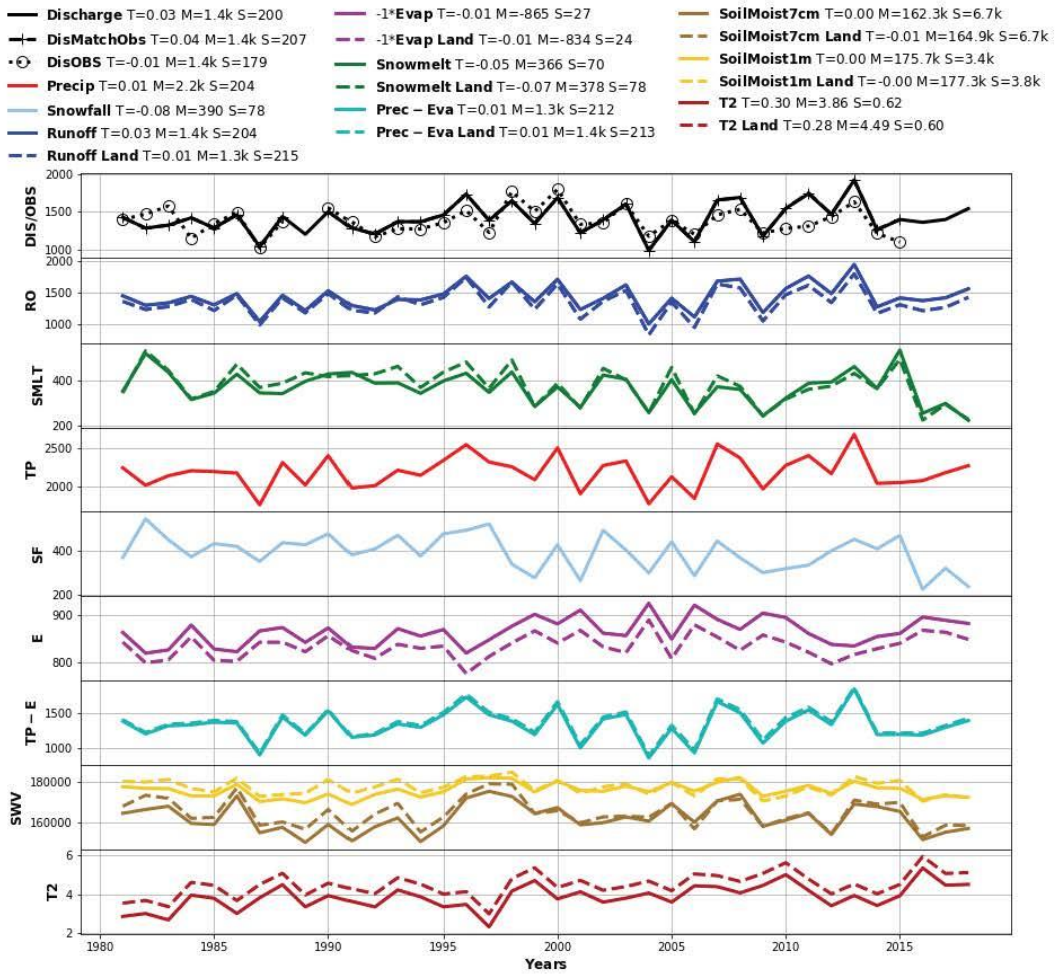
South Asia - No.3

River: Godavari, Outlet: [17.15N; 81.65E], Area: 311631 km2



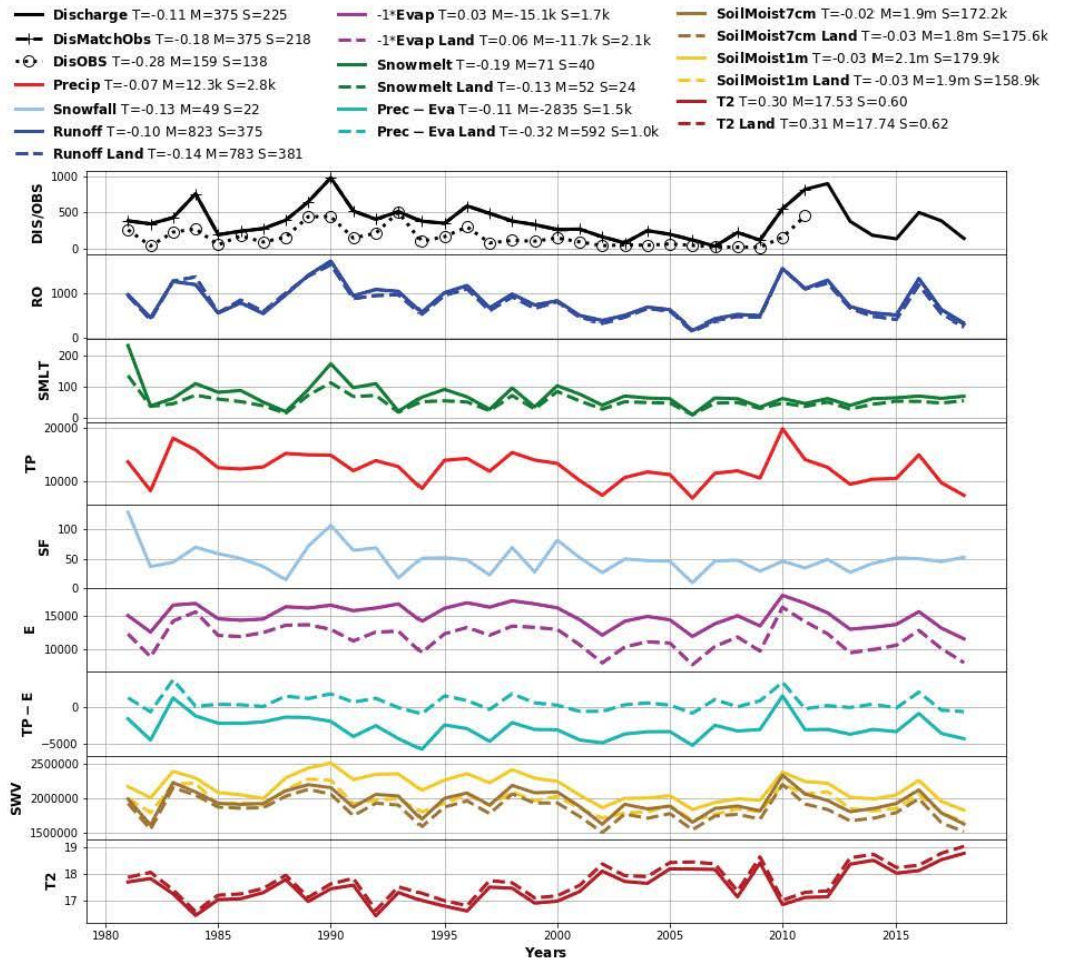
South Asia - No.4

River: Karnali, Outlet: [28.75N; 81.25E], Area: 46325 km2



Australia - No.1

River: Murray, Outlet: [-34.35S; 139.65E], Area: 713793 km²



A4: Using ensemble reforecasts to generate flood thresholds for improved global flood forecasting

This paper presents the published version of chapter 7 of this thesis, with the following reference:

Zsoter, E., C. Prudhomme, E. Stephens, F. Pappenberger and H. Cloke, 2020: Using ensemble reforecasts to generate flood thresholds for improved global flood forecasting, *J. Flood Risk Management*, doi:10.1111/jfr3.12658


Received: 19 February 2020 | Revised: 28 May 2020 | Accepted: 1 July 2020

DOI: 10.1111/jfr3.12658

ORIGINAL ARTICLE


 Chartered Institution of Water and Environmental Management
 
 Journal of Flood Risk Management
 
 WILEY

Using ensemble reforecasts to generate flood thresholds for improved global flood forecasting

Ervin Zsoter^{1,2}  | Christel Prudhomme^{1,3,4} | Elisabeth Stephens² | Florian Pappenberger¹ | Hannah Cloke^{2,5,6,7}

¹European Centre for Medium-Range Weather Forecasts, Reading, UK

²Department of Geography and Environmental Science, University of Reading, Reading, UK

³UK Centre for Ecology & Hydrology, Wallingford, UK

⁴Department of Geography, Loughborough University, Loughborough, UK

⁵Department of Meteorology, University of Reading, Reading, UK

⁶Department of Earth Sciences, Uppsala University, Uppsala, Sweden

⁷Centre of Natural Hazards and Disaster Science, CNDS, Uppsala, Sweden

Correspondence

Ervin Zsoter, European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, UK. Email: ervin.zsoter@ecmwf.int

Funding information

Natural Environment Research Council, Grant/Award Numbers: NE/K00896X/1, NE/P000525/1

Abstract

Global flood forecasting systems rely on predefining flood thresholds to highlight potential upcoming flood events. Existing methods for flood threshold definition are often based on reanalysis datasets using a single threshold across all forecast lead times, such as in the Global Flood Awareness System. This leads to inconsistencies between how the extreme flood events are represented in the flood thresholds and the ensemble forecasts. This paper explores the potential benefits of using river flow ensemble reforecasts to generate flood thresholds that can deliver improved reliability and skill, increasing the confidence in the forecasts for humanitarian and civil protection partners. The choice of dataset and methods used to sample annual maxima in the threshold computation, both for reanalysis and reforecast, is analysed in terms of threshold magnitude, forecast reliability, and skill for different flood severity levels and lead times. The variability of threshold magnitudes, when estimated from the different annual maxima samples, can be extremely large, as can the subsequent impact on forecast skill. Reanalysis-based thresholds should only be used for the first few days, after which ensemble-reforecast-based thresholds, that vary with forecast lead time and can account for the forecast bias trends, provide more reliable and skilful flood forecasts.

KEYWORDS

ensemble reforecasts, flood forecasting, flood thresholds, forecast lead times, global predictions, reanalysis, river discharge

1 | INTRODUCTION

Flood forecasting systems use meteorological data and hydrological modelling to deliver forecasts of river discharge and other hydrological variables such as inundation or soil moisture. They provide early flood warnings on time scales up to several weeks ahead, essential for

managing flood risk at local, regional, and recently also on the global scale (Emerton et al., 2016).

The state-of-the-art systems in use today provide an ensemble of equally likely solutions that can be used to define occurrence probabilities for certain flood events (Cloke & Pappenberger, 2009; Wu et al., 2020). These flood events are defined by comparing the forecast time

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Journal of Flood Risk Management* published by Chartered Institution of Water and Environmental Management and John Wiley & Sons Ltd.

J Flood Risk Management. 2020;13:e12658.
<https://doi.org/10.1111/jfr3.12658>

wileyonlinelibrary.com/journal/jfr3 | 1 of 14

series with flood thresholds, usually based on a return period magnitude or a quantile.

In the Global Flood Awareness System of the Copernicus Emergency Management Service (GloFAS; Alfieri et al., 2013, Hirpa et al., 2018), the severity of the predicted flood is defined according to a set of three thresholds, as shown in Figure 1 for the example of tropical cyclone Idai in Mozambique in March 2019. These thresholds are computed from a 40-year long river discharge reanalysis (Harrigan et al., 2020). The hydrograph in Figure 1 shows the predicted river discharge for the next 30 days, highlighting a severe flood event around 18–21 March with 10–15% chance of exceeding the 5% annual exceedance probability (AEP) threshold.

The flood thresholds, defined according to flood magnitude of selected return periods (or flood quantiles), and used in many of the existing flood prediction systems (GloFAS Alfieri et al., 2013; EFAS Thielen, Bartholmes, Ramos, & de Roo, 2009; WW-HYPE Arheimer et al., 2020), are determined by flood frequency analysis, usually by fitting an extreme value distribution on a set of annual maxima, sampled from a time series as long as possible. These quantities describe the likelihood of

different flood magnitudes occurring locally based on a “climatological” data set over a long period of time (preferably 30 years or more; World Meteorological Organisation [WMO], 2017). Traditionally, flood thresholds are produced from observations or deterministic model reanalysis (Alfieri et al., 2015). River discharge observations can provide a solution only at certain locations, whereas hydrological model simulations, forced with meteorological observations, can cover a whole geographical domain, delivering flood thresholds at every model river point or catchment.

Because the flood thresholds determine the severity of the forecasted flood signal, these flood thresholds should ideally represent extreme events the same way as they occur in the forecasts. If this is not the case and the different biases make an event of the same magnitude occur with a different frequency in the climatological data set that was used to compute the thresholds and in the forecasts (e.g., the 5% AEP flood magnitude happens more often in the forecasts than the expected 5% probability in a given year), then the flood forecast probabilities could become unreliable (e.g., leading to flood signals that often overestimate the flood severity). In the case of the

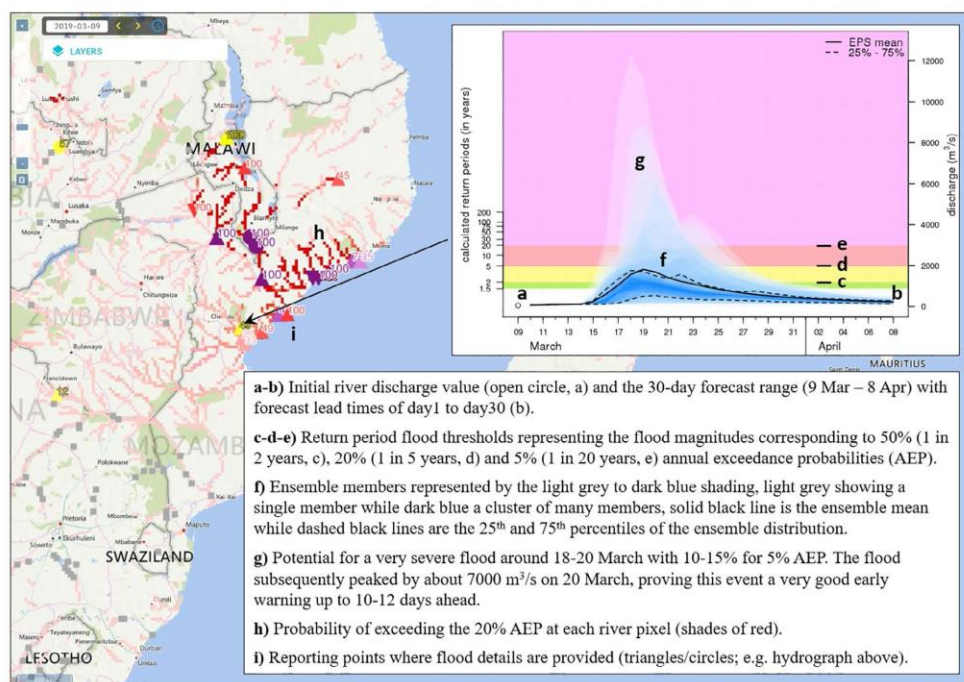


FIGURE 1 GloFAS forecast on March 9, 2019 for Mozambique showing flood predictions related to tropical cyclone Idai. As an example, the inset diagram shows the hydrograph for a river point near the coast in Mozambique for the 30-day period of 9 March to 8 April. GloFAS forecasts are openly accessible on www.globalfloods.eu

example in Figure 1, this could mean that the predicted severe flood event should in fact appear significantly less extreme as the high severity would only be a consequence of the unrealistically low thresholds.

The extreme event representation of flood thresholds can be heavily influenced by the data set and the method used to derive the thresholds. The value of the flood quantiles can be impacted by the choice of the extreme value statistical distribution (Papalexou & Koutsoyiannis, 2013), the data set that is used for the annual maxima extraction (observation, reanalysis or forecasts; see, for example, Hirpa et al., 2016) and also its length (Kjeldsen, Lamb, & Blazkova, 2014). These can all lead to potential differences in the flood threshold magnitudes, subsequently resulting in differences in the forecast probabilities to exceed the thresholds, and ultimately causing an impact on the quality of the flood warnings.

By definition, conventional observation- or reanalysis-based data sets provide a single time series to compute flood thresholds, meaning that only one set of thresholds, with different severities, is going to be applied to all lead times in the forecast range. This might cause further inconsistencies if the forecast biases have trends across lead times. For example, forecasts might show increasing river discharge overprediction with lead time, which would result in a growing number of forecasts exceeding the 5% AEP flood threshold (which stays unchanged as it is computed from the reanalysis time series), with an increasingly higher frequency than the expected 5% of the years occurrence on average (Alfieri et al., 2019). As trends and biases in a forecast are model specific, using different meteorological forcing models within the same forecasting system (such as in the European Flood Awareness System, EFAS, Thielen, Bartholmes, Ramos, & de Roo, 2009) might cause even more complex inconsistencies between the observation- or reanalysis-based flood thresholds and the forecasts.

Bias correction methods can help to achieve consistency between forecasts and thresholds (e.g., Verkade, Brown, Reggiani, & Weerts, 2013; Yuan & Wood, 2012). They have the potential to make the extreme event representation of the forecasts and the climatology, that is used to define the thresholds, similar. However, bias correction, even in its simplest form with only hydrological output postprocessing (without correction of the meteorological forcing data), would introduce further complexity into the river discharge production chain with its associated uncertainties.

Alternative approaches have been investigated (e.g., Alfieri et al., 2019). Generally, flood thresholds are not produced from forecasts. Part of the reason could be the limited sample of available historical forecasts and

also the convenience for the users to work with only a single threshold set that does not show evolution with lead time. The consequence is that, as said earlier, the same threshold is applied to all forecast lead times. However, Alfieri et al. (2019) showed that range-dependent, reforecast-based thresholds were substantially different from unique reanalysis-based thresholds in two thirds of the global rivers. Moreover, despite the recent advancement of ensemble-based forecast systems, ensemble forecasts are generally not considered in the flood threshold generation. However, this can be a problem as ensembles can have different biases to single deterministic forecasts (Leutbecher et al., 2017), which can further contribute to the extreme event representation inconsistencies between reanalysis-based thresholds and ensemble forecasts.

The use of ensemble reforecasts in generating the climatological sample can provide a range-dependent threshold set (e.g., as in Emerton et al., 2018 and Tsonevsky, Doswell, & Brooks, 2018), which has the potential to overcome the issues associated with extreme event identification. In addition, multi-value ensembles can also contribute to increased effective sample size, from which to define flood thresholds, and therefore help to improve the representation of extreme events (Zsoter, Pappenberger, & Richardson, 2014). This could be important for very extreme events which might not occur in the typical 30–50-year-long sample of traditional observation or reanalysis time series (e.g., the median length of the daily data in the Global Runoff Data Centre is 39 years, as of January 22, 2020 at www.bafg.de/GRDC).

In this study, the potential benefits of using river discharge ensemble reforecasts to define flood thresholds are analysed globally. Two main research questions were explored in our study, targeting specifically the sampling strategy to extract the annual maxima sample on which the flood frequency analysis is conducted:

- How adequate it is to use a reanalysis dataset to define flood thresholds and apply them for all forecast lead times?
- How best to use reforecast ensemble information in the flood threshold generation to improve flood forecast performance?

The work is carried out in the context of GloFAS, for a 30-day forecast range, with a selection of over 5,000 catchments. The impacts of the choice of data source (reanalysis or reforecasts) and of the annual maximum sampling strategies (from the reforecasts) are analysed by comparing the flood threshold magnitudes and the resulting forecast reliability and skill benefits for four different flood severity levels.

2 | SYSTEM DESCRIPTION, DATASETS, AND METHODS

This section describes the data sets, methods, and experimental set-up used to generate flood thresholds and compare their value and impact on the flood forecast skill.

2.1 | GloFAS

GloFAS is part of the Copernicus Emergency Management Service (CEMS) and has been developed by the Joint Research Centre of the European Commission and the European Centre for Medium-Range Weather Forecasts (ECMWF) with help from research institutions such as the University of Reading (UoR; e.g., Stephens, Day, Pappenberger, & Cloke, 2015, Emerton, Cloke, & Stephens, 2017 and Towner et al., 2019). It is a probabilistic hydrological prediction system, which has a 30-day (Alfieri et al., 2013) and a seasonal component (Emerton et al., 2018). This study is based on the 30-day component, which predicts daily flood occurrences on the global scale. In GloFAS, ensemble runoff outputs from the HTESSEL land surface model (the Hydrology-Tiled ECMWF Scheme for Surface Exchange over Land; Balsamo et al., 2009; Balsamo, Pappenberger, Dutra, Viterbo, & van den Hurk, 2011) are coupled to the Lisflood hydrological model (van der Knijff, Younis, & de Roo, 2010) to produce an ensemble of daily river discharge across a global river network at 0.1° resolution (Alfieri et al., 2013; Hirpa et al., 2018). To detect the likelihood of high flow situations, to forecast flood events, the real time river discharge forecasts are compared with a set of flood thresholds derived from a 40-year long climatological simulation, a daily river discharge reanalysis time series.

2.2 | River discharge reanalysis

The GloFAS-ERA5 river discharge reanalysis (Harrigan et al., 2020) is produced with ERA5 forcing, ECMWF's fifth generation global climate reanalysis (Hersbach et al., 2018; Hersbach et al., 2020), which is part of the EU-funded Copernicus Climate Change Service (C3S). ERA5 covers the period 1979 to present and is updated with two to 3 months delay. ERA5 is open access (<https://climate.copernicus.eu/>) and includes one high-resolution component and a lower resolution ensemble component with 10 members. The GloFAS-ERA5 uses the high-resolution ERA5 component at ~31 km horizontal resolution with the configuration of the GloFAS operational forecasting systems. GloFAS-ERA5 is a key component of GloFAS verification, serving as a proxy for

river discharge observations and it is also openly available from the Copernicus Climate Change Service Climate Data Store (Harrigan et al., 2020).

2.3 | Ensemble river discharge reforecasts

The ensemble river discharge reforecasts are GloFAS reforecasts produced for the 20-year period of 1997–2016. These are 30-day river discharge forecasts generated for past dates by the same GloFAS system that is used for the real time forecasts. They are initialised from GloFAS-ERA5 and forced by runoff from the twice weekly (Monday and Thursdays in 2017), 11-member, 20-year ECMWF meteorological ensemble reforecasts (Vitart, 2014). This data set includes a batch of 20 reforecasts (one for each year in 1997–2016) for each Monday and Thursday in 2017. Altogether 2080, 11-member, 30-day reforecasts were produced for the 20-year period (104 in each year).

2.4 | Flood thresholds

In the 30-day GloFAS, flood quantiles of three severity levels (2-, 5-, and 20-year return periods) are used as flood thresholds. Flood quantiles are commonly used in risk analysis, typically estimated using time series data of generally twice the length of the return period of interest. Because of the relatively short length of daily discharge data available, the 10-year return period severity was also considered in this study.

A return period T is an estimate of the likelihood of an event to occur (Gumbel, 1941), expressed as average number of years for an event of same or higher magnitude to occur. It can also be expressed as an Annual Exceedance Probability AEP (given by $AEP = 100/T$). To facilitate the interpretation, AEP is used in the rest of this study.

The flood thresholds were computed as currently done operationally in GloFAS: the Gumbel Extreme Value Distribution (EVD) is fitted to the annual maximum river discharge sample using the method of L-Moments (Hosking, 1990). This method is appropriate for relatively small sample sizes, such as used in GloFAS (Alfieri et al., 2019) and in this study (20 years of reforecast data from 1997 to 2016).

2.5 | River catchments

The study is based on the GloFAS network (Figure 2), a set of 6,122 catchments of which about

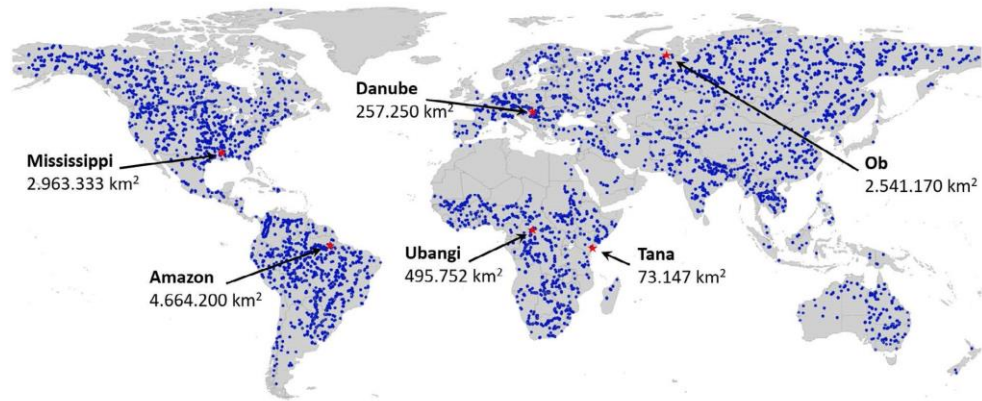
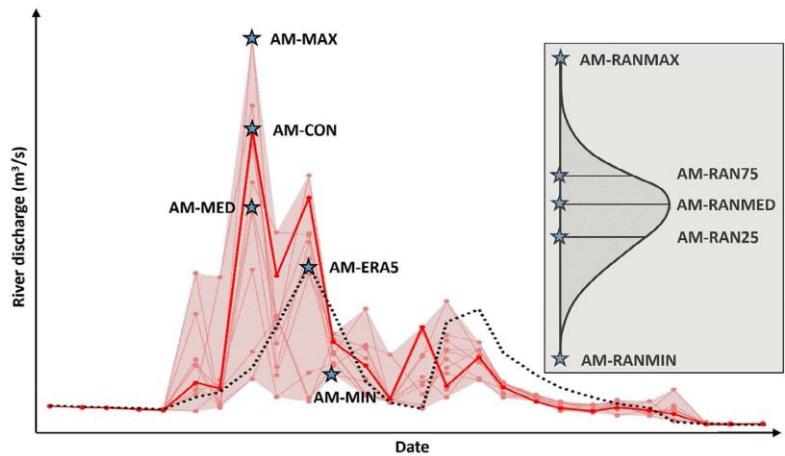


FIGURE 2 The 5,665 GloFAS stations used in this study. The six contrasting catchments of Figure 5 are indicated by red stars, along with the river names and GloFAS upstream areas

FIGURE 3 Schematic of the annual maximum sampling for flood threshold estimates from daily river discharge time series. Dotted black line: GloFAS-ERA5, solid red line: GloFAS reforecast control member, light red lines: GloFAS reforecast perturbed ensemble members. Small red dots show individual daily river discharge values in the ensemble reforecasts. The x-axis shows the date of the forecasts, while the y-axis the river discharge values



Version	Time series set	Version	Time series set
T-ERA5	Benchmark set	T-RANMIN	Minimum extended reforecast set
T-CON	Control member reforecast set	T-RAN25	25th percentile extended reforecast set
T-MIN	Minimum member reforecast set	T-RANMED	Median extended reforecast set
T-MED	Median member reforecast set	T-RAN75	75th percentile extended reforecast set
T-MAX	Maximum member reforecast set	T-RANMAX	Maximum extended reforecast set

one-third are always highlighted on GloFAS website as reporting points (www.globalfloods.eu). This network provides a global coverage and includes all points where daily historical river discharge observations are made available to the GloFAS team. Catchments that have 50% AEP magnitude below 20 m³/s in GloFAS-ERA5, that is, too dry or too small, were excluded from the study, resulting in 5,665 catchments in total for the analysis.

2.6 | Analysis methods

The impact of flood threshold estimation for all four severity levels was analysed for each catchment by direct comparison of the quantile magnitudes. Additionally, the flood forecast performance was evaluated for day 1 to day 30 lead times by:

- Comparing the number of events forecasted (i.e., when the discharge exceeds the flood threshold) with the

number of events identified in the benchmark set (i.e., GloFAS-ERA5 river discharge reanalysis which is the nearest equivalent to the “observations”), expressed as percentage occurrence frequency (or event/forecast probability). This step analyses the simplified forecast reliability with only one probability category;

- Calculating the Brier score (Murphy, 1973) and the reliability diagram (Hsu & Murphy, 1986). This step assesses both the skill and reliability in the resulting probability forecasts of exceeding the thresholds.

2.7 | Experimental set-up

For consistency and comparability, the annual maxima sampling for the flood threshold computation was done from daily time series containing only the calendar days corresponding to the dates of the day 1 to day 30 reforecast values (for all Monday and Thursday reforecast runs of 1997–2016). For each lead time, three sets of time series were used:

- *Benchmark set*: GloFAS-ERA5 river discharge reanalysis (independent from the lead time). This is as close as possible to the flood thresholds used operationally in GloFAS and can be considered as proxy observation-based thresholds;
- *Reforecast set*: the time series of the control member, plus three time series corresponding to the minimum, median and maximum values from each run of the 11-member GloFAS reforecasts;
- *Extended reforecast set*: 1,000 time series, each generated with randomly selecting one of the 11 ensemble members from each GloFAS reforecast.

After applying the flood threshold generation method, described earlier, this resulted in 1 + 4 + 1,000 threshold values summarised graphically in Figure 3 for the annual maxima selection differences. From the 1,000 random-member-based thresholds only the minimum, 25th percentile, median, 75th percentile, and maximum values were analysed further. The exercise was conducted on all study catchments, flood severity levels, and forecast lead times. The major methodological steps of this study are provided in Table 1.

3 | RESULTS

The impact of the data set and sampling strategy choice in the flood threshold generation was analysed globally on selected river catchments, with the flood threshold magnitude and forecast skill compared geographically.

TABLE 1 Major methodological steps of this study

Steps	Description
Setup	Ensemble reforecasts for day 1 to day 30 lead times, over 20 years (1997–2016), with 104 forecasts in each year, flood thresholds computed by fitting an extreme value distribution on the 20 annual maxima, for 5,665 global catchments and 4 return periods (50, 20, 10, and 5% AEP)
Benchmark (reanalysis) thresholds	Produce reanalysis-based reference thresholds (T-ERA5) for all lead times, always with the days of the reforecasts, to guarantee homogeneous samples
Reforecast thresholds	Produce ensemble-forecast-based alternative thresholds (T-CON, T-MIN, T-MED, T-MAX)
Extended reforecast thresholds	Produce random-ensemble-member-based thresholds 1,000 times (T-RAN)
Extended reforecast threshold distribution	Define the key statistics of the extended reforecast threshold distribution (T-RANMIN, T-RAN25, T-RANMED, T-RAN75, and T-RANMAX)
Probabilities	Compute the exceedance probabilities over the 20-year period with all threshold versions (10 in total)
Scores	Compute the Brier scores and produce the reliability diagrams with all threshold versions (10 in total), for all catchments including a global average

Note: For all lead times, catchments and return periods if not otherwise stated.

3.1 | How similar are the flood thresholds?

In this section, we analyse the impact of the annual maxima sampling strategy on the flood threshold magnitude for day 1 and day 30 lead times, focusing on the 10% AEP severity level.

Flood threshold magnitudes, derived from reforecasts, depend on lead-time with values less than 5% different from those derived from the reanalysis (T-ERA5) for day 1 (Figure 4a,c,e), but exceeding 50% difference over large parts of the world by day 30 (Figure 4b,d,f), regardless of the ensemble reforecast sampling strategy. However, there is a large spatial

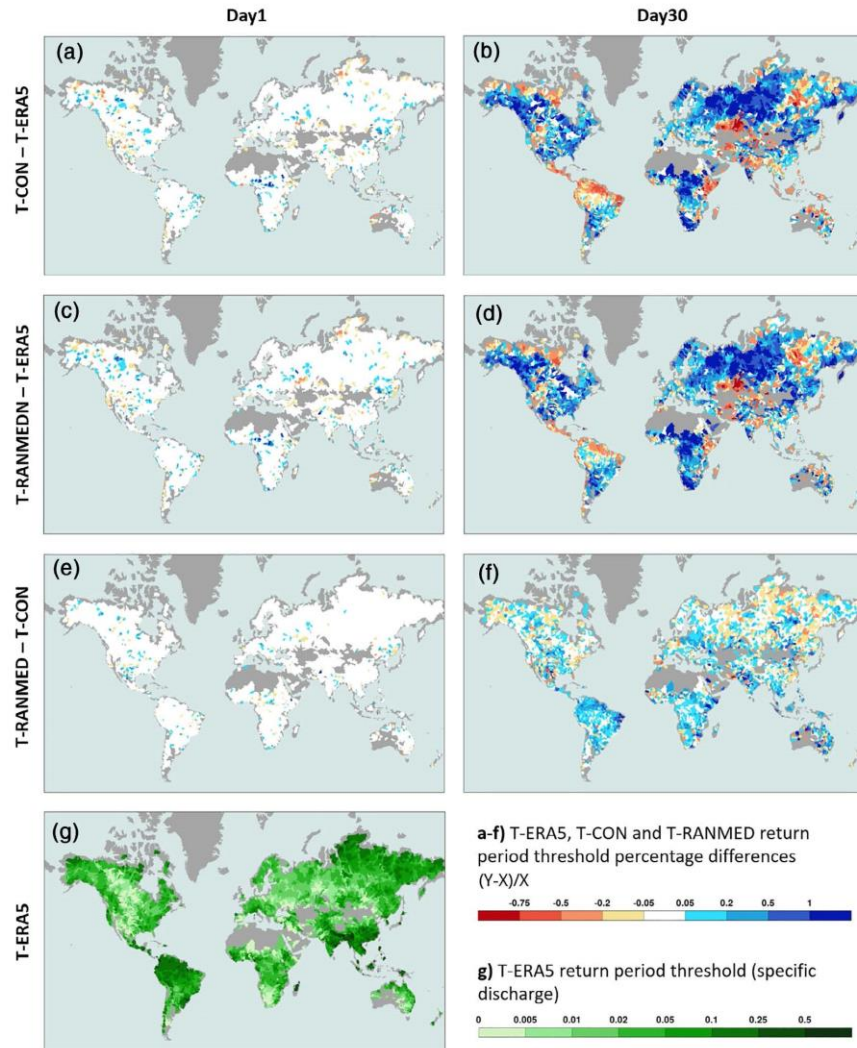


FIGURE 4 Percentage difference of 10% AEP flood thresholds between (a, b) T-CON and T-ERA5, (c, d) T-RANMED and T-ERA5 and (e, f) T-RANMED and T-CON based on the 1997–2016 period. The left column (a, c, e) is for day 1 while the right one (b, d, f) is for day 30 lead time. Percentage differences of orange (blue) colour palette mean lower (higher) flood thresholds respectively in T-CON (vs. T-ERA5) and in T-RANMED (vs. T-ERA5 and T-CON). Panel (g) shows the reference T-ERA5 threshold magnitudes as specific river discharge (river discharge divided by the upstream area in km^2 in order to scale better between different catchment sizes)

variability, with most of the world showing reforecast flood thresholds larger than T-ERA5, except in north Canada, Central and northern South America, Central Asia, the Horn of Africa, and some of east Russia. This confirms earlier finding from Alfieri et al. (2019) that for extended-range lead times, the flood frequency distribution of hydrological forecasts is not well represented by reanalysis simulations.

Figure 4 also shows that using the control member of the reforecasts to derive the flood threshold (T-CON) leads to different results than sampling the full reforecast ensemble (T-RAN family). Specifically, T-CON is systematically higher than T-RANMED in most regions (Figure 4e,f), with differences growing with lead time, generally below 20% but reaching 100% in some places by day 30. This suggests that the control

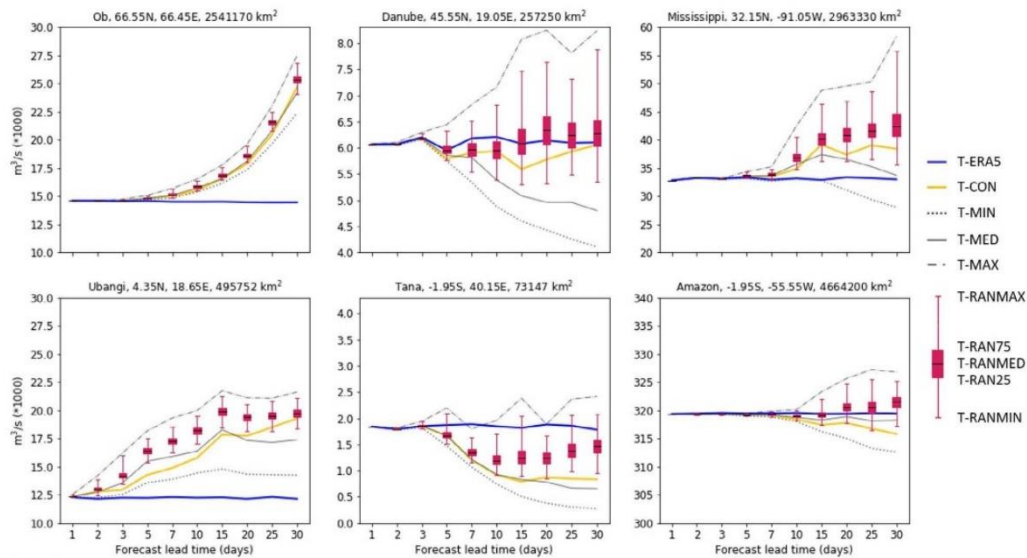


FIGURE 5 Flood thresholds of 10% AEP severity level based on the 1997–2016 period as function of forecast lead time for six contrasting catchments: benchmark T-ERA5 (blue), reforecast control T-CON (orange), minimum T-MIN (dotted grey), median T-MED (solid grey) and maximum T-MAX (dash-dotted grey) and also the extended reforecast T-RANMIN, T-RAN25, T-RANMED, T-RAN75, and T-RANMAX (red box whiskers)

forecasts do not fully represent the flood frequency distribution of the ensemble reforecasts and their use could potentially lower the forecast skill.

Results for other severity levels (5, 20, and 50% AEPs) show very similar behaviour. Although there are some variabilities across the severity levels with differences in flood threshold magnitude increasing with the severity level, the percentage differences appear to be in the same order of magnitude (Figures S1–S3).

Figure 5 shows the flood thresholds for the 10% AEP severity level as a function of lead time for six contrasting catchments (consult Figure 2 for the catchment locations). The influence of the ensemble reforecast sampling strategy on the flood threshold magnitudes gets larger with the increasing forecast lead time. For some catchments, such as the Ob and Amazon rivers, the impact is small (interquartile range of below 1% of T-ERA5 by day 30 as shown by the red boxes), but for some other catchments the difference could be as large as 10–20% of the T-ERA5 value at day 30 lead time (Tana and Mississippi rivers). Moreover, the flood thresholds, generated using the control member, are dominantly below the envelope of the ensemble reforecast (e.g., Ubangi and Tana rivers), confirming the general positive pattern already seen in Figure 4f.

Analysis on other flood threshold severity levels indicates that differences between reforecast- and reanalysis-

based thresholds and sampling strategies are generally increasing with both severity level and lead time (Figures S4–S6).

3.2 | How reliable are the forecast probabilities?

In this section, we investigate the match between the flood forecast probabilities and the flood occurrence frequencies, using the benchmark, the reforecast control, and the extended reforecast median flood thresholds, defined for the 10% AEP severity level (Figure 6).

At day 1 (Figure 6a–c), flood forecasts are very reliable regardless of the flood threshold generation used (points close to the diagonal line), but this is lost by day 30 (Figure 6d–f). The largest loss of reliability is found when using the benchmark flood threshold (T-ERA5), with many catchments showing too high flood forecast probability (points way above the diagonal line), suggesting that the T-ERA5 thresholds are too low. The performance using reforecasts-based thresholds shows a clear improvement over using T-ERA5, especially reducing the number of catchments with large flood forecast probability overestimation. Results based on T-RANMED are slightly better than those using T-CON with a larger cluster around the diagonal line (91 vs. 86% of the

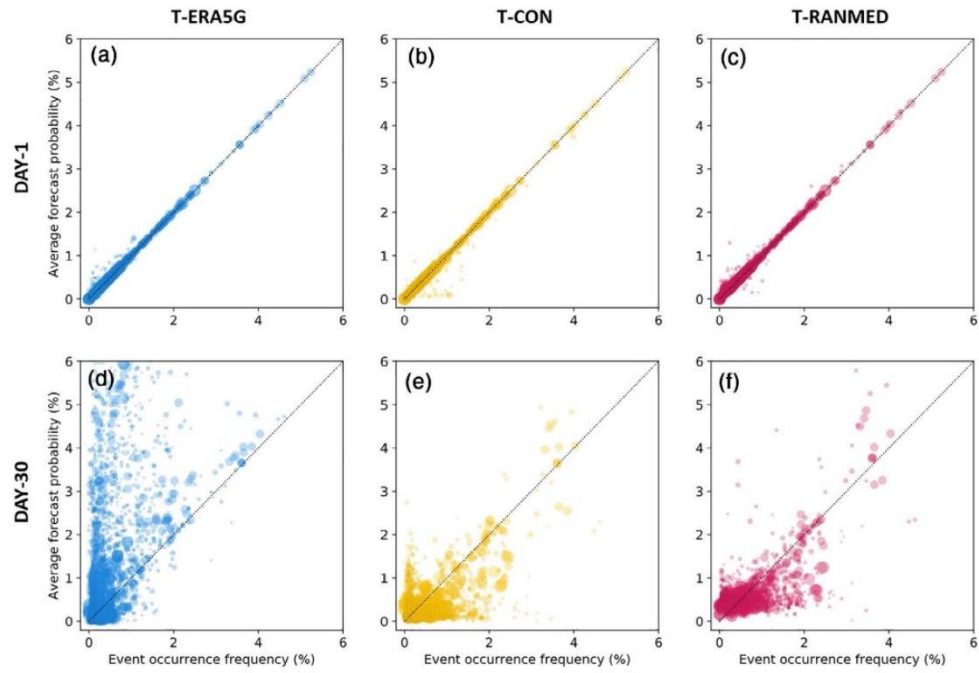


FIGURE 6 Scatter plot of day 1 (top) and day 30 (bottom) flood forecast probability (y-axis) against flood occurrence frequency (x-axis) using flood thresholds of T-ERA5 (a and c, blue), T-CON (b and e, orange), and T-RANMED (c and f, red) based on the 1997–2016 period. Dot size is proportional to catchment size

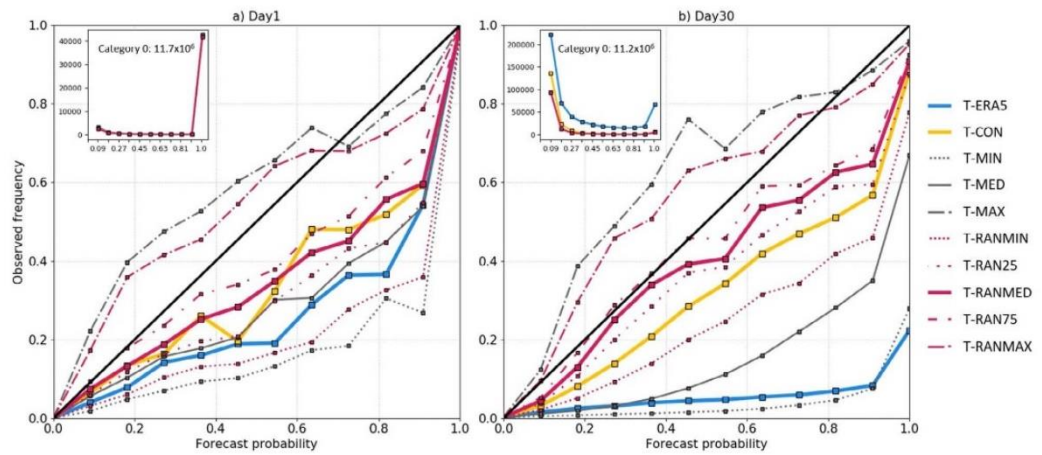


FIGURE 7 Reliability diagram for flood event forecast probabilities above 10% AEP based on the 1997–2016 period for (a) day 1 and (b) day 30 using flood thresholds based on the benchmark (T-ERA5), reforecast (T-CON, T-MIN, T-MED, and T-MAX), and extended reforecast (T-RANMIN, T-RAN25, T-RANMED, T-RAN75, and T-RANMAX) sets. The inset shows the distribution of number of cases in all 11 probability categories. The first category (0 ensemble member forecasting the event) is only indicated as a number

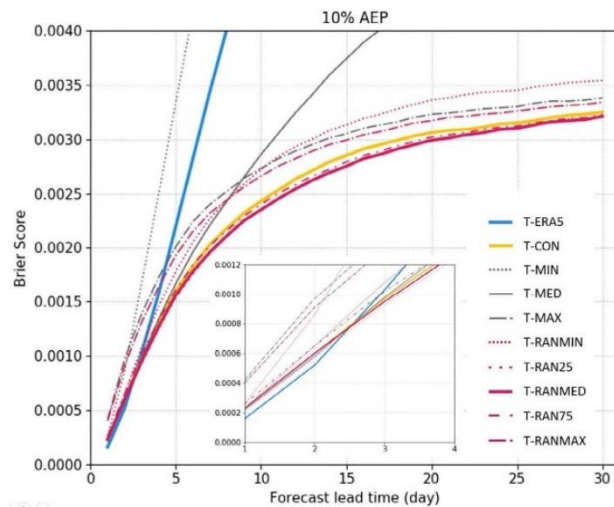


FIGURE 8 Brier score for flood event forecasts above 10% AEP over the 1997–2016 period for day 1 to day 30 using the benchmark (T-ERA5), reforecast (T-CON, T-MIN, T-MED, and T-MAX), and extended reforecast (T-RANMIN, T-RAN25, T-RANMED, T-RAN75, and T-RANMAX) flood thresholds. The inset shows the scores of the first 4 days only for better readability

catchments with less than 0.5% absolute difference between forecast probability and occurrence frequency), showing a stronger, more linear relationship.

3.3 | What is the impact on forecast reliability and skill?

Forecast reliability and skill were further examined using the reliability diagram and the Brier score for flood forecasts produced with the 10% AEP severity level.

As suggested by Figure 7, the reliability of the flood forecasts, based on the T-ERA5 thresholds, is low, especially for day 30 lead time, with an event frequency of less than 10% for almost all flood forecast probability categories (except the largest when it is just above 20%, see blue line close to x -axis). Using reforecast-based flood thresholds can greatly improve the flood forecast reliability, the only exceptions being T-MIN and also T-RANMIN for day 1 lead time. T-MAX and T-RANMAX tend to systematically underestimate flood event frequency up to 70–80% forecast probability, whilst overestimation of flood events is systematic for all other thresholds. The response is similar across all considered four severity levels except for flood thresholds of 50% AEP, where the reforecast thresholds become too high, making the flood forecast probabilities too low (points are above the diagonal; Figures S7–S9). Generally, thresholds based on the full ensemble can provide better reliability than using T-CON, this is especially clear by day 30, when all T-RANMED, T-RAN25, and T-RAN75 are closer to the diagonal line.

Figure 8 shows the general skill of the flood forecasts for the 10% AEP severity level from day 1 to day 30 lead

times, based on the Brier score. The benchmark flood thresholds (T-ERA5) can provide the lowest error only up to day 2 lead time (blue line below the other lines; for other flood severity levels this maximum lead time ranges from day 1 (5% AEP) to day 4 (50% AEP), Figures S10–S12). From day 3, the flood forecasts with reforecast-based thresholds become gradually more skilful than with T-ERA5 (the only exception is T-MIN), consistently with the conclusions of Figure 7b. In fact, both T-ERA5 and T-MIN show poor skill with multiple times higher Brier score values by day 30 (this higher section of the Brier score range is not shown in Figure 8 for better readability). The best performance is achieved by the median (T-RANMED) and the interquartile range boundaries (T-RAN25 and T-RAN25) of the extended reforecast threshold set and the reforecast control thresholds (T-CON), with skill slowly degrading with lead time. The skill improvement, using T-RANMED over T-CON, is statistically significant at the 99% level from day 5–6 lead time (tested by bootstrapping the dates in the verification sample, Figure S13). The pattern is similar for flood events of higher severity (5% AEP), whilst for less severe floods (20 and 50% AEP), the highest skill is achieved using T-RAN25 or T-RANMIN, but T-RANMED is still achieving high skill (Figures S10–S12).

4 | DISCUSSION

The global analysis conducted here showed that using flood thresholds based on reforecasts improved substantially the forecast performance after the first 1–4 days of the forecast range (depending on the flood severity levels)

compared with using thresholds based on reanalysis, as done operationally in most forecast systems. One of the key advantages is the lead time specific definition of thresholds, which accounts for the changing representation of extreme event frequencies in the forecasts. Overall, forecast errors are reduced by up to 2–4 times compared with reanalysis-based thresholds, depending on the flood severity and lead time. Results also showed that using the single unperturbed control member to define the thresholds is not sufficient and exploring the full ensembles of the reforecasts in the threshold derivation further increases the forecast reliability and skill.

4.1 | Ensemble member independence

Using ensemble members allows a better representation of the extreme events in the forecast climatology by increasing the sample size. The ensemble members are correlated to some extent by sharing the same initial condition, especially at the beginning of the 30-day forecast horizon. Correlation between ensemble members reduces with increased lead time when each ensemble member drifts towards becoming an independent and identical random sample from the mode climate. In addition, for each of the reforecast-based threshold methods, only one member was chosen from all the twice weekly reforecasts in the 20-year period in order to increase independence. This guaranteed that the correlation between the individual reforecast values in the climatological sample remained very small. This made them an effectively independent realisation of the true underlying model climate distribution, ultimately providing an appropriate basis for the extreme value distribution fitting in the flood threshold computation.

4.2 | Best performing thresholds

The median of the extended reforecast threshold set, produced by using one random ensemble member from the reforecasts, provides the best overall performance, however, for lower severity levels some other reforecast-based thresholds can be slightly better. This can be related to the nonlinear response between reforecast ensemble time series and flood quantile estimation. In particular, with increasing lead time, outliers associated with very high forecasted river discharge become more likely within the 11 ensemble members. The annual maximum selection then will over-represent the high outliers through the random member selection process, as even if only one very high forecast value is selected in 1 or 2 years, it is likely to shift the estimate of the 5–10% AEP flood

quantile to a high value. This potential increase of flood threshold value with lead time does not affect the forecast probabilities of flood event occurrence to the same extent, as the probabilities are calculated considering the full ensemble and are influenced much less by these relatively rare outliers in some of the reforecast members. This different effect of outliers on flood thresholds and flood forecast probabilities will translate into inconsistent reliability and skill impact associated with the various ways to sample the reforecasts to produce flood threshold, and could result, in some cases, in favouring a different sampling strategy than picking up the median.

4.3 | Biases in the forecasts

Using range-dependent flood thresholds, based on ensemble reforecasts, can account for the evolving biases in the forecasts across the forecast range. This study demonstrated that biases can grow large, affecting the extreme event representation and the use of flood thresholds in medium to extended range hydrological forecast systems like GloFAS 30-day. These biases can originate from the meteorological forcing and impact the hydrological simulations, mainly through precipitation and marginally also temperature, humidity, wind, and radiation, as shown by Zsoter et al. (2016) for the first 10 days of the forecasts. Another likely source for the biases is the land data assimilation (LDAS) impact documented by Zsoter et al. (2019). The LDAS can result in not conserving the water budget in coupled land surface models such as used in GloFAS, possibly contributing to biases seen in the GloFAS-ERA5 reanalysis across large parts of the world (Harrigan et al., 2020). In GloFAS, the reforecasts are initialised from GloFAS-ERA5, but with increasing lead time, the influence of LDAS on reforecast gradually decreases. This means that biases coming from the LDAS impact will remain present in any reanalysis-based flood threshold (in our case GloFAS-ERA5) but will slowly disappear with lead time in reforecasts-based flood thresholds. This inconsistency is likely to contribute to the large differences between the GloFAS-ERA5- and the ensemble-forecast-based thresholds shown in this study.

4.4 | Forecast post-processing

Post-processing of the forecasts against the reference dataset used to derive the flood thresholds (i.e., in our case, GloFAS-ERA5) is an alternative to ensemble reforecast-based thresholds. By removing biases in the forecasts (e.g., linear regression or quantile mapping; see Wentao et al., 2017 for a review of methods), the extreme

event representation of the forecasts would be expected to become similar to that of the reference dataset or climatology. The use of post-processing techniques to create a consistent system between forecasts and flood thresholds was beyond the scope of this paper but could be pursued in the future.

4.5 | Modelling system independence

Whilst the research was conducted on the GloFAS flood forecasting system (based on the HTESSEL land surface model), the main findings of this work are expected to be independent of the modelling system, the extreme value fitting method or the sampling period length used. Although a different fitting method or sampling length could inevitably change the flood thresholds locally, in the global context, they are expected to have a neutral impact on the relation of the threshold magnitudes amongst the different annual maxima sampling methods. In addition, the forecast biases are bound to be modelling system related, which will inevitably change the flood threshold behaviour across the forecast lead times. However, the benefit of using ensemble-based, lead-time-specific thresholds is expected to be general and not dependent on the actual underlying bias behaviour. This is supported by the consistent results found using the Lisflood hydrological model in Alfieri et al. (2019), where the reforecast-control-member-based flood thresholds showed significant biases compared with the ERA5-based thresholds, confirming the benefit of using ensemble reforecasts.

4.6 | Practical recommendations for flood applications

Severe problems can arise in flood forecasting because of the potential issue with inconsistencies between the representation of extreme event frequencies in the thresholds and the forecasts, due to the biases that might be present especially for longer lead times. We recommend that forecast system developers should evaluate these potential inconsistencies for themselves using the methodology presented in this paper. We further recommend that this should be carried out with the use of reforecasts where they are available. But even where this is not the case, attempts should be made to diagnose the biases in the climatological data and the available historical forecasts for potential inconsistencies. Without addressing this inconsistency issue, the reliability and skill of the forecast flood events, and thus the quality of the flood warnings, could be substantially reduced, which could

strongly impact on the decision-making process and ultimately lead to loss of confidence in the products.

In addition, even though the reanalysis-based flood thresholds are proven to be preferred in the first days of the forecast range, the difference in forecast skill to the reforecast-based thresholds is small. We recommend that it is both sensible and practical using the ensemble reforecasts for computing the flood thresholds for all forecast lead times and flood severity levels. Similarly, the best performing thresholds for the more impactful high floods (below 20% AEP) were generated from the median of a large number of random ensemble member selections from each reforecast. Although they are not necessarily the most favourable thresholds for smaller floods, they are the best overall choice and are recommended to be used for flood predictions across all flood severities and forecast lead times.

This study highlighted that flood forecasting applications, such as GloFAS, which use flood thresholds generated from a single time series (reanalysis or observation), can greatly benefit from using ensemble-forecast-based thresholds instead, as a practical and effective way to resolve inconsistencies between forecasts and flood thresholds, and therefore increasing the flood forecast skill.

5 | CONCLUSIONS

Using reliable thresholds in global flood forecasting, that truly reflect the flood event frequencies of the real-time ensemble forecasts across all forecast lead times, is very important. The generation of flood signals with such thresholds can provide the highest forecast reliability and skill, which then gives the best chance to create trust in the users for the application.

In this paper, different annual maxima sampling methods were analysed to generate flood thresholds, using both GloFAS-ERA5 river discharge reanalysis and ensemble reforecasts. The flood thresholds were compared and their impact on the forecast reliability and skill was evaluated.

Reanalysis-based thresholds were found appropriate for the first 1–4 days of the 30-day (depending on the flood severity level) forecast range only. For longer lead times, both global average forecast reliability and skill deteriorate, effectively due to the increasing forecast biases over large parts of the world not accounted for in the reanalysis-based thresholds. The ensemble-forecast-based thresholds provide increasing improvement over the reanalysis-based thresholds for up to the evaluated day 30 lead time. Additionally, using flood thresholds that sample the full ensemble in the reforecast, was

found to be overperforming a simple, single member sampling strategy (e.g., using the control reforecast), with generally better reliability and higher skill of the forecast probability.

The results of this study suggest that acknowledging the large uncertainty coming from the data sampling method in flood threshold generation is a crucial step in understanding and improving forecast skill, so that the system configuration that provides the highest reliability and lowest error globally can be found. In turn, better flood forecasts and better flood warnings could be delivered to the public, increasing the confidence and uptake of these products. Ultimately, the increase in confidence in the flood forecasts should result in better flood preparedness for humanitarian and civil protection partners, potentially reducing damages and casualties world-wide.

ACKNOWLEDGEMENTS

Ervin Zsoter's PhD is supported by the Wilkie Calvert Co-Supported PhD Studentships at the University of Reading. Ervin Zsoter and Christel Prudhomme were supported by the Copernicus Emergency Management Service—Early Warning Systems (CEMS-EWS [EFAS]). Hannah Cloke is supported by the TENDERLY project: Towards END-to End flood forecasting and a tool for Real-time catchment susceptibility UK NERC Flooding From Intense Rainfall (FFIR) programme, NE/K00896X/1. Elisabeth Stephens and Hannah Cloke are supported by the FATHUM project: Forecasts for Anticipatory HUMANitarian Action funded by UK NERC as part of their Science for Humanitarian Emergencies & Resilience (SHEAR) programme, NE/P000525/1.

AUTHORS' CONTRIBUTIONS

Ervin Zsoter designed the experiment, carried out the flood threshold analysis, and led the writing of the manuscript. Hannah Cloke and Liz Stephens assisted with posing the research question and designing the analysis. Christel Prudhomme and Florian Pappenberger helped designing the research methodology. All authors assisted with writing the manuscript.

DATA AVAILABILITY STATEMENT

The GloFAS-ERA5 river discharge reanalysis is openly available from the Copernicus Climate Change Service Climate Data Store. The GloFAS reforecasts will be made available through an ECMWF data repository in due course. The annual maxima time series and the related flood thresholds for all the analysed sampling methods are available upon request from the authors at ECMWF.

ORCID

Ervin Zsoter  <https://orcid.org/0000-0002-7998-0130>

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. *Hydrology and Earth System Sciences*, *17*, 1161–1175. <https://doi.org/10.5194/hess-17-1161-2013>
- Alfieri, L., Berenguer, M., Knechtel, V., Liechti, K., Sempere-Torres, D., & Zappa, M. (2015). Flash flood forecasting based on rainfall thresholds. In Q. Duan, F. Pappenberger, J. Thielen, A. Wood, H. Cloke, & J. Schaake (Eds.), *Handbook of hydro-meteorological ensemble forecasting*. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-40457-3_49-1
- Alfieri, L., Zsoter, E., Harrigan, S., Hirpa, F., Lavaysse, C., Prudhomme, C., & Salamon, P. (2019). Range-dependent thresholds for global flood early warning. *Journal of Hydrology*, *4*, 100034. <https://doi.org/10.1016/j.jhydro.2019.100034>
- Arheimer, B., Pimentel, R., Isberg, K., Crochemore, L., J.C.M. Andersson, J.C.M., Hasan A., & Pineda, L. (2020). Global catchment modelling using world-wide HYPE (WWH), open data and stepwise parameter estimation *Hydrology and Earth System Sciences*, *24*, 535–559. <https://doi.org/10.5194/hess-24-535-2020>.
- Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., & Betts, A. K. (2009). A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the integrated forecast system. *Journal of Hydrometeorology*, *10*, 623–643. <https://doi.org/10.1175/2008JHM1068.1>
- Balsamo, G., Pappenberger, F., Dutra, E., Viterbo, P., & 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*, 1046–1054. <https://doi.org/10.1002/hyp.7808>
- Cloke, H. L., & Pappenberger, F. (2009). Ensemble flood forecasting: A review. *Journal of Hydrology*, *375*(3–4), 613–626. <https://doi.org/10.1016/j.jhydrol.2009.06.005>
- Emerton, R. E., Stephens, E. M., Pappenberger, F., Pagano, T. C., Weerts, A. H., Wood, A. W., ... Cloke, H. L. (2016). Continental and global scale flood forecasting systems. *WIREs Water*, *3*, 391–418. <https://doi.org/10.1002/wat2.1137>
- Emerton, R., Cloke, H., & Stephens, E. (2017). Complex picture for likelihood of ENSO-driven flood hazard. *Nature Communications*, *8*, 14796. <https://doi.org/10.1038/ncomms14796>
- Emerton, R., Zsoter, E., Arnal, L., Cloke, H. L., Muraro, D., Prudhomme, C., ... Pappenberger, F. (2018). Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS—seasonal v1.0. *Geoscientific Model Development*, *11*, 3327–3346. <https://doi.org/10.5194/gmd-11-3327-2018>
- Gumbel, E. J. (1941). The return period of flood flows. *Annals of Mathematical Statistics*, *12*(2), 163–190.
- Hersbach, H., de Rosnay, P., Bell, B., Schepers, D., Simmons, A., Soci, C., ... Zuo, H. (2018). *Operational global reanalysis: progress, future directions and synergies with NWP*. ERA Report, ECMWF, UK. <https://doi:10.21957/tkic6g3wm>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horanyi, A., Muñoz-Sabater, J., ... Thepaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *1–51*. <https://doi.org/10.1002/qj.3803>
- Hirpa, F. A., Salamon, P., Alfieri, L., Pozo, J. T., Zsoter, E., & Pappenberger, F. (2016). The effect of reference climatology on

- global flood forecasting. *Journal of Hydrometeorology*, 17, 1131–1145. <https://doi.org/10.1175/JHM-D-15-0044.1>
- Hirpa, F. A., Salamon, P., Beck, H. E., Lorini, V., Alfieri, L., Zsoter, E., & Dadson, S. J. (2018). Calibration of the global flood awareness system (GloFAS) using daily streamflow data. *Journal of Hydrology*, 566, 595–606. <https://doi.org/10.1016/j.jhydrol.2018.09.052>
- Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society: Series B: Methodological*, 52, 105–124.
- Hsu, W. R., & Murphy, A. H. (1986). The attributes diagram: A geometrical framework for assessing the quality of probability forecast. *International Journal of Forecasting*, 2, 285–293.
- Kjeldsen, T. R., Lamb, R., & Blazkova, S. D. (2014). Uncertainty in flood frequency analysis. In K. Beven & J. Hall (Eds.), *Applied uncertainty analysis for flood frequency analysis*. London: Imperial College Press.
- Leutbecher, M., Lock, S.-J., Ollinaho, P., Lang, S. T. K., Balsamo, G., Bechtold, P., ... Smolarkiewicz, P. K. (2017). Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Quarterly Journal of the Royal Meteorological Society*, 143, 2315–2339. <https://doi.org/10.1002/qj.3094>
- Murphy, A. H. (1973). A new vector partition of the probability score. *Journal of Applied Meteorology*, 12(4), 595–600. [https://doi.org/10.1175/1520-0450\(1973\)012<0595:ANVPOT>2.0.CO;2](https://doi.org/10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2)
- Papalexiou, S. M., & Koutsogiannis, D. (2013). Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resources Research*, 49, 187–201. <https://doi.org/10.1029/2012WR012557>
- Stephens, E., Day, J. J., Pappenberger, F., & Cloke, H. (2015). Precipitation and floodiness. *Geophysical Research Letters*, 42, 10316–10323. <https://doi.org/10.1002/2015GL066779>
- Thielen, J., Bartholmes, J., Ramos, M.-H., & de Roo, A. (2009). The European flood alert system—Part 1: Concept and development. *Hydrology and Earth System Sciences*, 13, 125–140. <https://doi.org/10.5194/hess-13-125-2009>
- Towner, J., Cloke, H. L., Zsoter, E., Flamig, Z., Hoch, J. M., Bazo, J., ... 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. <https://doi.org/10.5194/hess-23-3057-2019>
- Tsonevsky, I., Doswell, C. A., & Brooks, H. E. (2018). Early warnings of severe convection using the ECMWF extreme forecast index. *Weather and Forecasting*, 33, 857–871. <https://doi.org/10.1175/WAF-D-18-0030.1>
- van der Knijff, J. M., Younis, J., & 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, 189–212. <https://doi.org/10.1080/13658810802549154>
- Verkade, J. S., Brown, J. D., Reggiani, P., & Weerts, A. H. (2013). Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. *Journal of Hydrology*, 501, 73–91. <https://doi.org/10.1016/j.jhydrol.2013.07.039>
- Vitart, F. (2014). Evolution of ECMWF sub-seasonal forecast scores. *Quarterly Journal of the Royal Meteorological Society*, 140, 1889–1899. <https://doi.org/10.1002/qj.2256>
- Wentao, L., Qingyun, D., Chiyuan, M., Aizhong, Y., Wei, G., & Zhenhua, D. (2017). A review on statistical postprocessing methods for hydrometeorological ensemble forecasting. *Wiley Interdisciplinary Reviews: Water*, 4, e1246. <https://doi.org/10.1002/wat2.1246>
- WMO Guidelines on the calculation of Climate Normals, WMO-No. 1203 (2017).
- 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(3), e1432. <https://doi.org/10.1002/wat2.1432>
- Yuan, X., Wood, E. F., & F. E. (2012). Downscaling precipitation or bias-correcting streamflow? Some implications for coupled general circulation model (CGCM)-based ensemble seasonal hydrologic forecast. *Water Resources Research*, 48, W12519. <https://doi.org/10.1029/2012WR012256>
- Zsoter, E., Pappenberger, F., & Richardson, D. (2014). Sensitivity of model climate to sampling configurations and the impact on the extreme forecast index. *Meteorological Applications*, 22, 236–247. <https://doi.org/10.1002/met.1447>
- Zsoter, E., Pappenberger, F., Smith, P., Emerton, R., Dutra, E., Wetterhall, F., ... Balsamo, G. (2016). Building a multi-model flood prediction system with the TIGGE archive. *Journal of Hydrometeorology*, 17, 2923–2940. <https://doi.org/10.1175/JHM-D-15-0130.1>
- Zsoter, E., Cloke, H., Stephens, E., de Rosnay, P., Muñoz-Sabater, J., Prudhomme, C., & Pappenberger, F. (2019). How well do operational numerical weather prediction setups represent hydrology. *Journal of Hydrometeorology*, 14, 1533–1552. <https://doi.org/10.1175/JHM-D-18-0086.1>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Zsoter E, Prudhomme C, Stephens E, Pappenberger F, Cloke H. Using ensemble reforecasts to generate flood thresholds for improved global flood forecasting. *J Flood Risk Management*. 2020;13:e12658. <https://doi.org/10.1111/jfr3.12658>

A5: Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0

This paper presents a co-author contribution arising through collaboration during this PhD, and has the following reference:

Emerton, R., E. Zsoter, L. Arnal, H L. Cloke, D Muraro, C. Prudhomme, E. M. Stephens, P. Salamon and F. Pappenberger, 2018: Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0, *Geosci. Model Dev.*, 11, 3327–3346, doi:10.5194/gmd-11-3327-2018

Geosci. Model Dev., 11, 3327–3346, 2018
 https://doi.org/10.5194/gmd-11-3327-2018
 © Author(s) 2018. This work is distributed under
 the Creative Commons Attribution 4.0 License.



Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0

Rebecca Emerton^{1,2}, Ervin Zsoter^{2,1}, Louise Arnal^{1,2}, Hannah L. Cloke^{1,3}, Davide Muraro⁶, Christel Prudhomme^{2,4,5}, Elisabeth M. Stephens¹, Peter Salamon⁷, and Florian Pappenberger²

¹Department of Geography & Environmental Science, University of Reading, Reading, UK

²European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK

³Department of Earth Sciences, Uppsala University, Uppsala, Sweden

⁴Centre for Ecology and Hydrology (CEH), Wallingford, UK

⁵Department of Geography and Environment, University of Loughborough, Loughborough, UK

⁶Image Recognition Integrated Systems (IRIS), Ispra, Italy

⁷European Commission, Joint Research Centre (JRC), Ispra, Italy

Correspondence: Rebecca Emerton (r.e.emerton@pgr.reading.ac.uk)

Received: 27 April 2018 – Discussion started: 14 May 2018

Revised: 7 August 2018 – Accepted: 9 August 2018 – Published: 21 August 2018

Abstract. Global overviews of upcoming flood and drought events are key for many applications, including disaster risk reduction initiatives. Seasonal forecasts are designed to provide early indications of such events weeks or even months in advance, but seasonal forecasts for hydrological variables at large or global scales are few and far between. Here, we present the first operational global-scale seasonal hydro-meteorological forecasting system: GloFAS-Seasonal. Developed as an extension of the Global Flood Awareness System (GloFAS), GloFAS-Seasonal couples seasonal meteorological forecasts from ECMWF with a hydrological model to provide openly available probabilistic forecasts of river flow out to 4 months ahead for the global river network. This system has potential benefits not only for disaster risk reduction through early awareness of floods and droughts, but also for water-related sectors such as agriculture and water resources management, in particular for regions where no other forecasting system exists. We describe the key hydro-meteorological components and computational framework of GloFAS-Seasonal, alongside the forecast products available, before discussing initial evaluation results and next steps.

1 Introduction

Seasonal meteorological forecasts simulate the evolution of the atmosphere over the coming months. They are designed to provide an early indication of the likelihood that a given variable, for example precipitation or temperature, will differ from normal conditions weeks or months ahead. Will a particular region be warmer or cooler than normal during the next summer? Or will a river have higher or lower flow than normal next winter? Seasonal forecasts of river flow have the potential to benefit many water-related sectors from agriculture and water resources management to disaster risk reduction and humanitarian aid through earlier indications of floods or droughts.

Many operational forecasting centres produce long-range (seasonal) global forecasts of meteorological variables, such as precipitation (Weisheimer and Palmer, 2014). However, at present, operational seasonal forecasts of hydrological variables, particularly for large or global scales, are few and far between. A number of continental-scale seasonal hydro-meteorological forecasting systems have begun to emerge around the globe over the past decade (Yuan et al., 2015a), using seasonal meteorological forecasts as input to hydrological models to produce forecasts of hydrological variables. These include the European Flood Awareness System (EFAS; Arnal et al., 2018; Cloke et al., 2013), the European Service for Water Indicators in Climate Change Adapta-

Published by Copernicus Publications on behalf of the European Geosciences Union.

tion (SWICCA; Copernicus, 2018b), the Australian Government Bureau of Meteorology Seasonal Streamflow Forecasts (Bennett et al., 2017; BoM, 2018), and the USA's National Hydrologic Ensemble Forecast Service (HEFS; Demargne et al., 2014; Emerton et al., 2016). There are also various ongoing research efforts using seasonal hydro-meteorological forecasting systems for forecast applications and research purposes at regional (Bell et al., 2017; Bennett et al., 2016; Crochemore et al., 2016; Meißner et al., 2017; Mo et al., 2014; Prudhomme et al., 2017; Wood et al., 2002, 2005; Yuan et al., 2013) and global (Candogan Yossef et al., 2017; Yuan et al., 2015b) scales. In addition to the ongoing research into improved seasonal hydro-meteorological forecasts at the global scale, an operational system providing consistent global-scale seasonal forecasts of hydrological variables could be of great benefit in regions where no other forecasting system exists and to organisations operating at the global scale (Coughlan De Perez et al., 2017).

Often, in the absence of hydrological forecasts, seasonal precipitation forecasts are used as a proxy for flooding. It has been shown that forecasts of seasonal total rainfall, the most often used seasonal precipitation forecasts, are not necessarily a good indicator of seasonal floodiness (Stephens et al., 2015), and other measures of rainfall patterns, or seasonal hydrological forecasts, would be better indicators of potential flood hazard (Coughlan De Perez et al., 2017).

While it seems a natural next step to produce global-scale seasonal hydro-meteorological forecasts, this is not a simple task, not only due to the complexities of geographical variations in rainfall–run-off processes and river regimes across the globe, but also due to the computing resources required and huge volumes of data that must be efficiently processed and stored and the challenge of effectively communicating forecasts for the entire globe. Indeed, global-scale forecasting for medium-range timescales has only become possible in recent years due to the integration of meteorological and hydrological modelling capabilities, improvements in data, satellite observations, and land-surface hydrology modelling, and increased resources and computer power (Emerton et al., 2016). In addition to continued improvements in computing capabilities, the recent move towards the development of coupled atmosphere–ocean–land models means that it is now becoming possible to produce seasonal hydro-meteorological forecasts for the global river network.

Despite the chaotic nature of the atmosphere (Lorenz, 1963), which introduces a limit of predictability (generally accepted to be ~ 2 weeks), seasonal predictions are possible as they rely on components that vary on longer timescales and are themselves predictable to an extent. This “second type predictability” (Lorenz, 1993) for seasonal river flow forecasts comes from the initial conditions and large-scale modes of climate variability. The most prominent pattern of climate variability is the El Niño–Southern Oscillation (ENSO; McPhaden et al., 2006), which is known to affect river flow and flooding across the globe (Chiew and McMa-

hon, 2002; Emerton et al., 2017; Guimarães Nobre et al., 2017; Ward et al., 2014a, b, 2016). Other teleconnections also influence river flow in various regions of the globe, such as the North Atlantic Oscillation (NAO), Southern Oscillation (SOI), Indian Ocean Dipole (IOD), and Pacific Decadal Oscillation (PDO), and contribute to the seasonal predictability of hydrologic variables (Yuan et al., 2015a). Coupled atmosphere–ocean–land models are key in representing these large-scale modes of variability in order to produce seasonal hydro-meteorological forecasts.

This motivates the development of an operational global-scale seasonal hydro-meteorological forecasting system as an extension of the Global Flood Awareness System (GloFAS; Alfieri et al., 2013), with openly available forecast products. GloFAS is developed by the European Centre for Medium-Range Weather Forecasts (ECMWF) and the European Commission Joint Research Centre (JRC) and has been producing probabilistic flood forecasts out to 30 days for the entire globe since 2012. In 2016, work began in collaboration with the University of Reading to implement a seasonal outlook in GloFAS, aiming to provide forecasts of both high and low river flow for the global river network up to several months in advance. On 10 November 2017, the first GloFAS seasonal river flow forecast was released. This paper introduces the modelling system, its implementation, and the available forecast products and provides an initial evaluation of the potential usefulness and reliability of the forecasts.

2 Implementation

The GloFAS seasonal outlooks are produced by driving a hydrological river routing model with meteorological forecasts from ECMWF. The forecasts are run operationally on the ECMWF computing facilities. This section provides an overview of the computing facilities, introduces the key hydro-meteorological components of the modelling platform (the meteorological forecast input, hydrological model, and reference climatology), and describes the computational framework of GloFAS-Seasonal.

2.1 ECMWF High-Performance Computing Facility

ECMWF's current High-Performance Computing Facility (HPCF) has been in operation since June 2016 and is used for both forecast production and research activities. The HPCF comprises two identical Cray XC40 supercomputers, each of which is self-sufficient with their own storage and each with equal access to the storage of the other. Each Cray XC40 consists of 20 cabinets of compute nodes and 13 storage nodes. One compute node has two Intel Broadwell processors, each with 18 cores, giving 192 nodes (6912 cores) per cabinet. The Cray Aries interconnect is used to connect the processing power. The majority of the nodes of the HPCF are run using the high-performance Cray Linux Environment, a stripped-

down version of Linux, as reducing the number of operating system tasks is critical for providing a highly scalable environment.

In terms of storage, each Cray XC40 has ~ 10 PB of storage, and the data handling system (DHS) also comprises two main applications: the Meteorological Archive and Retrieval System (MARS), which stores and provides access to meteorological data collected or produced by ECMWF, and ECFS, which stores data that are not suitable for storing on MARS. The DHS holds over 210 PB of primary data, and the archive increases by ~ 233 TB per day. The reader is referred to the ECMWF website at <https://www.ecmwf.int/> for further information on the HPCF and DHS.

In addition to the Cray XC40s, the ECMWF computing facility also includes four Linux clusters consisting of 60 servers and 1 PB of storage. The Linux clusters are currently used to run the river routing model used in GloFAS and to produce the forecast products, while the meteorological forcing and ERA5 reanalysis are produced on the HPCF. All data related to GloFAS-Seasonal are stored on the MARS and ECFS archives.

2.2 Hydro-meteorological components

2.2.1 Meteorological forcing

The first model component of the seasonal outlook is the meteorological forecast input from the ECMWF Integrated Forecast System (IFS, cycle 43r1; ECMWF, 2018b). GloFAS-Seasonal makes use of SEAS5, which is the latest version of ECMWF's long-range ensemble forecasting system made operational in November 2017 (ECMWF, 2017a; Stockdale et al., 2018). SEAS5 consists of 51 ensemble members (50 perturbed members and 1 unperturbed control member) and has a horizontal resolution of ~ 36 km (T_{CO319}). The system, which comprises a data assimilation system and a global circulation model, is run once a month, producing forecasts out to 7 months ahead. Initial pre-implementation testing of SEAS5 has suggested that in comparison to the previous version (System 4), SEAS5 better simulates sea surface temperatures (SSTs) in the Pacific Ocean, leading to improved forecasts of the El Niño–Southern Oscillation (ENSO; Stockdale et al., 2018), which is closely linked to river flow across the globe and can provide added predictability.

SEAS5 is a configuration of the ECMWF IFS (cycle 43r1), including atmosphere–ocean coupling to the NEMO ocean model. SEAS5 is run operationally on the HPCF. Each ensemble member is a complex, HPC-intensive, massively parallel code written in Fortran (version F90). In addition, further complex scripting systems are required to control, prepare, run, post-process, and archive all IFS forecasts. The data assimilation systems used to prepare the initial conditions for the forecasts also make use of Fortran and run on

the HPCF. For further information, the reader is referred to the IFS documentation (ECMWF, 2018b).

2.2.2 Land surface component

Within the IFS, which includes SEAS5, the Hydrology Tiled ECMWF Scheme of Surface Exchanges over Land, HTESSEL (Balsamo et al., 2011), is used to compute the land surface response to atmospheric forcing. HTESSEL simulates the evolution of soil temperature, moisture content, and snowpack conditions through the forecast horizon to produce a corresponding forecast of surface and subsurface run-off. This component allows for each grid box to be divided into tiles, with up to six tiles per grid box (bare ground, low and high vegetation, intercepted water, and shaded and exposed snow) describing the land surface. For a given precipitation, the scheme distributes the water as surface run-off and drainage, with dependencies on orography and soil texture. An interception layer accumulates precipitation until saturation is reached, with the remaining precipitation partitioned between surface run-off and infiltration. HTESSEL also accounts for frozen soil, redirecting the rainfall and snowmelt to surface run-off when the uppermost soil layer is frozen, and incorporates a snow scheme. Four soil layers are used to describe the vertical transfer of water and energy, with subsurface water fluxes determined by Darcy's law, and each layer has a sink to account for root extraction in vegetated areas. A detailed description of the hydrology of HTESSEL is provided by Balsamo et al. (2011).

HTESSEL comprises a Fortran library of $\sim 20\,000$ lines of code, using both F77 and F90 Fortran versions, and is implemented modularly. While HTESSEL can be run on diverse architectures from a workstation PC to the HPCF, operationally it is run on the HPCF.

2.2.3 River routing model

As HTESSEL does not simulate water fluxes through the river network, Lisflood (Van Der Knijff et al., 2010), driven by the surface and subsurface run-off output from HTESSEL interpolated to the 0.1° (~ 10 km) spatial resolution of Lisflood is used to simulate the groundwater (subsurface water storage and transport) processes and routing of the water through the river network. The initial conditions used to start the Lisflood model are taken from the ERA5-R river flow reanalysis (see Sect. 2.2.4).

Lisflood is a spatially distributed hydrological model, including a 1-D channel routing model. Groundwater processes are modelled using two linear reservoirs, the upper zone representing a quick run-off component, including subsurface flow through soil macropores and fast groundwater, and the lower zone representing a slow groundwater component fed by percolation from the upper zone. The routing of surface run-off to the outlet of each grid cell, and the routing of run-off produced by every grid cell from the surface, upper,

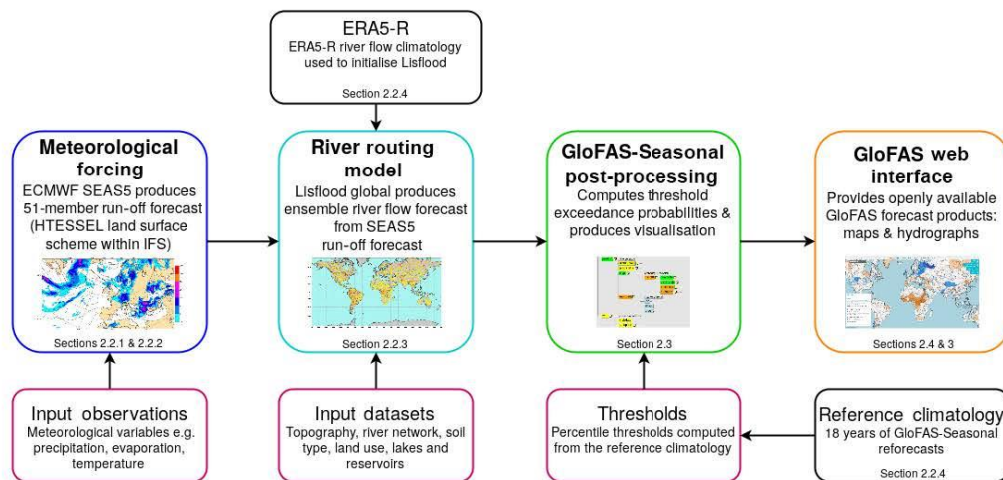


Figure 1. Flowchart depicting the key GloFAS-Seasonal forecasting system components.

and lower groundwater zones through the river network, is done using a four-point implicit finite-difference solution of the kinematic wave equations (Chow et al., 2010). The river network used is that of HydroSHEDS (Lehner et al., 2008), again interpolated to a 0.1° spatial resolution using the approach of Fekete et al. (2001). For a detailed account of the Lisflood model set-up within GloFAS, the reader is referred to Alfieri et al. (2013).

Lisflood is implemented using a combination of PCRaster GIS and Python and is currently run operationally on the Linux cluster at ECMWF.

2.2.4 Generation of reforecasts and reference climatology

In order to generate a reference climatology for GloFAS-Seasonal, the latest of ECMWF's reanalysis products, ERA5, was used. Reanalysis datasets combine historical observations of the atmosphere, ocean, and land surface with a data assimilation system; global models are used to “fill in the gaps” and produce consistent global best estimates of the atmosphere, ocean, and land state. ERA5 represents the current state of the art in terms of reanalysis datasets, providing a much higher spatial and temporal resolution (30 km, hourly) compared to ERA-Interim (79 km, 3-hourly) and better representations of precipitation, evaporation, and soil moisture (ECMWF, 2017b). In order to produce a river flow reanalysis (ERA5-R) for the global river network, the ERA5 surface and subsurface run-off variables were interpolated to 0.1° (~ 10 km) resolution and used as input to the Lisflood model (see Sect. 2.2.3). ERA5 is currently still in production, and while it will cover the period from 1950 to present

when completed, the full dataset will not be available until 2019. ERA5 is being produced in three “streams” in parallel; at the time of producing the ERA5-R reanalysis, 18 years of ERA5 data were available across the three streams (1990–1992, 2000–2007, and 2010–2016). In addition to the historical climatology, ERA5 is also produced in near real time, with a delay of just ~ 3 days, allowing its use as initial conditions for the river routing component of the GloFAS-Seasonal forecasts. The ERA5-R reanalysis is thus updated every month prior to producing the forecast. Figure 2 provides an overview of all datasets used in and produced for the development of GloFAS-Seasonal.

Once the ERA5-R reanalysis was obtained, a set of GloFAS-Seasonal reforecasts was produced. From the 25-ensemble-member SEAS5 reforecasts produced by ECMWF, the surface and subsurface run-off variables were used to drive the Lisflood model with initial conditions from ERA5-R. This generated 18 years of seasonal river flow reforecasts (one forecast per month out to 4 months of lead time, with 25 ensemble members at 0.1° resolution). It is the weekly averaged river flow from this reforecast dataset which is used as a reference climatology, including to calculate the high and low flow thresholds used in the real-time forecasts (described in Sect. 2.3).

2.3 GloFAS-Seasonal computational framework

The GloFAS-Seasonal real-time forecasts are implemented and run operationally on the ECMWF computing facilities using ecFlow (Bahra, 2011; ECMWF, 2012), an ECMWF work package used to run large numbers of programmes with dependencies on each other and on time. An ecFlow suite

is a collection of tasks and scheduling instructions with a user interface allowing for the interaction and monitoring of the suite, the code behind it, and the output. The GloFAS-Seasonal suite is run once per month and is used to retrieve the raw SEAS5 forecast data. It runs this through Lisflood and produces the final forecast products and visualisations using the newly developed GloFAS-Seasonal post-processing code.

The GloFAS-Seasonal suite performs tasks (detailed below) such as retrieving data, running Lisflood, computing weekly averages and forecast probabilities from the raw Lisflood river flow forecast data, and producing maps and hydrographs for the interface. It is primarily written in Python (version 2.7), with some elements written in R (version 3.1) and shell scripts incorporating climate data operators (CDOs). The code was developed and tested on OpenSUSE Leap 42 systems.

When a new SEAS5 forecast becomes available (typically on the 5th of the month at 00:00 UTC), the GloFAS-Seasonal ecFlow suite is automatically deployed. The structure of and tasks within the ecFlow suite are shown in Fig. 3. Each “task” represents one script from the GloFAS-Seasonal code. The suite first retrieves the latest raw SEAS5 forecast surface and subsurface variables for all 51 ensemble members (*stagefc* and *gefc* tasks), alongside the river flow reference climatology (see Sect. 2.2.4) for the corresponding month of the forecast (*copywb* task). The Lisflood river routing model (described in Sect. 2.2.3) is then run for each of the 51 ensemble members (*lisflood* task). Lisflood is initialised using the ERA5-R river flow reanalysis (see Sect. 2.2.4) and driven with the SEAS5 surface and subsurface run-off forecast to produce the 4-month ensemble river flow forecast at a daily time step, from which the weekly averaged ensemble river flow forecast is obtained (*average* task). The weekly averages are computed for every Monday–Sunday starting from the first Monday of each month so that the weekly averages correspond from one forecast to the next. While SEAS5 provides forecasts out to 7 months ahead, the first version of GloFAS-Seasonal uses only the first 4 months. This is in order to reduce the data volumes required and to allow for the assessment of the forecast skill out to 4 months ahead before possible extension of the forecasts out to 7 months ahead in the future.

Once the weekly averaging is complete, the *forecast product* section of the suite is deployed, which post-processes the raw forecast output to produce the final forecast products displayed on the web interface. The code behind the *forecast product* section is provided in the Supplement. For a full description of the forecast products, including examples, see Sect. 3. The suite computes the full forecast distribution (*distribution* task), followed by the probability of exceedance for each week of the forecast and for every grid point (*probability* task) based on the number of ensemble members exceeding the high flow threshold or falling below the low flow threshold. The high and low flow thresholds are defined

as the 80th and 20th percentiles of the reference climatology for the week of the year corresponding to the forecast week to use thresholds based on time of year of the forecast. From these weekly exceedance probabilities, the maximum probability of exceedance across the 4-month forecast horizon is calculated for each grid point (*maxprob* task). Basin-averaged maximum probabilities are also produced (*basinprob* task) by calculating the mean maximum probability of exceedance across every grid point at which the upstream area exceeds 1500 km² in each of the 306 major world river basins used in GloFAS-Seasonal (see Sect. 3.1). A minimum upstream area of 1500 km² is chosen, as the current resolution of the global model is such that reliable forecasts for very small rivers are not feasible.

These probabilities are used to produce the forecast visualisation for the web interface (Sect. 3). Firstly, the *map* task produces colour-coded maps of both the river network, again for grid points at which the upstream area exceeds 1500 km², and the major world river basins. The *reppoint* task then produces an ensemble hydrograph and persistence diagrams for a subset of grid points (the “reporting points”) across the globe. Further details on the location of reporting points are given in Sect. 3.3. Finally, the *web* task collates and subsequently transfers all data required for the web interface.

This process, from the time a new SEAS5 forecast becomes available, takes ~ 4 h on average to complete, with up to 10 tasks running in parallel (for example, running Lisflood for 10 ensemble members at the same time). It is possible to speed up this process by running more ensemble members in parallel; however, the speed is sufficient so that it is not necessary to use further resources to produce the forecast more quickly. GloFAS-Seasonal forecast products are typically produced by the 5th of the month at 05:00 UTC and made available via the web interface on the 10th of the month at 01:00 UTC. This is the earliest that the GloFAS-Seasonal forecasts can be provided publicly under the Copernicus licence agreement. Data are automatically archived at ECMWF as the suite runs in real time; ~ 285 GB of data from each SEAS5 forecast are used as input for GloFAS-Seasonal. Each GloFAS-Seasonal forecast run produces an additional ~ 1.8 TB of data and makes use of the ~ 18 TB reference climatology.

2.4 GloFAS web interface

The GloFAS website is based on a user-centred design (UCD), meaning that user needs are core to the design principles (ISO13407). The website uses Web 2.0 concepts such as simplicity, joy of use, and usability that are synonymous with engaging users. It is a rich internet application (RIA) aiming to provide the same level of interactivity and responsiveness as desktop applications. The website is designed for those engaged in flood forecasting and water resources, as users can browse various aspects of the current forecast or past forecasts in a simple and intuitive way, with spatially distributed

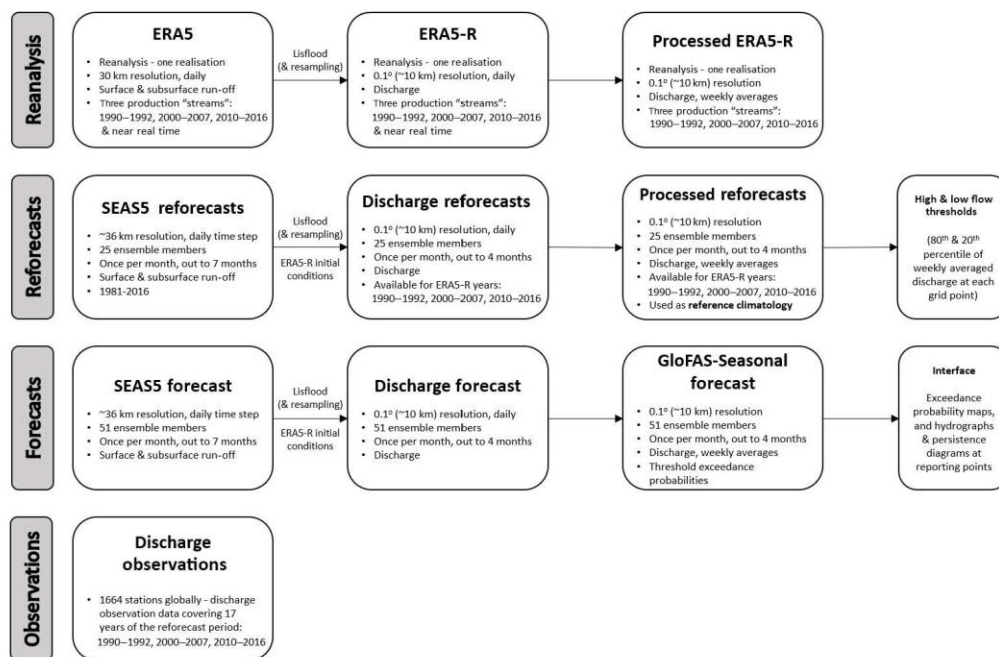


Figure 2. All datasets used and produced for GloFAS-Seasonal, including reanalysis, reforecasts, real-time forecasts, and observations.

information. Map layers containing different information, e.g. flood probabilities for different flood severities, precipitation forecasts, and seasonal outlooks, can be activated and the user can also choose to overlay other information such as land use, urban areas, or flood hazard maps. The interface consists of three principal modules: MapServer, GloFAS Web Map Service Time, and the Forecast Viewer. These are outlined below.

2.4.1 MapServer

MapServer (Open Source Geospatial Foundation, 2016) is an open source development environment for building spatially enabled internet applications developed by the University of Minnesota. MapServer has built-in functionality to support industry standard data formats and spatial databases, which is significant to this project, and the support of popular Open Geospatial Consortium (OGC) standards including WMS. In order to exploit the potential of asynchronous data transfer between server and client, the GloFAS raster data have to be divided into a grid of adequate dimensions and an optimal scale sequence.

2.4.2 GloFAS Web Map Service Time

The OpenGIS Web Map Service (WMS) is a standard protocol for serving geo-referenced map images over the internet. A web map service time (WMS-T) is a web service that produces maps in several raster formats or in vector format that may come simultaneously from multiple remote and heterogeneous sources. A WMS server can provide support to temporal requests (WMS-T) by providing a TIME parameter with a time value in the request.

The WMS specification (OGC, 2015) describes three HTTP requests; *GetCapabilities*, *GetMap*, and *GetFeatureInfo*. *GetCapabilities* returns an XML document describing the map layers available and the server's capabilities (i.e. the image formats, projections, and geographic bounds of the server). *GetMap* returns a raster map image. The request arguments, such as the layer ID and image format, should match those listed as available in the *GetCapabilities* return document. *GetFeatureInfo* is optional and is designed to provide WMS clients with more information about features in the map images that were returned by earlier *GetMap* requests. The response should contain data relating to the features nearest to an image coordinate specified in the *GetFeatureInfo* request. The structure of the data returned is not defined in the specification and is left up to the WMS server

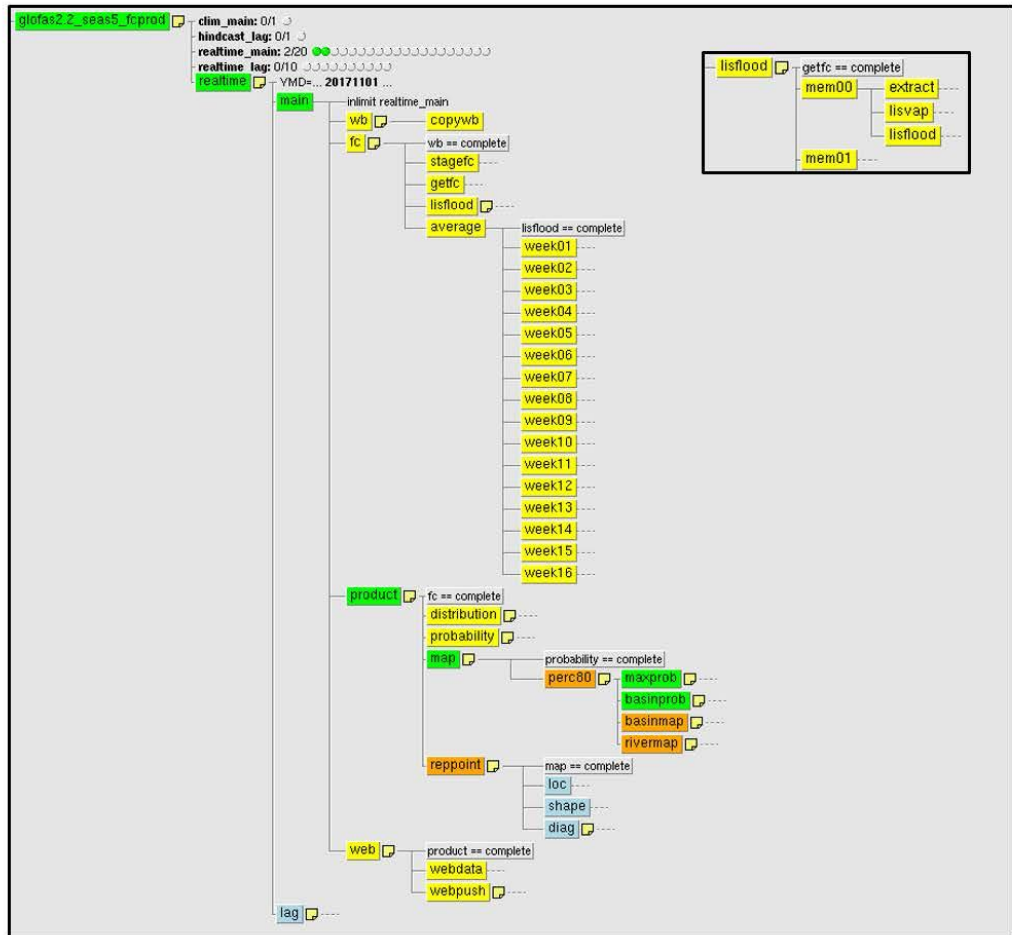


Figure 3. The GloFAS-Seasonal ecFlow suite. The inset image shows the sub-tasks within the Lisflood task for 1 of the 51 ensemble members. Colours indicate the status of each task. Yellow: complete, green: active, orange: suspended, pale blue: waiting, turquoise (not shown): queued, and red (not shown): aborted or failed. Grey boxes indicate dependencies; for example, “lisflood = complete” indicates that the Lisflood task and all Lisflood sub-tasks must have successfully completed in order for the average task to run.

implementation. The GloFAS WMS-T (GloFAS, 2018b) can be freely used, allowing access to the GloFAS layers in any GIS environment, such as QGIS (QGIS Development Team, 2017) or ArcMAP (Environmental Systems Research Institute, 2018). The user manual for the GloFAS WMS-T is available via the GloFAS website (GloFAS, 2018a).

2.4.3 Forecast Viewer

The GloFAS Forecast Viewer is based on the model view controller (MVC) architectural pattern used in software engineering. The pattern isolates “domain logic” (the applica-

tion logic for the user) from input and presentation (user interface, UI), permitting the independent development, testing, and maintenance of each. A fundamental part of this is the AJAX (asynchronous JavaScript and XML) technology used to enhance user-friendly interfaces for web mapping applications. AJAX technologies have a number of benefits; the essential one is removing the need to reload and refresh the whole page after every event. Careful application design and component selection results in a measurably smaller web server load in geodata rendering and publishing, as there is no need to link and send the whole html document, just the relevant part that needs to be changed.

GloFAS uses OpenLayers (OpenLayers, 2018) as a WMS client. OpenLayers is a JavaScript-based web mapping toolkit designed to make it easy to put a dynamic map on any web page. It does not depend on the server technology and can display a set of vector data, such as points, with aerial photographs as backdrop maps from different sources. Closely coupled to the map widget is a layer manager that controls which layers are displayed with facilities for adding, removing, and modifying layers. The new layers associated with GloFAS-Seasonal are described in the following section.

3 Forecast products

The GloFAS seasonal outlook is provided as three new forecast layers in the GloFAS Forecast Viewer: the basin overview, river network, and reporting point layers. Each of the three layers represents a different forecast product described in the following sections. Information on each of the layers is also provided for end users of the forecasts under the dedicated “Seasonal Outlook” page of the GloFAS website.

3.1 Basin overview layer

The first GloFAS seasonal outlook product is designed to provide a quick global overview of areas that are likely to experience unusually high or low river flow over the coming 4 months. The “basin overview” layer displays a map of 306 major world river basins colour coded according to the maximum probability of exceeding the high (blue) or low (orange) flow thresholds (the 80th and 20th percentiles of the reference climatology, respectively) during the 4-month forecast horizon. This value is calculated for each river basin by taking the average of the maximum exceedance probabilities at each grid cell within the basin (using only river pixels with an upstream area $> 1500 \text{ km}^2$). The three different shades of orange–blue indicate the probability: dark ($> 90\%$), medium ($75\%–90\%$), and light ($50\%–75\%$). Basins that remain white are those in which the probability of unusually high or low flow does not exceed 50% during the 4-month forecast horizon. An example is shown in Fig. 4.

As mentioned in Sect. 2.2.3, the Lisflood river network is based on HydroSHEDS (Lehner et al., 2008). In order to generate the river basins used in GloFAS-Seasonal, the corresponding HydroBASINS (Lehner and Grill, 2013) data were used. HydroBASINS consists of a suite of polygon layers depicting watershed boundaries at the global scale. These watersheds were manually merged using QGIS (QGIS Development Team, 2017) to create a global polygon layer of major river basins based on the river network used in the model.

3.2 River network layer

The second map layer provides similar information at the sub-basin scale by colour-coding the entire model river net-

work according to the maximum exceedance probability during the 4-month forecast horizon. This allows the user to zoom in to their region of interest and view the forecast maximum exceedance probabilities in more detail. Again, only river pixels with an upstream area $> 1500 \text{ km}^2$ are shown. The same colour scheme is used for both the basin overview and river network layers, with blue indicating high flow (exceeding the 80th percentile), orange low flow (falling below the 20th percentile), and darker colours indicating higher probabilities. In the river network layer, additional colours also represent areas where the forecast does not exceed 50% probability of exceeding either the high or low flow threshold (light grey) and where the river pixel lies in a climatologically arid area such that the forecast probability cannot be defined (darker grey–brown). Examples of the river network layer can be seen in both Fig. 4 (globally) and Fig. 5 (zoomed in).

3.3 Reporting points layer

In addition to the two summary map layers, reporting points are provided at both static and dynamic locations throughout the global river network, providing additional forecast information: an ensemble hydrograph and a persistence diagram.

Static points originally consisted of a selection of gauged river stations included in the Global Runoff Data Centre (GRDC; BfG, 2017); this set of points has since been expanded to further include points at locations of particular interest to GloFAS partners. There are now ~ 2200 static reporting points in the GloFAS interface.

Dynamic points are generated to provide the additional forecast information throughout the global river network, including river reaches for which there are no static points. These points are obtained for every new forecast based on a set of selection criteria adapted from the GloFAS flood forecast dynamic point selection criteria (Alfieri et al., 2013).

- The maximum probability of high (low) river flow (exceeding or falling below) the 80th (20th) percentile of the reference climatology) during the 4-month forecast horizon must be $\geq 50\%$ for at least five contiguous pixels of the river network.
- The upstream area of the selected point must be $\geq 4000 \text{ km}^2$.
- Dynamic reporting points are generated starting from the most downstream river pixel complying with the previous two selection criteria. A new reporting point is then generated every 300 km upstream along the river network, unless a static reporting point already exists within a short distance of the new dynamic point or the forecasts further upstream no longer comply with the previous two criteria.

Reporting points are displayed as black circles in the “reporting points” seasonal outlook layer. An example is shown in

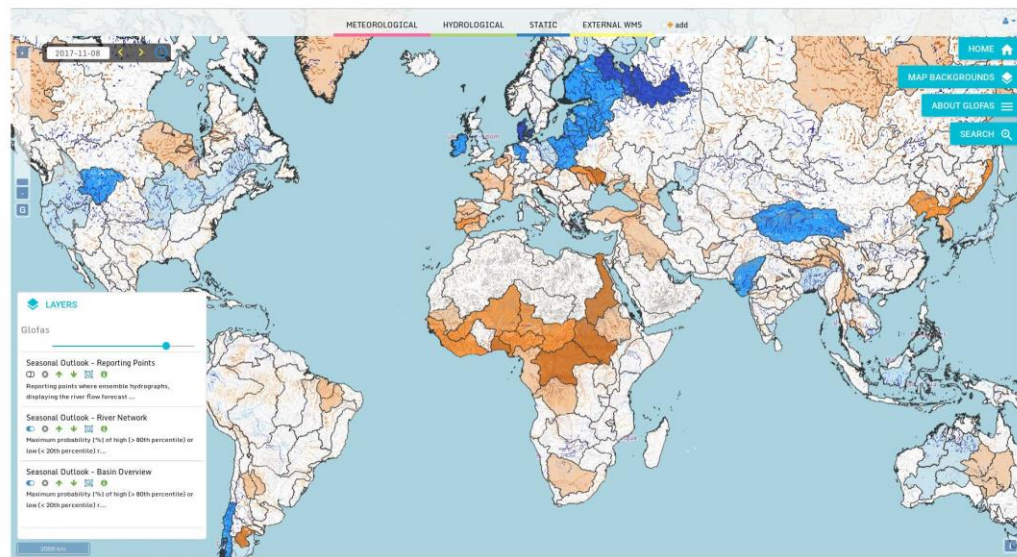


Figure 4. Example screenshot of the seasonal outlook layers in the GloFAS web interface. Shown here are both the “basin overview” layer and “river network” layer, both indicating the maximum probability of unusually high (blue) or low (orange) river flow during the 4-month forecast horizon. The darker the colour, the higher the probability: darkest shading indicates > 90 % probability, medium shading indicates 75 %–90 % probability, and light shading indicates 50 %–75 % probability. A white basin or light grey river pixel indicates that the forecast does not exceed 50 % probability of high or low flow during the forecast horizon. Legends providing this information are available for each layer by clicking on the green “i” next to the layer toggle (shown at the bottom left in this example).

Fig. 5. Clicking on a reporting point brings up a new window containing a hydrograph and persistence diagram alongside some basic information about the location, such as the latitude and longitude, and the upstream area of the point in the model river network. The number of dynamic reporting points can vary from one forecast to the next due to the criteria applied; for example, the March 2018 forecast included ~ 1600 dynamic points in addition to the static points, and thus ~ 3800 reporting points were available globally.

The ensemble hydrographs (also shown in Fig. 5) display a fan plot of the ensemble forecast of weekly averaged river flow out to 4 months, indicating the spread of the forecast and associated probabilities. Also shown are thresholds based on the reference climatology; the median and the 80th and 20th percentiles. These thresholds are displayed as a 3-week moving average of the weekly averaged river flow for the given threshold for the same months of the climatology as that of the forecast (i.e. a forecast for J–F–M–A also displays thresholds based on the reference climatology for J–F–M–A). This allows for a comparison of the forecast to typical and extreme conditions for the time of year.

Persistence diagrams (see Fig. 5) show the weekly probability of exceeding the high and low flow thresholds for the current forecast (bottom row) and previous three forecasts colour coded to match the probabilities indicated in the

map layers. These diagrams are provided in order to highlight the evolution of the forecast, which can indicate whether the forecast is progressing consistently or whether behaviour is variable from month to month.

4 Forecast evaluation

In this section, the GloFAS-Seasonal reforecasts are evaluated using historical river flow observations. Benchmarking a forecasting system is important to evaluate and understand the value of the system and in order to communicate the skill of the forecasts to end users (Pappenberger et al., 2015). This evaluation is designed to measure the ability of the forecasts to predict the correct category of an “event”, i.e. the ability of the forecast to predict that weekly averaged river flow will fall in the upper 80th or lower 20th percentile of climatology using a climatology of historical observations as a benchmark. This can be referred to as the potential usefulness of the forecasts and is of particular importance for decision-making purposes (Arnal et al., 2018). Another key aspect of probabilistic forecasts to consider is their reliability, which indicates the agreement between forecast probabilities and the observed frequency of events.

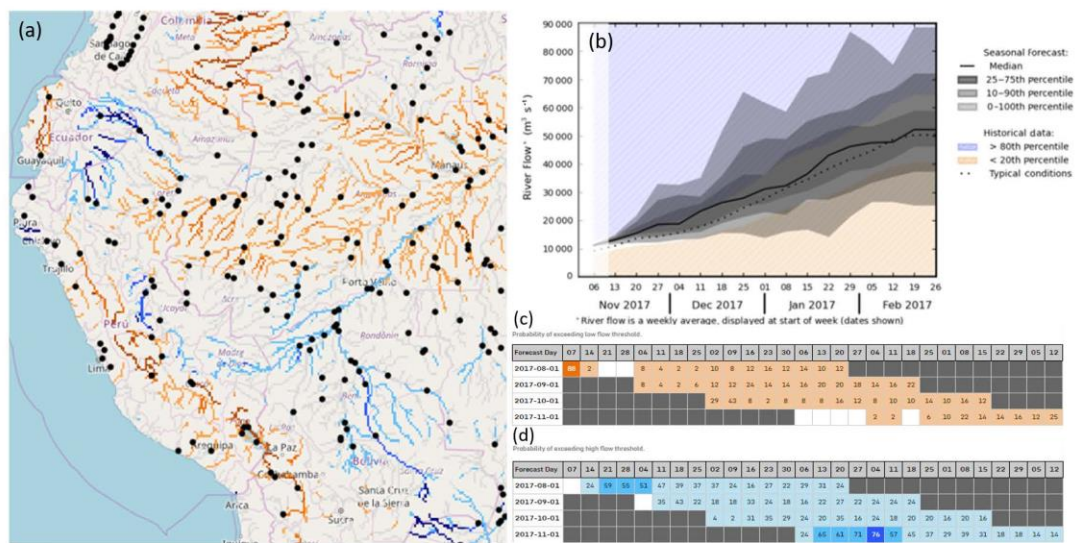


Figure 5. Example of the “reporting points” GloFAS seasonal outlook layer in the web interface (a). Black circles indicate the reporting points, which provide the ensemble hydrograph (b) and persistence diagrams for both low flow (c) and high flow (d). Also shown is an example section of the “river network” seasonal outlook layer indicating the maximum probability of high (blue) or low (orange) river flow during the 4-month forecast horizon. The darker the colour, the higher the probability.

The potential usefulness is assessed using the relative operating characteristic (ROC) curve, which is based on ratios of the proportion of events (the probability of detection, POD) and non-events (the false alarm rate, FAR) for which warnings were provided (Mason and Graham, 1999); in this case warnings are treated as forecasts of river flow exceeding the 80th or falling below the 20th percentile of the reference climatology (see Sect. 2.2.4). These ratios allow for the estimation of the probability that an event will be predicted.

For each week of the forecast (out to 16 weeks, corresponding to the forecasts provided via the interface; for example, the hydrograph shown in Fig. 5), the POD (Eq. 1) and FAR (Eq. 2) are calculated for both the 80th and 20th percentile events at each observation station:

$$\text{POD} = \frac{\text{hits}}{\text{hits} + \text{misses}}, \quad (1)$$

$$\text{FAR} = \frac{\text{false alarms}}{\text{hits} + \text{false alarms}}, \quad (2)$$

where a hit is defined when the forecast correctly exceeded (fell below) the 80th (20th) percentile of the reference climatology during the same week that the observed river flow exceeded (fell below) the 80th (20th) percentile of the observations at that station. It follows that a miss is defined when an event was observed but the forecast did not exceed the threshold, and a false alarm when the forecast exceeded the threshold but no event was observed. From these, the area un-

der the ROC curve (AROC) is calculated, again for both the 80th and 20th percentile events. The AROC ($0 \leq \text{AROC} \leq 1$, where 1 is perfect) indicates the skill of the forecasts compared to the long-term average climatology (which has an AROC of 0.5) and is used here to evaluate the potential usefulness of the forecasts. The maximum lead time at which forecasts are more skilful than climatology ($\text{AROC} > 0.5$) is identified; a forecast with an $\text{AROC} < 0.5$ would be less skilful than climatology and thus not useful.

The reliability of the forecasts is assessed using attributes diagrams, which show the relationship between the forecast probability and the observed frequency of the events. While the ROC measures the ability of a forecasting system to predict the correct category of an event, the reliability assesses how closely the forecast probabilities correspond to the actual chance of observing the event. As such, these evaluation metrics are useful to consider together. As with the ROC calculations, the reliability is assessed for each week of the forecast (out to 16 weeks) and for both the 80th and 20th percentile events. The range of forecast probabilities is divided into 10 bins (0%–10%, 10%–20%, etc.), and the forecast probability is plotted against the frequency at which an event was observed for forecasts in each probability bin. Perfect reliability is exhibited when the forecast probability and the observed frequency are equal; for example, if a forecast predicts that an event will occur with a probability of 60%, then the event should occur on 60% of the occasions that this fore-

cast was made. Attributes diagrams can also be used to assess the sharpness and resolution of the forecasts. Forecasts that do not discriminate between events and non-events are said to have no resolution (a forecast of climatology would have no resolution), and forecasts which are capable of predicting events with probabilities that differ from the observed frequency, such as forecasts of high or 0 probability, are said to have sharpness.

The GloFAS-Seasonal reforecasts (of which there are 216 covering 18 years, as described in Sect. 2.2.4 and Fig. 2) are compared to river flow observations that have been made available to GloFAS, covering 17 years of the study period up to the end of 2015 when the data were collated (see Fig. 2). To ensure a large enough sample size for this analysis, alongside the best possible spatial coverage, the following criteria are applied to the data.

- The weekly river flow data record available for each station must contain no more than 53 % (9 years) missing data. The high and low flow thresholds (the 80th and 20th percentile, respectively) are calculated using the observations for each station and for each week across the 17 years of data, so a sample size of 17 is the maximum possible. A threshold of (up to) 53 % missing data allows for a minimum sample size of eight. Selecting a smaller threshold reduced the number of stations and the spatial coverage across the globe significantly. The percentage of missing data is calculated at each station and for each week of the dataset independently, and as such the number of stations used can vary slightly with time.
- The upstream area of the corresponding grid point in the model river network must be at least 1500 km².

These criteria allow for the use of 1140 ± 14 stations globally. While the dataset contains 6122 stations, just 1664 of these contain data during the 17-year period, and none have the full 17 years of data available. Data from human-influenced rivers have not been removed, as in this study we are interested in identifying the ability of the forecasting system in its current state to predict observed events rather than the ability of the hydrological model to represent natural flow.

4.1 Potential usefulness

In order to gain an overview of the potential usefulness of the GloFAS-Seasonal forecasts across the globe, we map the maximum lead time at which the forecasts are more skilful than climatology (i.e. AROC > 0.5) at each observation station averaged across all forecast months. These results are shown in Fig. 6, and it is clear that forecasts of both high and low flow events are more skilful than climatology across much of the globe, with potentially useful forecasts at many stations out to 4 months ahead. However, there are regions where the forecasts are (on average across all fore-

cast months) not useful (i.e. AROC < 0.5), such as the western USA and Canada (excluding coastlines), much of Africa, and additionally across parts of Europe for low flow events. As forecasts with an AROC larger than but close to 0.5 could be deemed as only marginally more skilful than climatology, we apply a skill buffer, setting the threshold to AROC > 0.6 for a forecast to be deemed as potentially useful. These results are mapped in Fig. 7 and clearly indicate the reduction in the lead time at which forecasts are potentially useful (for both high and low flow events) at many stations, implying that in some locations, forecasts beyond the first 1–2 months are only marginally more skilful than climatology. There are, however, stations in some rivers with an AROC > 0.6 out to 4 months of lead time and many locations across the globe that still indicate that forecasts are potentially useful 1–2 months ahead for both high and low flow events.

These results can be further broken down by season, indicating whether the forecasts are more potentially useful at certain times of the year. Maps showing the maximum lead time at which AROC > 0.6 for each season (for forecasts started during the season; e.g. DJF indicates the average results for forecasts produced on 1 December, 1 January, and 1 February) are provided for high and low flow events in Figs. S1 and S2 in the Supplement, respectively.

The following paragraphs provide an overview of these results for each continent; for further detail please refer to the maps.

South America. For high flow events, forecasts for the Amazon basin in DJF and MAM are potentially useful out to longer lead times (up to 3–4 months) and at more stations than in JJA and SON, with similar results in MAM for low flow events. In contrast, further south, forecasts are most potentially useful JJA and SON up to 4 months ahead. In the more mountainous regions of western South America, forecasts in JJA and SON are generally less skilful than climatology for high and low flow events. In the north-west, however, for some stations, forecasts started in DJF and MAM are potentially useful up to 3 months ahead.

North America. In eastern North America, JJA and SON forecasts are most potentially useful, with more stations indicating an AROC > 0.6 out to 2–3 months ahead. However, during all seasons there are several stations in the east showing skill out to varying lead times. Much of the western half of the continent (excluding coastal areas) sees forecasts that are less skilful than climatology during all seasons, although some stations do indicate skill up to 4 months ahead for high flow, for forecasts started in MAM and JJA, and for low flow in MAM. At many coastal stations in the west, forecasts of high flow events started in DJF, MAM, and JJA indicate skill out to 3–4 months and out to ~ 6 weeks in SON.

Europe. Forecasts for European rivers generally perform best for high flow events in SON and DJF, with the exception of some larger rivers in eastern Europe, for which the forecasts are more potentially useful in JJA and SON. In MAM and JJA, the number of stations indicating no skill is gener-

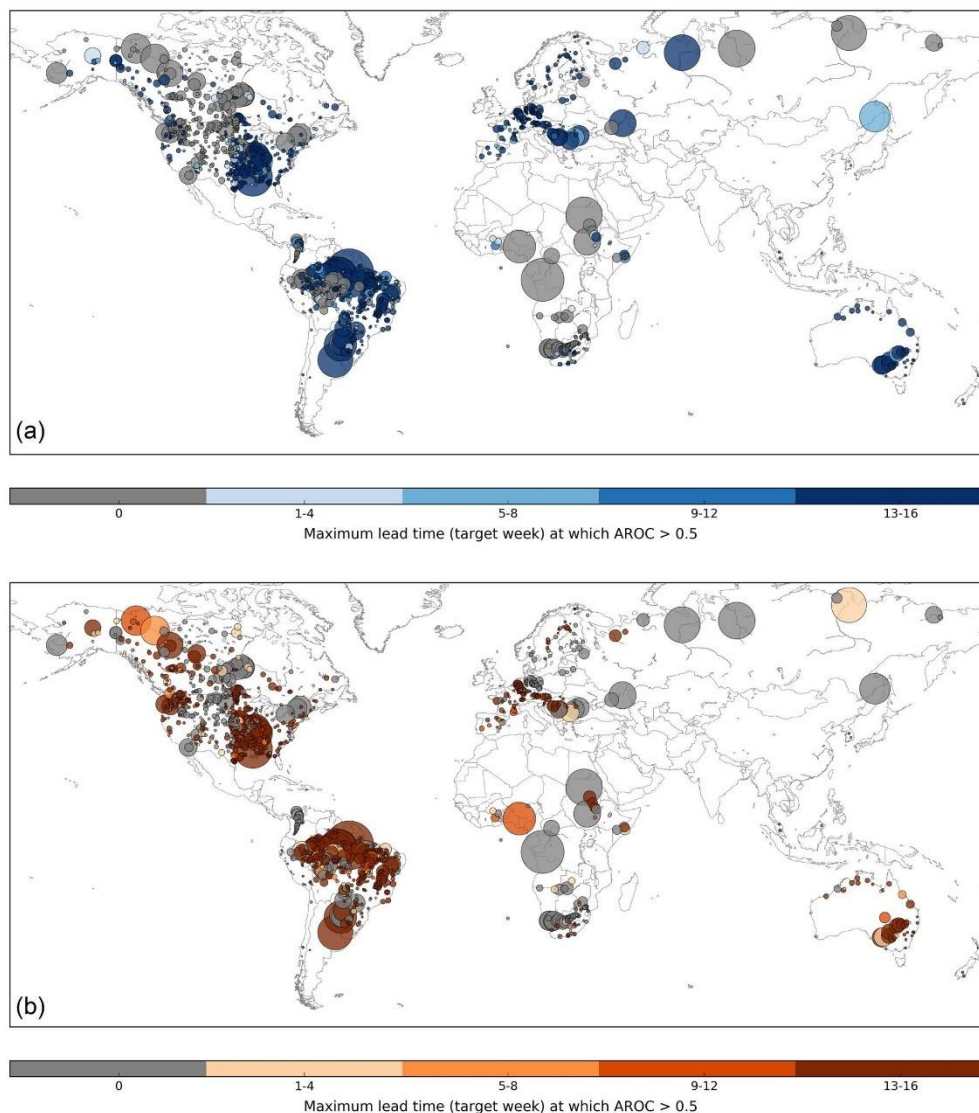


Figure 6. Maximum forecast lead time (target week, averaged across all months) at which the area under the ROC curve (AROC) is greater than 0.5 **(a)** for high flow events (flow exceeding the 80th percentile of climatology) and **(b)** low flow events (flow below the 20th percentile of climatology) at each observation station. This is used to indicate the maximum lead time at which forecasts are more skillful than the long-term average. Dot size corresponds to the upstream area of the location – thus larger dots represent larger rivers and vice versa. Grey dots indicate that (on average, across all months) forecasts are less skillful than climatology at all lead times.

ally higher. In contrast, forecasts for low flow events are less skillful than climatology across much of Europe. Particularly in north-east Europe and Scandinavia, forecasts produced in the summer months of JJA have an $\text{AROC} < 0.6$ at all sta-

tions, with only a few stations indicating any skill in other seasons, whereas in central and south-east Europe forecasts of low flow events are most skillful in JJA and SON out to 3–4 months ahead in the larger rivers. These results are similar

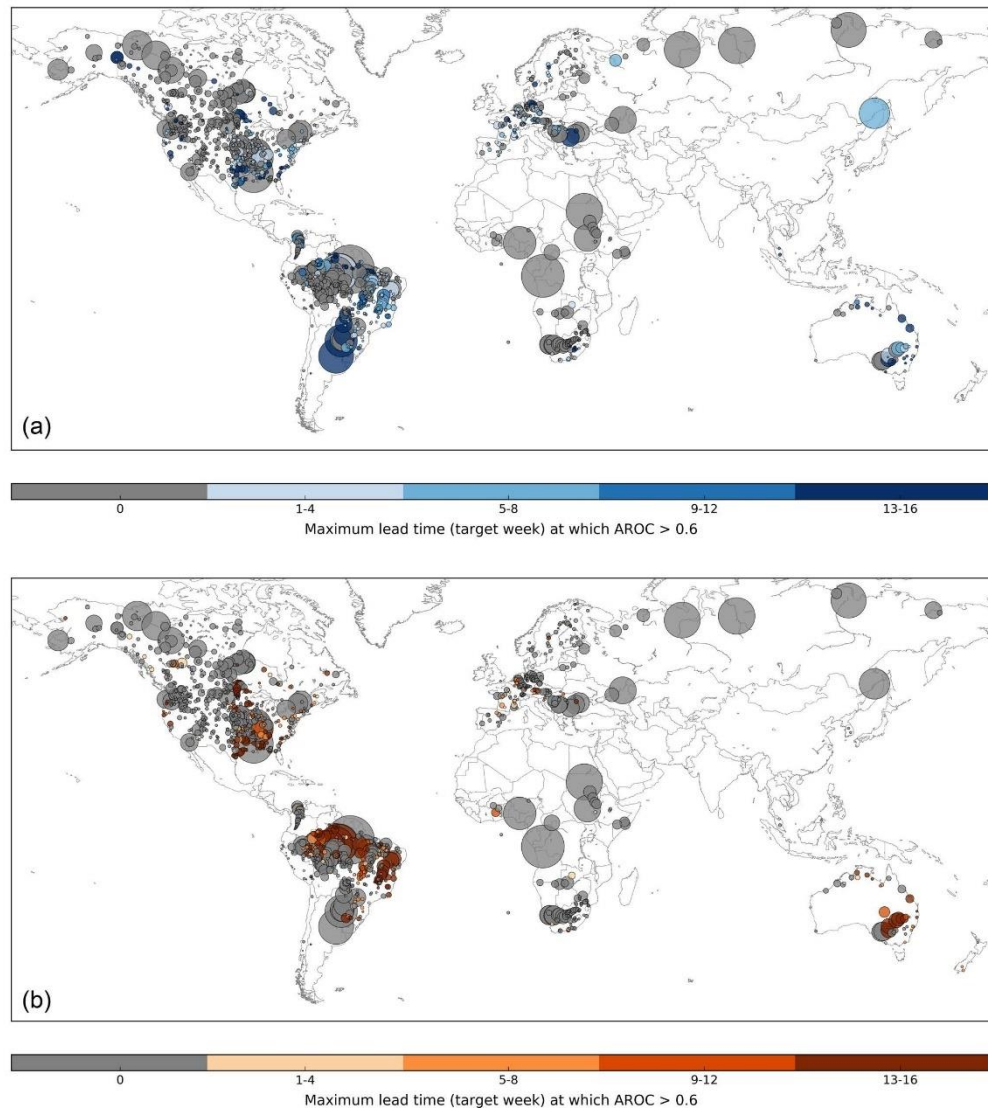


Figure 7. Maximum forecast lead time (target week, averaged across all months) at which the area under the ROC curve (AROC) is greater than 0.6 for **(a)** high flow events (flow exceeding the 80th percentile of climatology) and **(b)** low flow events (flow below the 20th percentile of climatology) at each observation station. This is used to indicate the maximum lead time at which forecasts are deemed skilful. Dot size corresponds to the upstream area of the location – thus larger dots represent larger rivers and vice versa. Grey dots indicate that (on average, across all months) forecasts are less skilful than climatology at all lead times. Maps for each season are provided in the Supplement.

to those of Arnal et al. (2018) for the potential usefulness of the EFAS seasonal outlook.

Asia. Although the number of available stations is very limited, the few stations available in South East Asia indicate

that the forecasts are potentially useful out to 3–4 months ahead, particularly for forecasts started in DJF and MAM preceding the start of the wet season. For low flow events, this skill extends into JJA, whereas forecasts made in SON

towards the end of the wet season tend to be less skilful than climatology.

Australia and New Zealand. Forecasts are most skilful out to longer lead times in the Murray–Darling river basin in the south-east, in particular for forecasts started in JJA and SON during the Southern Hemisphere winter and spring. In northern Australia, forecasts started in DJF and MAM for high flow events and MAM and JJA for low flow events are potentially useful out to 3–4 months ahead. This corresponds with the assessment of the skill of the Bayesian joint probability modelling approach for sub-seasonal to seasonal streamflow forecasting in Australia by Zhao et al. (2016), who found that forecasts in northern Australian catchments tend to be more skilful for the dry season (May to October) than the wet season (December to March). At the three stations in New Zealand, forecasts are only skilful for high flow events during the first month of lead time in DJF and MAM; however, for low flow events forecasts made in SON for the southern stations are potentially useful out to 4 months ahead.

Africa. While the spatial distribution of stations is limited, for high flow events forecasts are seen to be potentially useful at some of the stations in eastern Africa, particularly in SON and to a lesser extent in DJF. In southern Africa, there is skill in DJF and MAM, although the maximum lead time varies significantly from station to station. For low flow, there is little variation between the seasons; forecasts are generally less skilful than climatology across the continent, with some stations in DJF in southern and western Africa indicating skill in the first 1–2 months only.

4.2 Reliability

To provide an overall picture of the reliability of the GloFAS-Seasonal forecasts, attributes diagrams are produced for forecasts aggregated across all observation stations globally for both the 80th and 20th percentile events. In order to assess geographical differences in forecast reliability, attributes diagrams are also produced for forecasts aggregated across the stations within each of the major river basins used in the GloFAS-Seasonal forecast products (see Sect. 3.1). Many of these river basins do not contain a large enough number of stations to produce useful attributes diagrams, and as such the results in this section are presented for one river basin per continent for this initial evaluation. The river basin chosen for each continent is that which contains the largest number of observation stations.

The globally aggregated results (Fig. 8) indicate that, in general, the forecasts have more reliability than a forecast of climatology, though the reliability is less than perfect. It is important to note that the globally aggregated results shown in Fig. 8 mask any variability between river basins. Overall, the reliability appears to be slightly better for forecasts of high flow events than low flow events, and for lower probabilities, indicated by the steeper positive slope showing that as the forecast probability increases, so does the verified

chance of the event. The forecasts for both high and low flow events exhibit sharpness, although more so for high flow events, meaning that they have the ability to forecast probabilities that differ from the climatological average. This is indicated by the histograms inset within the attributes diagrams in Fig. 8; a forecast with sharpness will show a range of forecast probabilities differing from the climatological average (20 %), and a forecast with perfect sharpness will show peaks in the forecast frequency at 0 % and 100 %. Forecasts with no or low sharpness will show a peak in the forecast frequency near the climatological average. A forecast can have sharpness but still be unreliable. Figure 8 also suggests that in general, GloFAS-Seasonal forecasts have a tendency to over-predict the likelihood of an event occurring.

The following paragraphs summarise the forecast reliability for one river basin per continent; for a map of the location of these river basins, please refer to Fig. S3. The attributes diagrams for these river basins for both the 80th and 20th percentile events and for each season are provided in Figs. S4–S8. Each attributes diagram displays the results for forecast weeks 4, 8, 12, and 16, representing the reliability out to 1, 2, 3, and 4 months ahead. There are no river basins in Asia containing enough stations to produce an attributes diagram.

South America, Tocantins River (Fig. S4). For high flow events, forecasts for the Tocantins River indicate good reliability in all seasons, particularly up to 50 % probability. Forecasts in the higher-probability bins tend to over-predict, and this over-prediction worsens with lead time. In MAM and JJA, the forecasts tend to slightly under-predict in the lower-probability bins. The forecasts have sharpness, but it is clear that the sample size of high-probability forecasts is limited. There is a tendency to over-predict the likelihood of low flow events in all seasons, but the forecasts show good reliability for the lower-probability bins, particularly in SON and DJF. In JJA, the resolution of the forecasts is low.

North America, Lower Mississippi River (Fig. S5). For high flow events, the sample size of high-probability forecasts is small, and as such it is difficult to evaluate the reliability of these forecasts. The forecasts at lower probabilities have good reliability, particularly out to 2 months ahead in MAM and JJA. In SON and DJF, forecasts are more reliable at longer lead times. There is a tendency to under-predict at low probabilities and over-predict at high probabilities. For low flow events, the forecasts have a tendency to over-predict in all seasons, and the resolution of the forecasts is lower than for high flow events. At higher probabilities, forecasts of low flow events are more reliable than climatology, but the resolution is particularly low for probabilities up to 50–60 %. The forecasts for both high and low flow events have sharpness.

Europe, River Rhône (Fig. S6). For the River Rhône, the reliability is better than climatology at all lead times for high flow events, although there is a lack of forecasts of higher probabilities, particularly in MAM and JJA, as may be expected in the summer months. In SON, the reliability of forecasts up to 60–70 % is good at all lead times, and in DJF the

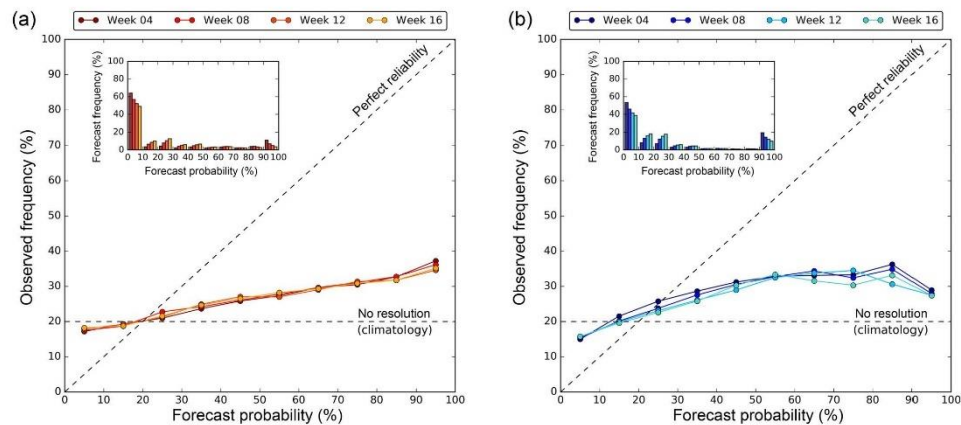


Figure 8. Attributes diagram for forecasts of (a) low flow events (flow below the 20th percentile of climatology) and (b) high flow events (flow exceeding the 80th percentile of climatology) aggregated across all observation stations globally. Results are shown for lead time weeks 4, 8, 12, and 16 and indicate the reliability of the forecasts. The histograms (inset) show the frequency at which forecasts occur in each probability bin and are used to indicate forecast sharpness. Attributes diagrams for selected river basins are provided in the Supplement.

forecasts are more reliable in the first 2 months of lead time for most probability bins. The reliability is less good for low flow events, but is generally better than climatology, particularly in summer (JJA). In winter (DJF), the resolution and reliability of the forecasts is poor. For all seasons and lead times and for both events, the forecasts have sharpness.

Australia, Murray River (Fig. S7). The attributes diagrams for both high and low flow events indicate that forecasts are often over-confident in this river basin, with probabilities of 0%–10% for low flow events and 0%–30% and 90%–100% for high flow events, occurring frequently. As such, the sample size of forecasts in several of the bins is low. For high flow events, forecasts tend to over-predict at high probabilities and under-predict at low probabilities. The reliability is very good up to ~30%, after which the sample size is too small. For low flow events, there is a tendency to under-predict, but based on the forecasts available, the reliability is better than climatology at all lead times. The reliability for low flow events is better in SON and DJF (spring and summer) than MAM and JJA (autumn and winter), and for high flow events there is less differentiation between the seasons.

Africa, Orange River (Fig. S8). For the Orange River, forecasts of high flow events exhibit good reliability for lower probabilities in SON, DJF, and MAM (spring through autumn), particularly at longer lead times in SON and DJF, with a tendency to over-predict at higher probabilities. Resolution and reliability are poor for high flow events in JJA (winter), with probabilities of 90%–100% predicted too frequently. For low flow events, forecasts of 0%–10% are very frequent, and the forecasts under-predict in all seasons, although the reliability is better than climatology at all lead times (based

on a limited sample of forecasts for most probability bins). Reliability for low flow events is best in DJF (summer).

4.3 Discussion

The results presented provide an initial evaluation of the potential usefulness and reliability of GloFAS-Seasonal forecasts. For decision-making purposes, it is important to measure the ability of a forecasting system to predict the correct category of an event. As such, an event-based evaluation of the forecasts is used to assess whether the forecasts were able to correctly predict observed high and low river flow events over a 17-year period and whether it is able to do so with good reliability. The initial results are promising, indicating that the forecasts are, on average, potentially useful up to 1–2 months ahead in many rivers worldwide and up to 3–4 months ahead in some locations. The GloFAS-Seasonal forecasts have sharpness, i.e. they are able to predict forecasts with probabilities that differ from climatology, and overall have better reliability than a forecast of climatology, but with a tendency to over-predict at higher probabilities. It is also clear that there is a frequency bias in the reliability results, as often there is a small sample of high-probability forecasts. Typically, the reliability is seen to be better when there is a higher forecast frequency on which to base the results. As would be expected, the potential usefulness and reliability of the forecasts vary by region, season, and forecast lead time.

Considering the evaluation results by season allows for further analysis of the times of year in which the forecasts are potentially useful and/or reliable. For example, in southeast Australia, forecasts are seen to be potentially useful up to 4 months ahead in JJA and SON, but for forecasts produced

in DJF the skill only extends to 1 month ahead, and forecasts are less skilful than climatology at several of the stations in MAM. In many rivers across the globe, it is the case that forecasts are potentially useful in some seasons, but not in others, and may be more reliable in certain seasons than others. As such, the maps provided in Figs. S1 and S2 are intended to highlight where and when the forecasts are likely to be useful, information that is key in terms of decision-making.

It is clear that there are regions and seasons in which the forecasts are less skilful than climatology and do not have good reliability, and thus in these rivers it would be more useful to use a long-term average climatology than seasonal hydro-meteorological forecasts of river flow. This lack of skill could be due to several factors, such as certain hydrological regimes that may not be well-represented in the hydrological model or may be difficult to forecast at these lead times (for example, snow-dominated catchments or regions where convective storms produce most of the rainfall in some seasons), poor skill of the meteorological forecast input, poor initial conditions from the ERA5-R reanalysis, extensive management of rivers that cannot be represented by the current model, or the lack of model calibration. While this initial evaluation is designed to provide an overview of whether the forecasts are potentially useful and reliable in predicting high and low flow events, more extensive analysis is required to diagnose the sources of predictability in the forecasts and the potential causes of poor skill. Additionally, it is evident that observations of river flow, particularly covering the reforecast period, are both spatially and temporally limited across large areas of the globe. A more extensive analysis should make use of the globally consistent ERA5-R river flow reanalysis as a benchmark in order to fully assess the forecast skill worldwide, including in regions where no observations are available.

The verification metrics used also require that a high or low flow event is predicted with the correct timing in the same week as that in which it occurred. This is asking a lot of a seasonal forecasting system and for many applications, such as water resources and reservoir management, a forecast of the exact week in which an event is expected at a lead time of several months ahead may not be necessary. That such a system shows real skill despite this being a tough test for the model and is able to successfully predict observed high or low river flow in a specific week, several weeks or months ahead, provides optimism for the future of global-scale seasonal hydro-meteorological forecasting. Further evaluation should aim to assess the skill of the forecasts with a more relaxed constraint on the event timing and also make use of alternative skill measures to cover different aspects of the forecast skill, such as the spread and bias of the forecasts. It will also be important to assess whether the use of weekly averaged river flow is the most appropriate way to display the forecasts. While this is commonly used for applications such as drought early awareness and water resources management, there may be other aspects of decision-making,

such as flood forecasting, for which other measures may be more appropriate, for example daily averages or floodiness (Stephens et al., 2015).

Future development of GloFAS-Seasonal will aim to address these evaluation results and improve the skill and reliability of the current forecasts; it will also aim to overcome some of the grand challenges in operational hydrological forecasting, such as seamless forecasting and the use of data assimilation. Seamless forecasting will be key in the future development of GloFAS; the use of two different meteorological forecast inputs for the medium-range and seasonal versions of the model means that discrepancies can occur between the two timescales, thus producing confusing and inconsistent forecast information for users. Additionally, the use of river flow observations could lead to significant improvements in skill through calibration of the model using historical observations and assimilation of real-time data to adjust the forecasts. This remains a grand challenge due to the lack of openly available river flow data, particularly in real time.

5 Conclusions

In this paper, the development and implementation of a global-scale operational seasonal hydro-meteorological forecasting system, GloFAS-Seasonal, was presented, and an event-based forecast evaluation was carried out using two different but complementary verification metrics to assess the capability of the forecasts to predict high and low river flow events.

GloFAS-Seasonal provides forecasts of high or low river flow out to 4 months ahead for the global river network through three new forecast product layers via the openly available GloFAS web interface at <http://www.globalfloods.eu> (last access: 16 August 2018). Initial evaluation results are promising, indicating that in many rivers, forecasts are both potentially useful, i.e. more skilful than a long-term average climatology out to several months ahead in some cases, and overall more reliable than a forecast of climatology. Forecast skill and reliability vary significantly by region and by season.

The initial evaluation, however, also indicates a tendency of the forecasts to over-predict in general, and in some regions forecasts are currently less skilful than climatology; future development of the system will aim to improve the forecast skill and reliability with a view to providing potentially useful forecasts across the globe. Development of GloFAS-Seasonal will continue based on results of the forecast evaluation and on feedback from GloFAS partners and users worldwide in order to provide a forecast product that remains state of the art in hydro-meteorological forecasting and caters to the needs of its users. Future versions are likely to address some of the grand challenges in hydro-meteorological forecasting in order to improve forecast skill, such as data assim-

ilation, and will also include more features, such as flexible percentile thresholds and indication of the forecast skill via the interface. A further grand challenge that is important in terms of global-scale hydro-meteorological forecasting, and indeed for the development of GloFAS, is the need for more observed data (Emerton et al., 2016), which is essential not only for providing initial conditions to force the models, but also for evaluation of the forecasts and continuous improvement of forecast accuracy.

While such a forecasting system requires extensive computing resources, the potential for use in decision-making across a range of water-related sectors, and the promising results of the initial evaluation, suggest that it is a worthwhile use of time and resources to develop such global-scale systems. Recent papers have highlighted the fact that seasonal forecasts of precipitation are not necessarily a good indicator of potential floodiness and called for investment in better forecasts of seasonal flood risk (Coughlan De Perez et al., 2017; Stephens et al., 2015). Coughlan de Perez et al. (2017) state that “ultimately, the most informative forecasts of flood hazard at the seasonal scale could be seasonal streamflow forecasts using hydrological models” and that better seasonal forecasts of flood risk could be hugely beneficial for disaster preparedness.

GloFAS-Seasonal represents a first attempt at overcoming the challenges of producing and providing openly available seasonal hydro-meteorological forecast products, which are key for organisations working at the global scale and for regions where no other forecasting system exists. We provide, for the first time, seasonal forecasts of hydrological variables for the global river network by driving a hydrological model with seasonal meteorological forecasts. GloFAS-Seasonal forecasts could be used in addition to other forecast products, such as seasonal rainfall forecasts and short-range forecasts from national hydro-meteorological centres across the globe, to provide useful added information for many water-related applications from water resources management and agriculture to disaster risk reduction.

Code availability. The ECMWF IFS source code is available subject to a licence agreement, and as such access is available to the ECMWF member-state weather services and other approved partners. The IFS code is also available for educational and academic purposes as part of the OpenIFS project (ECMWF, 2011, 2018a), with full forecast capabilities and including the HTESSSEL land surface scheme, but without modules for data assimilation. Similarly, the GloFAS river routing component source code is not openly available; however, the “forecast product” code (prior to implementation in ecFlow) that was newly developed for GloFAS-Seasonal and used for a number of tasks such as computing exceedance probabilities and producing the graphics for the interface is provided in the Supplement.

Data availability. ECMWF’s ERA5 reanalysis and SEAS5 reforecasts are available through the Copernicus Climate Data Store (Copernicus, 2018a). The ERA5-R river flow reanalysis and the GloFAS-Seasonal reforecasts (daily data) are currently available from the authors on request and will be made available through ECMWF’s data repository in due course. The majority of the observed river flow data were provided by the Global Runoff Data Centre (GRDC; BfG, 2017). These data are freely available from <https://www.bafg.de/> (last access: 16 August 2018). Additional data were provided by the Russian State Hydrological Institute (SHI, 2018), the European Flood Awareness System (EFAS, 2017), Somalia Water and Land Information Management (SWALIM, 2018), South Africa Department for Water and Sanitation (DWA, 2018), Colombia Institute of Hydrology, Meteorology and Environmental Studies (IDEAM, 2014), Nicaragua Institute of Earth Studies (INETER, 2016), Dominican Republic National Institute of Hydraulic Resources (INDRHI, 2017), Brazil National Centre for Monitoring and Forecasting of Natural Hazards (Cemaden, 2017), Environment Canada Water Office (Environment Canada, 2014), Nepal Department of Hydrology and Meteorology (DHM, 2017), Red Cross Red Crescent Climate Centre (RCCC, 2018), Chile General Water Directorate (DGA, 2018), and the Historical Database on Floods (BDHI, 2018).

Supplement. The supplement related to this article is available online at: <https://doi.org/10.5194/gmd-11-3327-2018-supplement>.

Author contributions. FP proposed the development of GloFAS-Seasonal, RE wrote the GloFAS-Seasonal forecast product code, and RE and LA designed the forecast products. EZ built the ecFlow suite and produced ERA5-R and the GloFAS-Seasonal reforecasts, and DM provided technical support for the website and operational implementation. RE evaluated the forecasts and wrote the paper, with the exception of Sect. 2.4, written by DM. All authors were involved in discussions throughout development, and all authors commented on the paper.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work has been funded by the Natural Environment Research Council (NERC) as part of the SCENARIO Doctoral Training Partnership under grant NE/L002566/1. Ervin Zsoter, Davide Muraro, Christel Prudhomme, and Peter Salamon were supported by the Copernicus Emergency Management Service – Early Warning Systems (CEMS-EWS; EFAS). Louise Arnal, Hannah L. Cloke, and Florian Pappenberger acknowledge financial support from the Horizon 2020 IMPREX project (grant agreement no. 641811). Elisabeth M. Stephens is thankful for support from NERC and the Department for International Development (grant number NE/P000525/1) under the Science for Humanitarian Emergencies and Resilience (SHEAR) research programme (project FATHUM: Forecasts for Anticipatory HUMANitarian action).

Edited by: Jeffrey Neal
Reviewed by: two anonymous referees

References

- Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger, F.: 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>, 2013.
- Arnal, L., Cloke, H. L., Stephens, E., Wetterhall, F., Prudhomme, C., Neumann, J., Krzeminski, B., and Pappenberger, F.: Skilful seasonal forecasts of streamflow over Europe?, *Hydrol. Earth Syst. Sci.*, 22, 2057–2072, <https://doi.org/10.5194/hess-22-2057-2018>, 2018.
- Balra, A.: Managing work flows with ecFlow, *ECMWF Newsl.*, 129, 30–32 available from: <https://www.ecmwf.int/sites/default/files/elibrary/2011/14594-newsletter-no129-autumn-2011.pdf> (last access: 18 April 2018), 2011.
- Balsamo, G., Pappenberger, F., Dutra, E., Viterbo, P., and van den Hurk, B.: A revised land hydrology in the ECMWF model: a step towards daily water flux prediction in a fully-closed water cycle, *Hydrol. Process.*, 25, 1046–1054, <https://doi.org/10.1002/hyp.7808>, 2011.
- BDHI: Base de Donnees Historiques sur les Inondations, available at: <http://bdhi.fr/appli/web/welcome>, last access: 23 April 2018.
- Bell, V. A., Davies, H. N., Kay, A. L., Brookshaw, A., and Scaife, A. A.: A national-scale seasonal hydrological forecast system: development and evaluation over Britain, *Hydrol. Earth Syst. Sci.*, 21, 4681–4691, <https://doi.org/10.5194/hess-21-4681-2017>, 2017.
- Bennett, J. C., Wang, Q. J., Li, M., Robertson, D. E., and Schepen, A.: Reliable long-range ensemble streamflow forecasts: Combining calibrated climate forecasts with a conceptual runoff model and a staged error model, *Water Resour. Res.*, 52, 8238–8259, <https://doi.org/10.1002/2016WR019193>, 2016.
- Bennett, J. C., Wang, Q. J., Robertson, D. E., Schepen, A., Li, M., and Michael, K.: Assessment of an ensemble seasonal streamflow forecasting system for Australia, *Hydrol. Earth Syst. Sci.*, 21, 6007–6030, <https://doi.org/10.5194/hess-21-6007-2017>, 2017.
- BFG: The GRDC, available at: http://www.bafg.de/GRDC/EN/Home/homepage_node.html (last accessed: 23 April 2018), 2017.
- BoM: Seasonal Streamflow Forecasts: Water Information: Bureau of Meteorology, available at: <http://www.bom.gov.au/water/ssf/about.shtml>, last access: 24 April 2018.
- Candogan Yossef, N., van Beek, R., Weerts, A., Winsemius, H., and Bierkens, M. F. P.: Skill of a global forecasting system in seasonal ensemble streamflow prediction, *Hydrol. Earth Syst. Sci.*, 21, 4103–4114, <https://doi.org/10.5194/hess-21-4103-2017>, 2017.
- Cemaden: Cemaden – Centro Nacional de Monitoramento e Alertas de Desastres Naturais, available at: <http://www.cemaden.gov.br/> (last access: 23 April 2018), 2017.
- Chiew, F. H. S. and McMahon, T. A.: Global ENSO-streamflow teleconnection, streamflow forecasting and interannual variability, *Hydrol. Sci. J.*, 47, 505–522, <https://doi.org/10.1080/02626660209492950>, 2002.
- Chow, V. Te, Maidment, D. R., and Mays, L. W.: Applied hydrology, Tata McGraw-Hill Education, available at: https://books.google.co.uk/books/about/Applied_Hydrology.html?id=RRwidSsBJrEC&redir_esc=y (last access: 17 November 2017), 2010.
- Cloke, H., Pappenberger, F., Thielen, J., and Thiemeig, V.: Operational European Flood Forecasting, in *Environmental Modelling*, John Wiley & Sons, Ltd, Chichester, UK, 415–434, 2013.
- Copernicus: Copernicus Climate Data Store, available at: <https://climate.copernicus.eu/climate-data-store>, last access: 23 April 2018a.
- Copernicus: SWICCA, Service for Water Indicators in Climate Change Adaptation, available at: <http://swicca.climate.copernicus.eu/>, last access: 12 January 2018b.
- Coughlan de Perez, E., Stephens, E., Bischiniotis, K., van Aalst, M., van den Hurk, B., Mason, S., Nissan, H., and Pappenberger, F.: Should seasonal rainfall forecasts be used for flood preparedness?, *Hydrol. Earth Syst. Sci.*, 21, 4517–4524, <https://doi.org/10.5194/hess-21-4517-2017>, 2017.
- Crochemore, L., Ramos, M.-H., and Pappenberger, F.: Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts, *Hydrol. Earth Syst. Sci.*, 20, 3601–3618, <https://doi.org/10.5194/hess-20-3601-2016>, 2016.
- Demargne, J., Wu, L., Regonda, S. K., Brown, J. D., Lee, H., He, M., Seo, D.-J., Hartman, R., Herr, H. D., Fresch, M., Schaake, J., Zhu, Y., Demargne, J., Wu, L., Regonda, S. K., Brown, J. D., Lee, H., He, M., Seo, D.-J., Hartman, R., Herr, H. D., Fresch, M., Schaake, J., and Zhu, Y.: The Science of NOAA's Operational Hydrologic Ensemble Forecast Service, *J. Am. Meteorol. Soc.*, 95, 79–98, <https://doi.org/10.1175/BAMS-D-12-00081.1>, 2014.
- DGA: Ministerio de Obras Públicas – Dirección de General de Aguas, available at: <http://www.dga.cl/Paginas/default.aspx>, last access: 23 April 2018.
- DHM: Department of Hydrology and Meteorology, available at: <http://www.dhm.gov.np/> (last access: 23 April 2018), 2017.
- DWA: Department: Water and Sanitation, available at: <http://www.dwa.gov.za/default.aspx>, last access: 23 April 2018.
- ECMWF: OpenIFS, available at: <https://www.ecmwf.int/en/research/projects/openifs> (last access: 16 August 2018), 2011.
- ECMWF: ecFlow Documentation, available at: <https://software.ecmwf.int/wiki/display/ECFLOW/Documentation> (last access: 18 April 2018), 2012.
- ECMWF: SEAS5 user guide, available at: https://www.ecmwf.int/sites/default/files/medialibrary/2017-10/System5_guide.pdf (last access: 18 April 2018), 2017a.
- ECMWF: What are the changes from ERA-Interim to ERA5? – Copernicus Knowledge Base – ECMWF Confluence Wiki, available at: <https://software.ecmwf.int/wiki/pages/viewpage.action?pageId=74764925> (last access: 24 April 2018), 2017b.
- ECMWF: About OpenIFS, available at: <https://software.ecmwf.int/wiki/display/OIFS/About+OpenIFS>, last access: 26 April 2018a.
- ECMWF: ECMWF IFS Documentation CY43R1, available at: https://www.ecmwf.int/search/elibrary/IFS?secondary_title=IFSDocumentationCY43R1, last access: 18 April 2018b.
- EFAS: European Flood Awareness System (EFAS), available at: <https://www.efas.eu/> (last access: 23 April 2018), 2017.

- Emerton, R., Cloke, H. L., Stephens, E. M., Zsoter, E., Woolnough, S. J., and Pappenberger, F.: Complex picture for likelihood of ENSO-driven flood hazard, *Nat. Commun.*, 8, 14796, <https://doi.org/10.1038/ncomms14796>, 2017.
- 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.: Continental and global scale flood forecasting systems, *Wiley Interdiscip. Rev. Water*, 3, 391–418, <https://doi.org/10.1002/wat2.1137>, 2016.
- Environment Canada: Water Level and Flow – Environment Canada, available at: <https://wateroffice.ec.gc.ca/> (last access: 23 April 2018), 2014.
- Environmental Systems Research Institute: ArcMap, ArcGIS Desktop, available at: <http://desktop.arcgis.com/en/arcmap/>, last access: 26 April 2018.
- Fekete, B. M., Vörösmarty, C. J., and Lammers, R. B.: Scaling gridded river networks for macroscale hydrology: Development, analysis, and control of error, *Water Resour. Res.*, 37, 1955–1967, <https://doi.org/10.1029/2001WR900024>, 2001.
- GloFAS: GloFAS Web Map Service Time (WMS-T) User Manual, available at: http://www.globalfloods.eu/static/downloads/GloFAS-WMS-T_usermanual.pdf last access: 26 April 2018a.
- GloFAS: GloFAS WMS-T, available at: <http://globalfloods-ows.ecmwf.int/glofas-ows/?service=WMS&request=GetCapabilities>, last access: 16 August 2018b.
- IDEAM: IDEAM, available at: <http://www.ideam.gov.co/> (last access: 23 April 2018), 2014.
- INDRHI: INDRHI – National Institute of Hydraulic Resources, available at: <http://indrhi.gob.do/> (last access: 23 April 2018), 2017.
- INETER: Ineter, Instituto Nicaragüense de Estudios Territoriales, available at: <http://www.ineter.gob.ni/> (last access: 23 April 2018), 2016.
- Lehner, B. and Grill, G.: Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems, *Hydrol. Process.*, 27, 2171–2186, <https://doi.org/10.1002/hyp.9740>, 2013.
- Lehner, B., Verdin, K., and Jarvis, A.: New Global Hydrography Derived From Spaceborne Elevation Data, *Eos, Trans. Am. Geophys. Union*, 89, 93–94, <https://doi.org/10.1029/2008EO100001>, 2008.
- Lorenz, E. N.: Deterministic Nonperiodic Flow, *J. Atmos. Sci.*, 20, 130–141, [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2), 1963.
- Lorenz, E. N.: The essence of chaos, University of Washington Press, 1993.
- Mason, S. J. and Graham, N. E.: Conditional Probabilities, Relative Operating Characteristics, and Relative Operating Levels, *Weather Forecast.*, 14, 713–725, [https://doi.org/10.1175/1520-0434\(1999\)014<0713:CPROCA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1999)014<0713:CPROCA>2.0.CO;2), 1999.
- McPhaden, M. J., Zebiak, S. E., and Glantz, M. H.: ENSO as an integrating concept in earth science., *Science*, 314, 1740–1745, <https://doi.org/10.1126/science.1132588>, 2006.
- Meißner, D., Klein, B., and Ionita, M.: Development of a monthly to seasonal forecast framework tailored to inland waterway transport in central Europe, *Hydrol. Earth Syst. Sci.*, 21, 6401–6423, <https://doi.org/10.5194/hess-21-6401-2017>, 2017.
- Mo, K. C., Lettenmaier, D. P., Mo, K. C., and Lettenmaier, D. P.: Hydrologic Prediction over the Conterminous United States Using the National Multi-Model Ensemble, *J. Hydrometeorol.*, 15, 1457–1472, <https://doi.org/10.1175/JHM-D-13-0197.1>, 2014.
- OGC: OGC Web Map Service v1.3.0, <https://doi.org/10.3173/air.21.76>, 2015.
- OpenLayers: OpenLayers, available at: <http://openlayers.org/>, last access: 18 April 2018.
- Open Source Geospatial Foundation: MapServer 7.0.1 documentation, available at: <http://mapserver.org/uk/index.html> (last access: 26 April 2018), 2016.
- Pappenberger, F., Ramos, M. H., Cloke, H. L., Wetterhall, F., Alfieri, L., Bogner, K., Mueller, A., and Salamon, P.: How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction, *J. Hydrol.*, 522, 697–713, <https://doi.org/10.1016/J.JHYDROL.2015.01.024>, 2015.
- Prudhomme, C., Hannaford, J., Harrigan, S., Boorman, D., Knight, J., Bell, V., Jackson, C., Svensson, C., Parry, S., Bachiller-Jareno, N., Davies, H., Davis, R., Mackay, J., McKenzie, A., Rudd, A., Smith, K., Bloomfield, J., Ward, R., and Jenkins, A.: Hydrological Outlook UK: an operational streamflow and groundwater level forecasting system at monthly to seasonal time scales, *Hydrolog. Sci. J.*, 62, 2753–2768, <https://doi.org/10.1080/02626667.2017.1395032>, 2017.
- QGIS Development Team: Quantum GIS Geographical Information System, available at: <https://www.qgis.org/>, last access: 4 December 2017.
- RCCC: Home – Red Cross Red Crescent Climate Centre, available at: <http://www.climatecentre.org/>, last access: 23 April 2018.
- SHI: “State Hydrological Institute” (SHI), Russian Federal State Budgetary Organization, available at: <http://www.hydrology.ru/en>, last access: 23 April 2018.
- Stephens, E., Day, J. J., Pappenberger, F., and Cloke, H.: Precipitation and floodiness, *Geophys. Res. Lett.*, 42, 10316–10323, <https://doi.org/10.1002/2015GL066779>, 2015.
- Stockdale, T., Johnson, S., Ferranti, L., Balmaseda, M., and Briceag, S.: ECMWF's new long-range forecasting system SEAS5, *ECMWF Newsl.*, 154, 15–20, available at: <http://www.ecmwf.int/en/about/news-centre/media-resources>, last access: 18 April 2018.
- SWALIM: FAO SWALIM: Somalia Water and Land Information Management, available at: <http://www.faoswalim.org/>, last access: 23 April 2018.
- Van Der Knijff, J. M., Younis, J., and De Roo, A. P. J.: LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation, *Int. J. Geogr. Inf. Sci.*, 24, 189–212, <https://doi.org/10.1080/13658810802549154>, 2010.
- Ward, P. J., Eisner, S., Flörke, M., Dettlinger, M. D., and Kumm, M.: Annual flood sensitivities to El Niño–Southern Oscillation at the global scale, *Hydrol. Earth Syst. Sci.*, 18, 47–66, <https://doi.org/10.5194/hess-18-47-2014>, 2014a.
- Ward, P. J., Jongman, B., Kumm, M., Dettlinger, M. D., Sperna Weiland, F. C., and Winsemius, H. C.: Strong influence of El Niño Southern Oscillation on flood risk around the world, *P. Natl. Acad. Sci. USA*, 111, 15659–15664, <https://doi.org/10.1073/pnas.1409822111>, 2014b.
- Ward, P. J., Kumm, M., and Lall, U.: Flood frequencies and durations and their response to El Niño Southern Oscillation: Global analysis, *J. Hydrol.*, 539, 358–378, <https://doi.org/10.1016/J.JHYDROL.2016.05.045>, 2016.

- Weisheimer, A. and Palmer, T. N.: On the reliability of seasonal climate forecasts, *J. R. Soc. Interface*, 11, 20131162, <https://doi.org/10.1098/rsif.2013.1162>, 2014.
- Wood, A. W., Maurer, E. P., Kumar, A., and Lettenmaier, D. P.: Long-range experimental hydrologic forecasting for the eastern United States, *J. Geophys. Res.*, 107, 4429, <https://doi.org/10.1029/2001JD000659>, 2002.
- Wood, A. W., Kumar, A., and Lettenmaier, D. P.: A retrospective assessment of National Centers for Environmental Prediction climate model-based ensemble hydrologic forecasting in the western United States, *J. Geophys. Res.*, 110, D04105, <https://doi.org/10.1029/2004JD004508>, 2005.
- Yuan, X., Wood, E. F., Chaney, N. W., Sheffield, J., Kam, J., Liang, M., and Guan, K.: Probabilistic Seasonal Forecasting of African Drought by Dynamical Models, *J. Hydrometeorol.*, 14, 1706–1720, <https://doi.org/10.1175/JHM-D-13-054.1>, 2013.
- Yuan, X., Wood, E. F., and Ma, Z.: A review on climate-model-based seasonal hydrologic forecasting: physical understanding and system development, *Wiley Interdiscip. Rev. Water*, 2, 523–536, <https://doi.org/10.1002/wat2.1088>, 2015a.
- Yuan, X., Roundy, J. K., Wood, E. F., Sheffield, J., Yuan, X., Roundy, J. K., Wood, E. F., and Sheffield, J.: Seasonal Forecasting of Global Hydrologic Extremes: System Development and Evaluation over GEWEX Basins, *B. Am. Meteorol. Soc.*, 96, 1895–1912, <https://doi.org/10.1175/BAMS-D-14-00003.1>, 2015b.
- Zhao, T., Schepen, A., and Wang, Q. J.: Ensemble forecasting of sub-seasonal to seasonal streamflow by a Bayesian joint probability modelling approach, *J. Hydrol.*, 541, 839–849, <https://doi.org/10.1016/j.jhydrol.2016.07.040>, 2016.

A6: GloFAS-ERA5 operational global river discharge reanalysis 1979-present

This paper presents a co-author contribution arising through collaboration during this PhD, and has the following reference:

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

Earth Syst. Sci. Data, 12, 2043–2060, 2020
<https://doi.org/10.5194/essd-12-2043-2020>
 © Author(s) 2020. This work is distributed under
 the Creative Commons Attribution 4.0 License.



Open Access
 Earth System
 Science
 Data

GloFAS-ERA5 operational global river discharge reanalysis 1979–present

Shaun Harrigan¹, Ervin Zsoter^{1,2}, Lorenzo Alfieri^{3,4}, Christel Prudhomme^{1,5,6}, Peter Salamon³, Fredrik Wetterhall¹, Christopher Barnard¹, Hannah Cloke^{2,7,8,9}, and Florian Pappenberger¹

¹Forecast Department, European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK

²Department of Geography and Environmental Science, University of Reading, Reading, UK

³Disaster Risk Management Unit, European Commission Joint Research Centre (JRC), Ispra, Italy

⁴CIMA Research Foundation, Savona, Italy

⁵Centre for Ecology and Hydrology (CEH), Wallingford, UK

⁶Department of Geography and Environment, University of Loughborough, Loughborough, UK

⁷Department of Meteorology, University of Reading, Reading, UK

⁸Department of Earth Sciences, Uppsala University, Uppsala, Sweden

⁹Centre of Natural Hazards and Disaster Science, CNDS, Uppsala, Sweden

Correspondence: Shaun Harrigan (shaun.harrigan@ecmwf.int)

Received: 25 November 2019 – Discussion started: 7 January 2020

Revised: 28 June 2020 – Accepted: 1 July 2020 – Published: 7 September 2020

Abstract. Estimating how much water is flowing through rivers at the global scale is challenging due to a lack of observations in space and time. A way forward is to optimally combine the global network of earth system observations with advanced numerical weather prediction (NWP) models to generate consistent spatio-temporal maps of land, ocean, and atmospheric variables of interest, which is known as a reanalysis. While the current generation of NWP models output runoff at each grid cell, they currently do not produce river discharge at catchment scales directly and thus have limited utility in hydrological applications such as flood and drought monitoring and forecasting. This is overcome in the Global Flood Awareness System (GloFAS; <http://www.globalfloods.eu/>, last access: 28 June 2020) by coupling surface and sub-surface runoff from the Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land (HTESSEL) land surface model used within ECMWF's latest global atmospheric reanalysis (ERA5) with the LISFLOOD hydrological and channel routing model. The aim of this paper is to describe and evaluate the GloFAS-ERA5 global river discharge reanalysis dataset launched on 5 November 2019 (version 2.1 release). The river discharge reanalysis is a global gridded dataset with a horizontal resolution of 0.1° at a daily time step. An innovative feature is that it is produced in an operational environment so is available to users from 1 January 1979 until near real time (2 to 5 d behind real time). The reanalysis was evaluated against a global network of 1801 daily river discharge observation stations. Results found that the GloFAS-ERA5 reanalysis was skilful against a mean flow benchmark in 86 % of catchments according to the modified Kling–Gupta efficiency skill score, although the strength of skill varied considerably with location. The global median Pearson correlation coefficient was 0.61 with an interquartile range of 0.44 to 0.74. The long-term and operational nature of the GloFAS-ERA5 reanalysis dataset provides a valuable dataset to the user community for applications ranging from monitoring global flood and drought conditions to the identification of hydroclimatic variability and change and as raw input for post-processing and machine learning methods that can add further value. The dataset is openly available from the Copernicus Climate Change Service Climate Data Store: <https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-glofas-historical?tab=overview> (last access: 28 June 2020) with the following DOI: <https://doi.org/10.24381/cds.a4fdd6b9> (C3S, 2019).

Published by Copernicus Publications.

1 Introduction

A key challenge in hydrology is estimating past, present, and future hydrological conditions in rivers around the world. This is largely due to severe temporal and spatial gaps in the global river discharge observing network. In many parts of the world, there is simply not enough long-term river discharge observations at high enough spatial density, and in the vast majority of countries hydrometric data are not available in real time (Lavers et al., 2019). The lack of observations is therefore a major barrier in our ability to provide monitoring and early warning of hydrological extremes such as floods and droughts, which have, for example, implications for progressing international disaster risk reduction (UNDRR, 2015). A way forward pioneered in the field of meteorology and climate has been to optimally combine in situ and satellite earth system observations, together with advanced numerical weather prediction (NWP) models, to generate a reanalysis of land, ocean, and atmospheric variables of interest, thus providing consistent spatio-temporal “maps without gaps” (Hersbach et al., 2020). Several global hydrological products have been developed that provide estimates of runoff or river discharge with a wide range of forcing and methodological approaches (e.g. Fekete et al., 2002; Döll et al., 2003; Qian et al., 2006; Sperna Weiland et al., 2010; Reichle et al., 2011; Yamazaki et al., 2011; Beck et al., 2017; Ghiggi et al., 2019; Lin et al., 2019). While these datasets can be used to understand past variability and change in the terrestrial hydrological cycle, they are currently not produced in an operational environment in near real time and so cannot be used for monitoring current global river conditions or providing initial conditions to hydrometeorological forecasting systems.

A long-term and near-real-time river discharge reanalysis is produced operationally as part of the Global Flood Awareness System (GloFAS; <http://www.globalfloods.eu/>) which bridges this gap. GloFAS is the global flood service of the European Commission’s Copernicus Emergency Management Service (CEMS), an operational system for monitoring and forecasting floods across the world with over 5000 registered users. GloFAS was developed together by the Joint Research Centre (JRC) of the European Commission, the University of Reading, and the European Centre for Medium-Range Weather Forecasts (ECMWF). The system went pre-operational in July 2011 (Alfieri et al., 2013), becoming a fully operational, 24/7 supported service in April 2018 (version 1.0, upgraded to version 2.0 in November 2018). GloFAS is provided through a free and open licence and is designed for decision makers and forecasters in national and international water authorities, water resources management, hydropower companies, civil protection authorities, and international humanitarian aid organizations. A recent example of the use of GloFAS was for supporting the humanitarian

response to the devastating floods that affected large parts of Mozambique, Malawi, and Zimbabwe in the wake of tropical cyclone Idai in March 2019 (Magnusson et al., 2019). Given the large amount of openly available data that is generated by GloFAS, including a long-term, near-real-time river discharge reanalysis, a large set of reforecasts, and real-time flood and seasonal forecasts, it is also used by researchers and commercial industries for a wide range of projects and for developing value-added products.

In GloFAS, ensemble river discharge forecasts are produced each day at a daily time step and provide probabilities of flood thresholds being exceeded for a given river section with a lead up to 30 d ahead (GloFAS 30 d; Alfieri et al., 2013). There is also a seasonal component, GloFAS-Seasonal (Emerton et al., 2018), that provides forecasts once per month at a weekly time step with a lead time up to 4 months ahead. The river discharge reanalysis is used for two core tasks within GloFAS. First, flood thresholds at 2-, 5-, and 20-year return periods for each river cell are derived from the long-term reanalysis series. This allows for the magnitude of the real-time ensemble river discharge forecasts to be directly compared to the magnitude of the long-term flood thresholds and thus awareness of a flood signal if the threshold is exceeded. Second, it provides the basis to derive initial hydrometeorological conditions for both GloFAS 30 d and GloFAS-Seasonal real-time forecasts. Estimating initial conditions is a key step to determine the current status of soil moisture, groundwater, snow cover, and initial state of water within rivers and other waterbodies, and it has been identified as one of the major challenges in continental- and global-scale flood forecasting given the limited availability of observational data at these scales (Emerton et al., 2016).

The aim of this data paper is to describe the newly produced operational river discharge reanalysis dataset as part of the launch of GloFAS v2.1 on 5 November 2019 (see GloFAS technical documentation for details on upgrades: <https://confluence.ecmwf.int/display/COPSRV/Global+Flood+Awareness+System>, last access: 28 August 2020). GloFAS river discharge reanalysis is based on ERA5 (Hersbach et al., 2020), ECMWF’s latest global atmospheric reanalysis which extends back to 1979 and was officially released in January 2019. An innovation of ERA5 is that it is produced in near real time in an operational environment, allowing for the production of GloFAS-ERA5 reanalysis with a latency of 2 to 5 d behind real time. This has the major advantage for GloFAS in that the initial hydrometeorological conditions can now be derived from the same product from which the long-term flood thresholds are derived, which will ensure a much better consistency with real-time forecasts compared to previous GloFAS model configurations. Uniquely, the global river discharge product is over 40 years long, produced in near real time, and is freely available to download for the community through the Copernicus

nicus Climate Change Service (C3S) Copernicus Climate Data Store (CDS): <https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-glofas-historical?tab=overview> (C3S, 2019), opening multitudes of hydroclimate applications across the world.

Section 2 outlines the production of the dataset and Sect. 3 describes its main attributes including available variables and file format. An evaluation of the dataset against a global network of observations is conducted in Sect. 4. The dissemination of the data through the CDS is shown in Sect. 5 before key conclusions and future work are offered in Sect. 6.

2 Data production

Pappenberger et al. (2010) first demonstrated that it was possible to achieve useful river discharge predictions by coupling a river routing scheme with the land surface model of the ECMWF global numerical weather prediction (NWP) system. The GloFAS-ERA5 river discharge reanalysis uses this concept and is produced by coupling the land surface model runoff component of the ECMWF ERA5 global reanalysis (Hersbach et al., 2020) with the LISFLOOD hydrological and channel routing model (van der Knijff et al., 2010). In ERA5, the runoff (m d^{-1}) from one cell is not connected to neighbouring cells; hence, it is not possible to estimate river discharge ($\text{m}^3 \text{s}^{-1}$) at the catchment scale. Coupling ERA5 runoff with LISFLOOD allows for the lateral connectivity of grid cells with runoff routed through the river channel to produce river discharge. A schematic of the key components in the production of the GloFAS-ERA5 reanalysis is provided in Fig. 1. The open-access scientific publications and model documentation that describe the full methodological detail for each key component is provided in Table 1 and summarized below.

2.1 ERA5 runoff

ERA5 runoff is produced from the HTESSEL land surface model (Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land; Balsamo et al., 2009) as used within the ECMWF Integrated Forecasting System (IFS). HTESSEL computes the surface water and energy fluxes and the temporal evolution of soil temperature, soil moisture, and snowpack. Excess precipitation and snowmelt are partitioned as surface runoff or infiltrated into a four-layer soil column (7 cm depth for top layer and then 21, 72, and 189 cm) at each ERA5 grid cell before draining from the bottom of the soil column as sub-surface runoff (Balsamo et al., 2009). ERA5 uses an advanced land data assimilation system to assimilate conventional in situ and satellite observations for land surface variables such as soil moisture, soil temperature, snow water equivalent, snow density, and snow temperature, as outlined in de Rosnay et al. (2014).

ERA5 benefits from a decade worth of numerical weather prediction developments in model physics, numerics, and

data assimilation by using ECMWF IFS model cycle 41r2 (2016) compared to model cycle 31r2 (2006) as used in its predecessor, ERA-Interim (Dee et al., 2011). ERA5 has a horizontal resolution of approximately 31 km at the Equator (native octahedral grid) and since January 2019 has been openly available from 1979 to present. A key novelty of ERA5 is its operational production that makes available an intermediate timely product, ERA5T, in near real time, allowing the production of the GloFAS-ERA5 river discharge reanalysis operationally with a latency of between 2 and 5 d behind real time.

2.2 LISFLOOD river discharge

River discharge is currently not calculated by HTESSEL. Instead, surface and sub-surface runoff from the HTESSEL land surface model coupled with a simplified global version of LISFLOOD, a spatially distributed grid-based hydrological and channel routing model. The details of the global version of LISFLOOD used within GloFAS v2.1 and its calibration can be found in Hirpa et al. (2018) but are briefly summarized here for context. The sub-surface runoff from HTESSEL is used as input for the LISFLOOD groundwater module, which consists of two parallel linear reservoirs that store and subsequently transport water to the river channel with a time delay. The upper zone represents quick groundwater and sub-surface flow, while the lower zone represents slow groundwater flow that generates base flow. In Hirpa et al. (2018), the upper zone time constant was given a default value of 10 d with a lower (upper) bound of 3 d (40 d) during calibration, and the lower zone time constant given a default value of 200 d with a lower (upper) bound of 40 d (500 d). The surface runoff from HTESSEL is used as input for the LISFLOOD river channel routing module. This is a two-stage process whereby the surface runoff for each cell is first routed to the nearest downstream river channel cell, then the water in the channel is routed through the river network using the kinematic wave approach. Groundwater and river routing parameters in GloFAS were calibrated against daily river discharge observations for 1287 catchments globally by Hirpa et al. (2018). A key feature of LISFLOOD is the ability to represent features that can severely alter the timing and magnitude of river discharge, such as lakes, reservoirs, and human water use (Burek et al., 2013). A total of 463 of the largest lakes (surface area $> 100 \text{ km}^2$) and 667 of the largest reservoirs were incorporated into the GloFAS river network by Zajac et al. (2017).

To generate the GloFAS-ERA5 river discharge reanalysis, the LISFLOOD model is forced with daily HTESSEL surface and sub-surface runoff from ERA5 starting from 1 January 1979 (Fig. 1). In order to be consistent with the operational GloFAS procedure, the runoff fields from ERA5 were downscaled using the simple nearest neighbour method from the native ERA5 to the 0.1° GloFAS grid. To avoid the need for very long spin-up periods, LISFLOOD calcu-

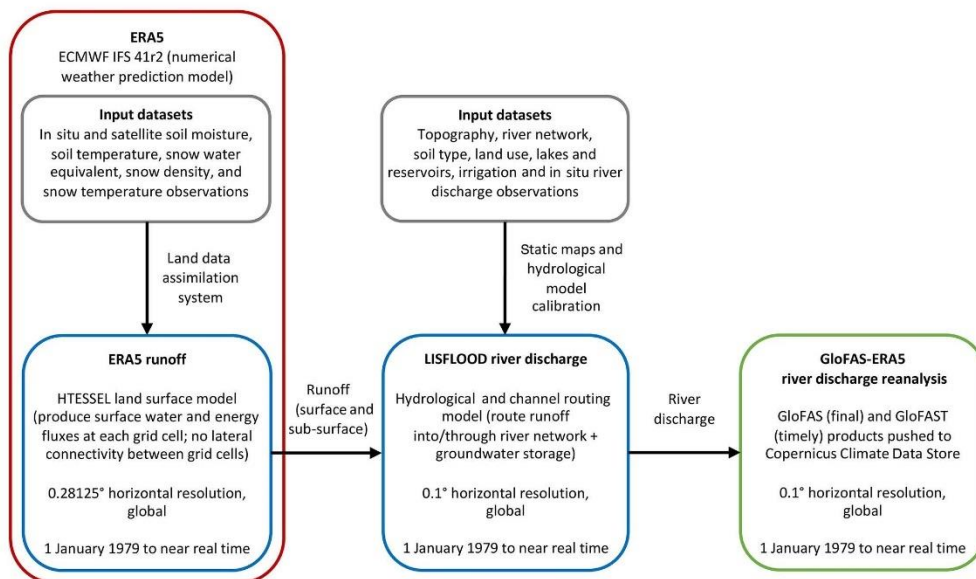


Figure 1. A schematic of the key components in the production of GloFAS-ERA5 v2.1 river discharge reanalysis dataset.

Table 1. Scientific papers and model documentation for the key components in the production of GloFAS-ERA5 v2.1 river discharge reanalysis dataset.

GloFAS-ERA5 component	Description	Reference
ERA5	Global reanalysis dataset using ECMWF Integrated Forecast System (IFS) model cycle 41r2 from 1979 to present	Hersbach et al. (2020)
ERA5 runoff	Surface and sub-surface runoff within ERA5 generated using the HTESSEL land surface model	Balsamo et al. (2009)
LISFLOOD river discharge	River discharge generated using LISFLOOD hydrological and channel routing model to route runoff into and through the river network and provide groundwater storage, including lake, reservoir, and human water use routines	Burek et al. (2013)
Lakes and reservoirs used in GloFAS	Incorporated 463 lakes and 667 reservoirs into the GloFAS river network	Zajac et al. (2017)
Calibration of LISFLOOD used in GloFAS	LISFLOOD was calibrated against daily river discharge from 1287 observation stations worldwide	Hirpa et al. (2018)

lates a steady-state storage amount for the lower groundwater zone during a long-term “pre-run” and thus reduces the lower zone’s spin-up time (Burek et al., 2013). LISFLOOD was therefore given a 1-year model spin-up using preliminary ERA5 output for 1978. To produce GloFAS-ERA5 reanalysis in near real time operationally, the latest available ERA5T data are used.

3 Data description

The key attributes of the current operational version (v2.1) of the GloFAS-ERA5 river discharge reanalysis dataset are shown in Table 2. The daily reanalysis is global in coverage, except for Antarctica, with a horizontal grid resolution of 0.1° (approximately 11 km at the Equator). The dataset is over 40 years long starting on 1 January 1979. An innovative

aspect of the dataset is its operational production allowing it to be available 2 to 5 d behind real time, shortly after ERA5T becomes available. The intermediate ERA5T data are not quality assured due to its timely nature. Consequently, there will be two reanalysis streams available: GloFAS (consolidated) is the final product based on the consolidated ERA5 from 1 January 1979 until 2 to 3 months behind real time, updated on the CDS on a monthly basis; and GloFAST (intermediate) is the timely product based on the intermediate ERA5T from 1 August 2019 until 2 to 5 d behind real time, updated on the CDS on a daily basis whenever ERA5T becomes available.

The GloFAS-ERA5 reanalysis dataset includes the variables river discharge and the upstream area for each GloFAS grid cell (Table 3). Data are stored in NetCDF format with one file per day containing the 24 h mean river discharge (00:00 UTC to 00:00 UTC). Each daily filename follows the convention “CEMS_ECMWF_dis24_<YYYYMMDD>_glofas-<T>_v2.1.nc” whereby the date stamp represents the end of the 24 h averaging period. So, for example, the file “CEMS_ECMWF_dis24_20190101_glofas_v2.1.nc” contains the daily mean flow for the 24 h period 00:00 UTC 31 December 2018 to 00:00 UTC 1 January 2019. Appendix A shows the header metadata information contained within the example NetCDF file. Each daily NetCDF file for the whole globe has an uncompressed size of ~ 21.7 MB; therefore, the estimated size of the dataset from January 1979 to October 2019 is ~ 320 GB.

Figure 2 maps the mean daily river discharge from 1979 to 2018 for each GloFAS river with an upstream area greater than 1000 km², revealing the main river arteries of the world. An example hydrograph of the long-term near-real-time reanalysis against available river discharge observations is shown in Fig. 3 for the Teles Pires River in the Amazon basin, Brazil.

4 Evaluation and limitations

GloFAS-ERA5 v2.1 river discharge reanalysis was evaluated against a global network of daily river discharge observations. As part of GloFAS, a database of global hydrological observations for 2042 stations is held, consisting predominantly (i.e. $\sim 75\%$) of data from the Global Runoff Data Centre (GRDC) and supplemented by data collected through collaboration with GloFAS partners worldwide to improve spatial coverage. A number of criteria were used to select stations for the evaluation:

- at least 4 years of daily data available between 1979 and 2018 (not necessarily contiguous) (78 stations removed);
- minimum upstream area of 500 km² (4 stations removed);

- error in catchment area supplied by data provider and upstream area for corresponding cell on the GloFAS river network within 20 % (93 stations removed);
- first order visual quality check on observed river discharge time series to remove stations with erroneous data; for example, time series truncated above a threshold, severe inhomogeneities, or series monitoring an artificial canal instead of a river (39 stations removed);
- station with the longest record retained when multiple observation stations were matched to the same GloFAS river cell (27 stations removed).

This filtering procedure resulted in the selection of 1801 stations with drainage areas ranging between 575 to 4 664 200 km² and a median of 30 046 km². Individual metadata of all 1801 stations are given in Table S1 in the Supplement. Care must be taken in spatial representativeness of the following evaluation results as the observation network is sparse in some regions of the world, particularly in large parts of Africa and Asia.

Performance at the daily scale was assessed using the modified Kling–Gupta efficiency metric (KGE' ; Gupta et al., 2009; Kling et al., 2012). The KGE' is gaining popularity as the standard performance metric in hydrology (e.g. Beck et al., 2017; Harrigan et al., 2018; Lin et al., 2019) and can be decomposed into three components important for assessing hydrological dynamics: temporal errors through correlation, bias errors, and variability errors:

$$KGE' = 1 - \sqrt{(r - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2}, \quad (1)$$

$$\beta = \frac{\mu_s}{\mu_o}, \quad (2)$$

$$\gamma = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o}, \quad (3)$$

where r is the Pearson correlation coefficient between reanalysis simulations (s) and observations (o), β is the bias ratio, γ is the variability ratio, μ is the mean discharge, and σ is the discharge standard deviation. The KGE' and its three decomposed components (correlation, bias ratio, and variability ratio) are all dimensionless with an optimum value of 1. In order to evaluate the hydrological simulation *skill* of GloFAS-ERA5 reanalysis, its performance is compared against a simpler benchmark. Here the observed mean flow is used as a benchmark as proposed by Knoben et al. (2019). This is not a difficult benchmark to beat but should arguably be the minimum reference for any hydrological system to be compared against. Here we represent KGE' as a skill score, $KGESS$, to evaluate the performance of GloFAS-ERA5 river discharge reanalysis against the mean flow benchmark simulation, given as

$$KGESS = \frac{KGE'_{\text{reanalysis}} - KGE'_{\text{bench}}}{KGE'_{\text{perf}} - KGE'_{\text{bench}}}, \quad (4)$$

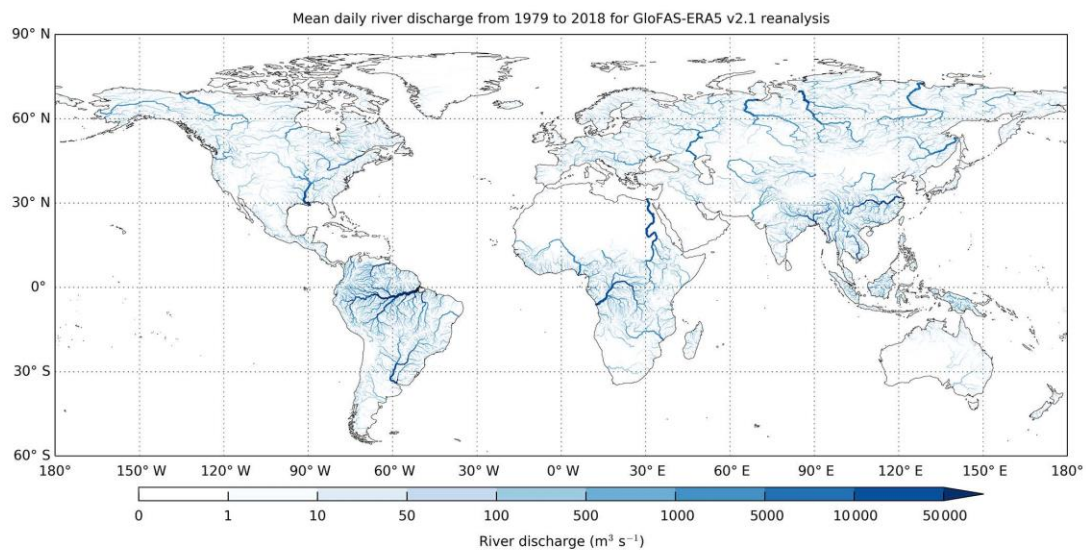


Figure 2. Mean GloFAS-ERA5 daily river discharge from 1979 to 2018 for each GloFAS river grid cell with an upstream area greater than 1000 km². Darker blue river sections have larger river discharge.

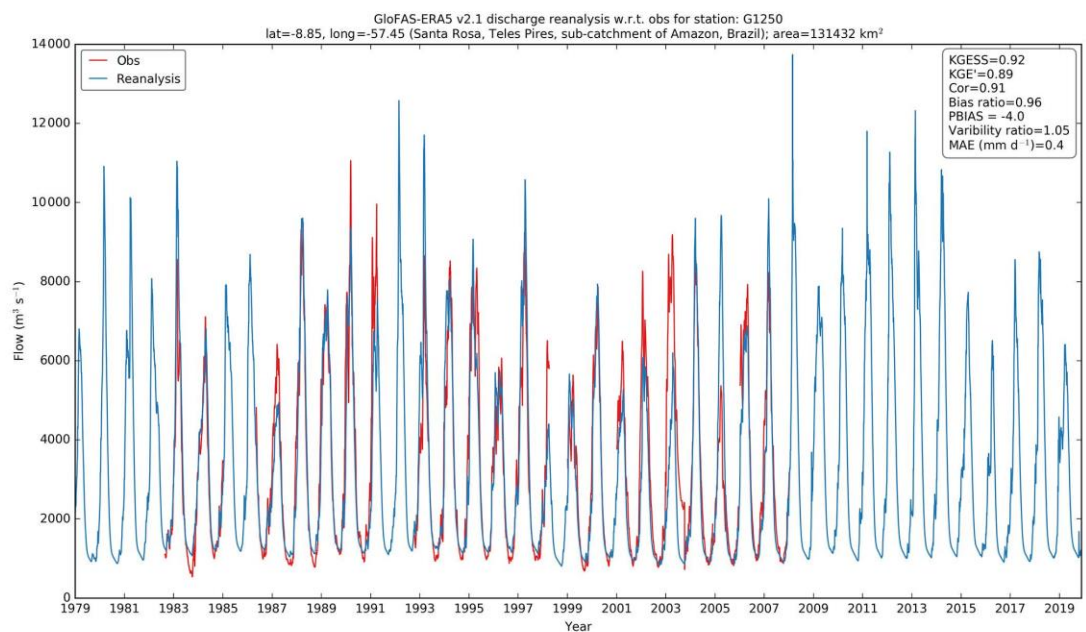


Figure 3. Hydrograph for GloFAS-ERA5 river discharge reanalysis (blue line) from 1 January 1979 to 12 November 2019 and observations (red line), when available, for the Santa Rosa gauging station on the Teles Pires River, a sub-catchment of the Amazon, Brazil (GloFAS ID = 1250; GRDC ID = 3629770). Summary statistics from the evaluation of the reanalysis against observations in top right box as used in Sect. 4.

Table 2. Summary of GloFAS-ERA5 dataset attributes in the C3S Climate Data Store.

Dataset attribute	Details
Horizontal coverage	Global except for Antarctica (90° N–60° S, 180° W–180° E)
Horizontal resolution	0.1° × 0.1°
Spatial reference system	Latitude/longitude (WGS 84; EPSG:4326)
Vertical resolution	Surface level for river discharge
Temporal resolution	Daily data
Temporal coverage	1 January 1979 to near real time
Availability behind real time	(i) GloFAS (consolidated): 2 to 3 months, updated on CDS monthly (final product following availability of officially released quality assured ERA5 data) (ii) GloFAST (intermediate): 2 to 5 d, updated on CDS daily (timely product following availability of non-quality assured ERA5T data)
Update frequency	A new river discharge reanalysis will be published with every major update of the GloFAS system. The latest version will always be the version used in operations.
File format	NetCDF
Data type	Grid
Data size on disk	Approximately 21.7 MB uncompressed per global NetCDF file for 1 d (full dataset currently ~ 320 GB uncompressed)
Version	GloFAS-ERA5 v2.1
File naming convention	“CEMS_ECMWF_dis24_<YYYYMMDD>_glofas<T>_v2.1.nc” where YYYY is year, MM is month, DD is day, and T is for timely (i.e. GloFAST). The date stamp, <YYYYMMDD>, represents the end of the 24 h averaging period.

Table 3. Variables available within GloFAS-ERA5 dataset in the C3S Climate Data Store.

Variable type	Name	Units	Description
Primary variable	River discharge	m ³ s ⁻¹	Volume rate of water flow, including sediments and chemical and biological material, in the river channel averaged over a time step through a cross section. The value is an average over a 24 h period.
Related variable	Upstream area	m ²	Static file (“upArea.nc”), upstream area for the point in the river network

where $KGE'_{\text{reanalysis}}$ is the KGE' value for the GloFAS-ERA5 reanalysis against observations, KGE'_{bench} is the KGE' value for the observed mean flow benchmark against observations (i.e. $KGE'(\overline{Q}_{\text{obs}}) = 1 - \sqrt{2} \approx -0.41$ from Knoben et al., 2019), and KGE'_{perf} is the value of KGE' for a perfect simulation which is 1. A $KGESS = 0$ means the GloFAS-ERA5 reanalysis is no better than the mean flow benchmark and so has no skill, $KGESS > 0$ means the reanalysis is considered skilful, and $KGESS < 0$ means the performance is worse than the benchmark and so has negative skill. Performance metrics for all 1801 stations are included in Table S1.

4.1 Overall performance

Results for overall performance show that the GloFAS-ERA5 river discharge reanalysis is skilful in 86 % of catchments (Fig. 4a). The global median $KGESS$ (KGE') is 0.51 (0.31) with an interquartile range (IQR) of 0.30 (0.00) to 0.66 (0.52). Performance is best in Brazil (particularly the Amazon basin), central Europe, and the eastern and western regions of the United States (Fig. 5). GloFAS-ERA5 reanalysis performance is poor (i.e. $KGESS < 0$) in many catchments in Africa and the North American Great Plains extending into Mexico with notable patches in eastern Brazil, Thailand, and southern Spain. Results will be biased towards regions with a larger number of stations, especially when well performing

large basins contain many sub-catchments (e.g. Amazon and Rhine basins).

4.2 Decomposition into correlation, bias, and variability

An advantage with the KGE' is that it can be decomposed into three constituent components so that greater insights can be gained into which aspects of the GloFAS-ERA5 reanalysis are driving poor and good skills. Almost all (99 %) catchments show a positive correlation (Figs. 4b and 6a) with a global median Pearson correlation coefficient of 0.61 (IQR = 0.44, 0.74). Figure 4c shows that river discharge reanalysis is negatively biased in 64 % of catchments (i.e. bias ratio < 1) with a global median bias ratio of 0.84 (IQR = 0.62, 1.21). In the evaluation of their global river simulation, Lin et al. (2019) consider a percentage bias within ± 20 % (equivalent to a bias ratio within 0.8 to 1.2) to be very good. Whilst only 28 % of stations meet this criterion for the GloFAS-ERA5 reanalysis, results are in line with simulations in Lin et al. (2019). The worst performing catchments (dark red KGESS dots in Fig. 5) are predominantly driven by very large positive biases (dark blue dots in Fig. 6b) in dryer rivers of the central United States, Africa, and eastern Brazil, as well as the western coast of South America; in total 12 % of catchments have a bias ratio > 2 (equivalent to a percent bias > 100 %). Figure 4d (shown spatially in Fig. 6c) shows lower variability in GloFAS-ERA5 reanalysis than observations in 61 % of catchments (i.e. variability ratio < 1), but errors in variability are less severe than bias errors with a global median variability ratio of 0.91 (IQR = 0.69, 1.15).

It is important to also look at the average magnitude of errors as a small over/under estimation in dry rivers can produce large percentage biases (and hence bias ratios). This was done by converting the units of both the reanalysis and observation time series from cubic metres per second ($\text{m}^3 \text{s}^{-1}$) to runoff depth across the catchment area in millimetres per day (mm d^{-1}) to allow direct comparison between catchments of different sizes and then computing the mean absolute error (MAE) metric (Fig. 7). The global median MAE is 0.41 mm d^{-1} (IQR = 0.18 mm d^{-1} , 0.72 mm d^{-1}). Most areas with a bias ratio > 2 (in Fig. 6b), namely much of Africa, the central United States, and eastern Brazil, have in fact a low absolute magnitude of errors given their dry locations. Other notable areas with a low absolute magnitude of errors include large parts of India, South East Asia, and Australia. There are, however, catchments on the western coast of South America, Sudan, and Ethiopia and tributaries of the River Ganges with a large MAE.

4.3 Performance by month

Figure 8 shows the global performance of GloFAS-ERA5 reanalysis for each month across all 1801 stations. Hydrological simulation skill is relatively consistent across each month with median KGESS ranging between 0.32 to 0.41

(Fig. 8a). The April to October months have the highest skill, with November to March having a higher proportion of catchments with negative skill. When the KGE' is decomposed into correlation, bias, and variability components at the monthly scale (Fig. 8b–d, respectively), it shows that the months with higher incidence of negative KGESS are driven by a higher proportion of catchments with large positive biases in those months. Correlation and variability error metrics do not vary much from one month to the next in comparison to bias errors.

Results are grouped into Northern Hemisphere ($n = 1268$ stations) and Southern Hemisphere ($n = 533$ stations) in Fig. 9. The overall GloFAS-ERA5 monthly performance in each hemisphere does not change substantially from the global analysis (Fig. 8). Nevertheless, there are some differences. The KGESS and bias ratio from the Northern Hemisphere (Fig. 9a and c, respectively) tend to follow the global analysis most strongly (i.e. Fig. 8a and c, respectively), which is not surprising given 70 % of all stations are located in the Northern Hemisphere. However, a higher proportion of Southern Hemisphere stations show large positive biases from April to June compared to November to March in the Northern Hemisphere. The largest proportion of stations with negative KGESS in the Southern Hemisphere is found from August to October (Fig. 9a). These months correspond with a lower Southern Hemisphere correlation (Fig. 9b) and a higher proportion of stations with large positive variability ratios (i.e. GloFAS-ERA5 has higher variability than observed river discharge).

4.4 Performance by catchment area

The skill of GloFAS-ERA5 river discharge reanalysis grouped into seven catchment area categories is shown in Fig. 10. In general, skill is lowest for catchments in the three categories < 10 000 km^2 with median KGESS = 0.21 ($n = 39$), 0.4 ($n = 41$), and 0.42 ($n = 53$), respectively. Performance improves as catchment size increases with median KGESS = 0.56 for catchments > 50 000 km^2 . It must be noted that results are affected by uneven samples of catchment sizes available within the GloFAS observation database, with catchments between 10 000 and 50 000 km^2 being dominant ($n = 1013$) and smaller catchments being under-represented.

4.5 Limitations

This first evaluation has found the dataset to be hydrologically skilful in the vast majority of catchments tested, although the strength of skill can vary considerably depending on location. The degradation in skill, as defined using KGESS, is the combination of (lower) correlation, (larger) bias errors, and (larger) variability errors. The evaluation provides users with an overview of the global-scale quality of the dataset, although users are advised to undertake a more

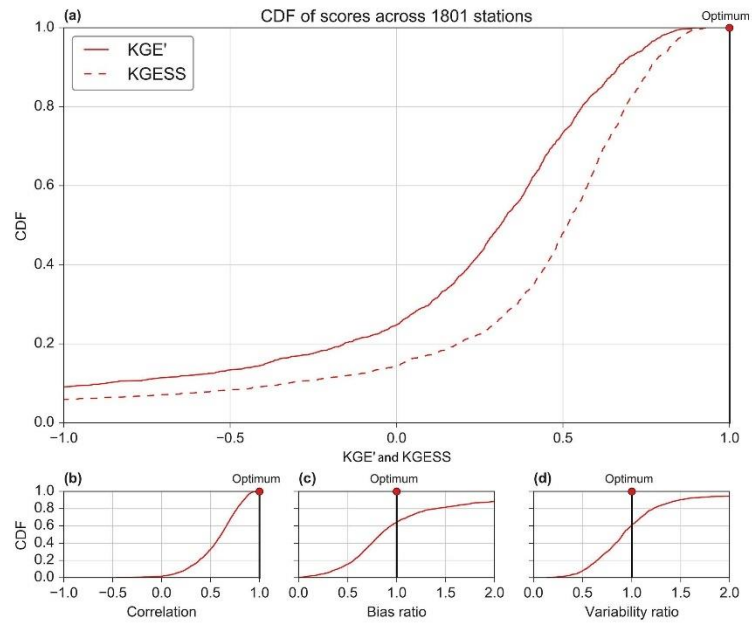


Figure 4. Cumulative distribution function (CDF) of performance metrics across all 1801 stations. Modified Kling–Gupta efficiency (KGE') and skill score (KGESS) (a) with decomposition of KGE' into Pearson correlation (b), bias ratio (c), and variability ratio (d). The red dot marks the optimum value for each metric.

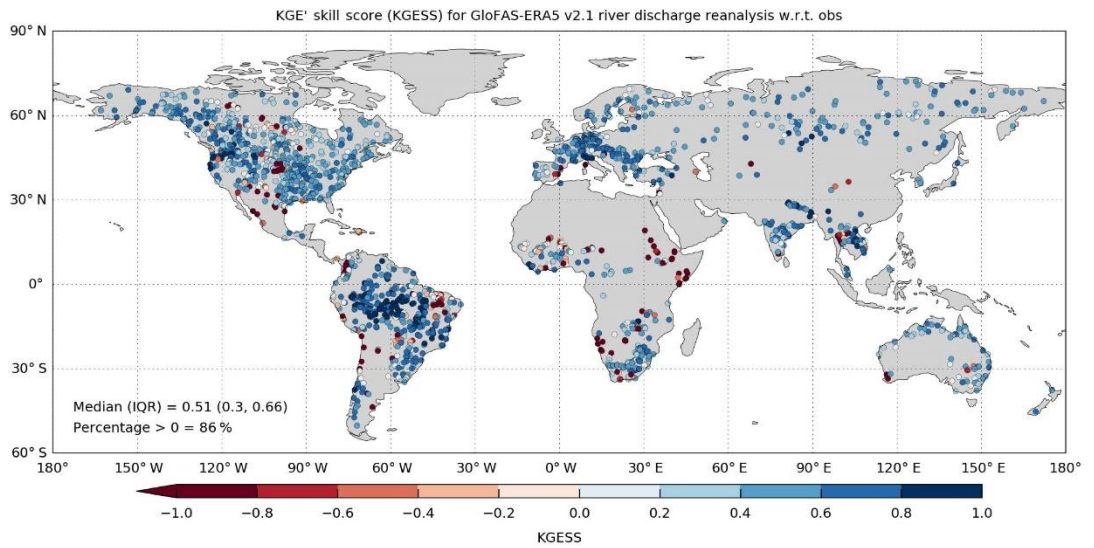


Figure 5. Modified Kling–Gupta efficiency skill score (KGESS) for GloFAS-ERA5 river discharge reanalysis against 1801 observation stations. Optimum value of KGESS is 1. Blue (red) dots show catchments with positive (negative) skill.

<https://doi.org/10.5194/essd-12-2043-2020>

Earth Syst. Sci. Data, 12, 2043–2060, 2020

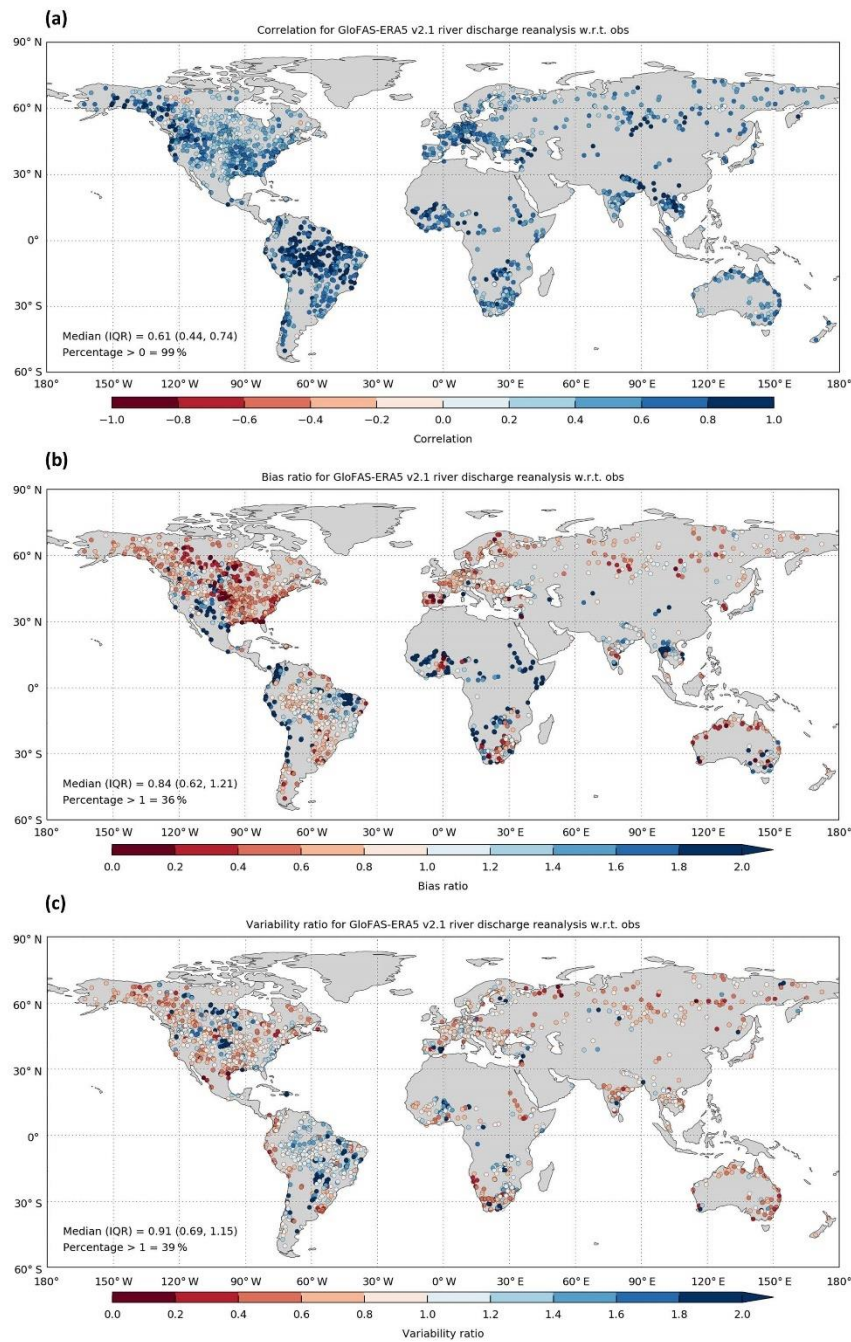


Figure 6. Decomposition of the Modified Kling–Gupta efficiency (KGE') into its three components: Pearson correlation (a), bias ratio (b), and variability ratio (c) for GloFAS-ERA5 river discharge reanalysis against 1801 observation stations. The optimum value for each of the three KGE' components is 1. Blue (red) dots represent positive (negative) values.

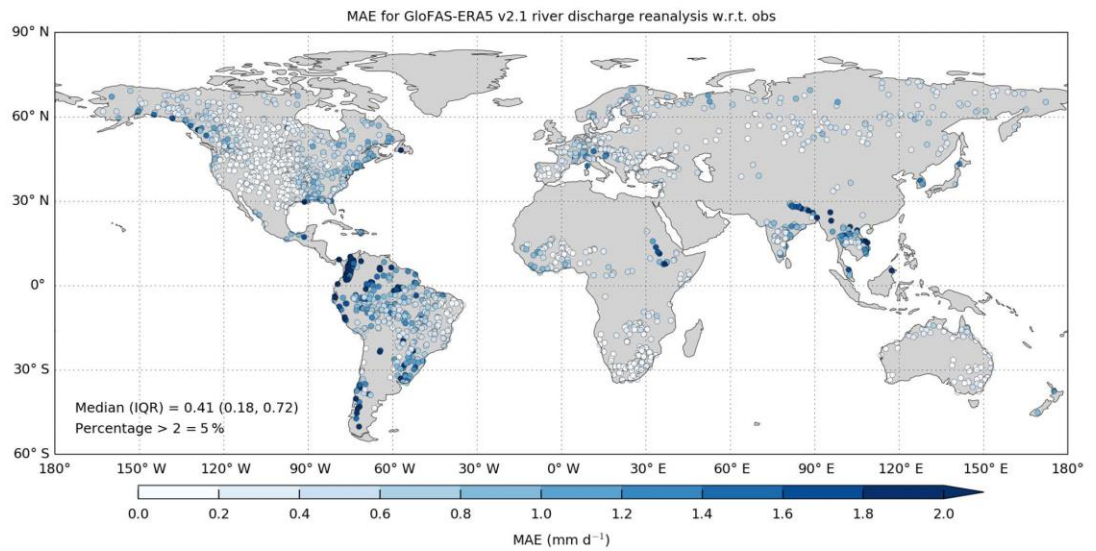


Figure 7. Mean absolute error (MAE) for GloFAS-ERA5 reanalysis against 1801 observation stations. Units for both reanalysis and observations have been converted from cubic metres per second ($\text{m}^3 \text{s}^{-1}$) to runoff depth across the catchment area (mm d^{-1}) to allow direct comparison of the magnitude of errors. Optimum value of MAE is 0; catchments with larger magnitude of errors are darker shades of blue dots.

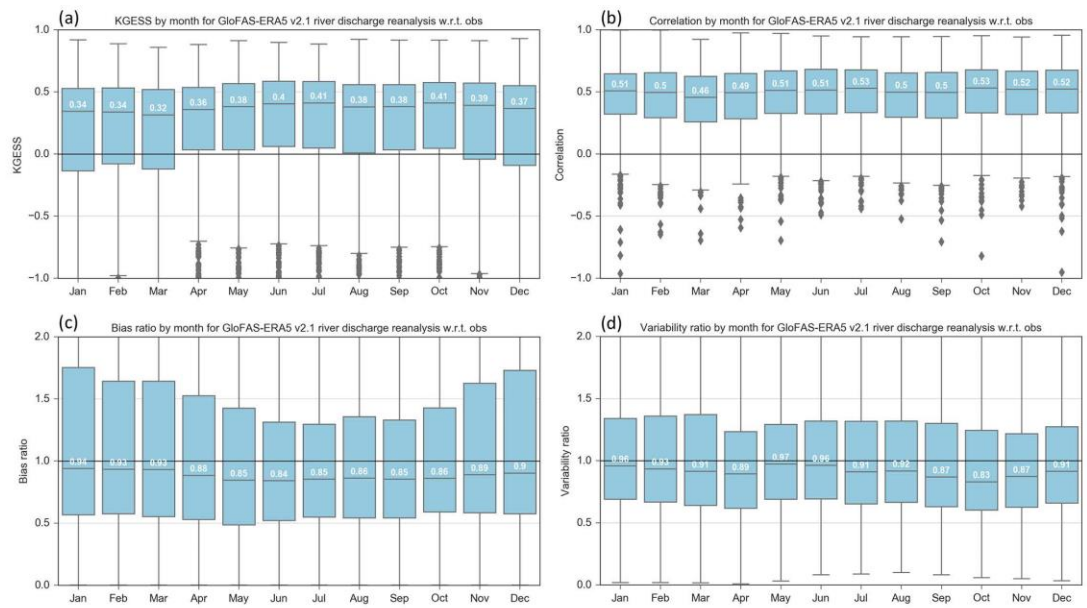


Figure 8. Performance metrics for each month for all 1801 stations. Modified Kling–Gupta efficiency skill score (KGE_{SS}) (a) with decomposition of KGE into Pearson correlation (b), bias ratio (c), and variability ratio (d). Boxes represent the IQR and horizontal grey line the median. Whiskers extend to the most extreme data point unless the data point is more than 1.5 times the IQR from the box and is instead represented as an outlier (grey diamond).

<https://doi.org/10.5194/essd-12-2043-2020>

Earth Syst. Sci. Data, 12, 2043–2060, 2020

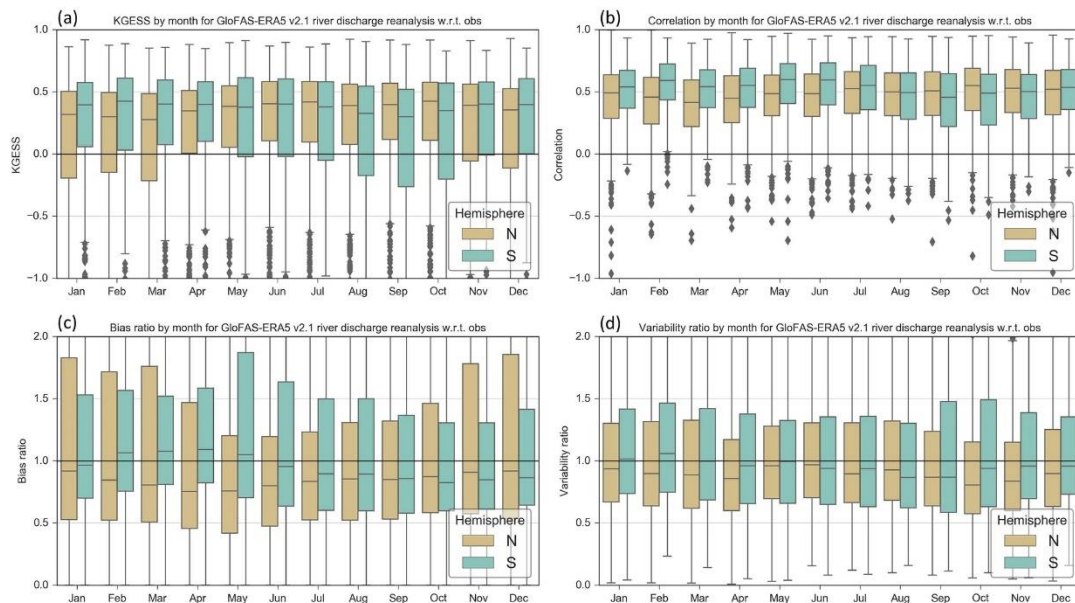


Figure 9. As in Fig. 8 but by hemisphere: Northern Hemisphere ($n = 1268$ stations) as brown boxes and Southern Hemisphere ($n = 533$ stations) as green boxes.

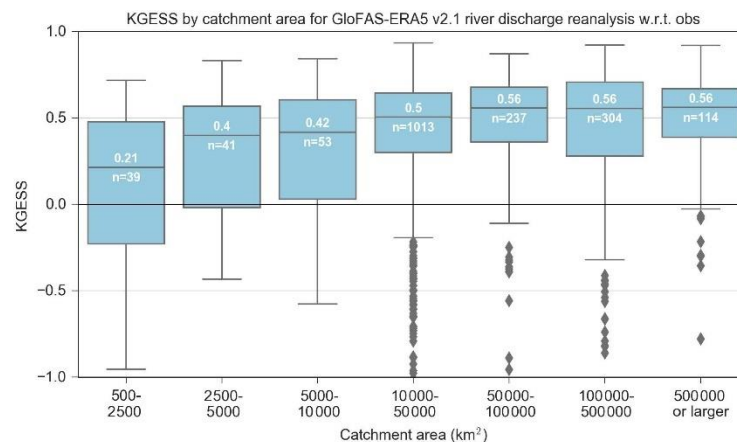


Figure 10. Modified Kling–Gupta efficiency skill score (KGESS) grouped into seven catchment area categories. Box and whisker descriptions are as in Fig. 8.

in-depth evaluation of the dataset for their region of interest. A key limitation of the dataset is the large biases identified in several regions (see above). The attribution of such biases in the GloFAS-ERA5 reanalysis is outside the scope of this data paper, but ongoing investigations such as Zsoter et al. (2019) have shown that biases can be introduced by

the real-time land data assimilation within the HTESSEL land surface model. Another expected cause of differences between river discharge reanalysis and observations is due to human modification within catchments and river channels (e.g. Harrigan et al., 2014). It is estimated that just 37 % of rivers remain free-flowing globally with the construction

DATA DESCRIPTION

Horizontal coverage	Global except for Antarctica (90N–90S, 180W–180E)
Horizontal resolution	0.1° × 0.1°
Vertical resolution	Surface level for river discharge
Temporal coverage	1979-01-01 to near real time for the most recent version.
Temporal resolution	Daily data.
Update frequency	A new river discharge reanalysis will be published with every major update of the GLOFAS system. The latest version will always be the version used in operations. For more information on the model versions, we refer to the documentation.
File format	netCDF
Data type	GRID
Versions	GloFAS v2.1

MAIN VARIABLES

Name	Units	Description
River discharge	m ³ s ⁻¹	Volume rate of water flow, including sediments, chemical and biological material, in the river channel averaged over a time step through a cross-section. The value is an average over a 24-hour period.

RELATED VARIABLES

Name	Units	Description
Upstream area	m ²	Static file - upArea.nc, Upstream area for the point in the river network

Record updated 2019-11-05 13:49:02 UTC

Figure 11. The GloFAS-ERA5 river discharge reanalysis landing page in the C3S Climate Data Store (CDS; <https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-glofas-historical?tab=overview>).

of reservoirs and dams the main contributor to loss of connectivity (Grill et al., 2019). While GloFAS-ERA5 reanalysis does represent major dams and reservoirs on the modelled river network, simplified reservoir operating parameters were used based on expert opinion (outlined in Zajac et al., 2017) due to lack of availability of global operational release records. Given the fundamental dependence of the dataset on ERA5, it would be pertinent for users to be aware of the known ERA5 issues, which can be found in the ERA5 documentation: <https://confluence.ecmwf.int/display/CKB/ERA5> (last access: 28 June 2020). In particular, “rain bombs” are known to occur from time to time in the numerical weather prediction model used by ERA5 whereby extremely large rainfall totals are generated, although these are rare (~ 10 episodes per year) and happen mostly in isolated grid points over orographic areas in Africa (Hersbach et al., 2020). However, their impact on hydrology has not been assessed. As with any reanalysis product, care must be taken when calculating long-term trends in river discharge as discontinuities may be present in the record due to changes in the global observing system entering ERA5.

5 Data availability

The GloFAS-ERA5 river discharge reanalysis is provided through the European Commission Copernicus Emergency Management Service (CEMS) and follows the Copernicus open data policy that users shall have free, full, and open access to Copernicus service information. With the drive for

open data comes challenges. In the era of big data, it is clear that traditional ways of hosting and disseminating large earth system datasets is no longer fit for purpose. An exciting development in the way large climate datasets are discovered, accessed, and used is the Copernicus Climate Change Service (C3S) Climate Data Store (CDS; <https://cds.climate.copernicus.eu/cdsapp#!/home>, last access: 28 June 2020). The CDS hosts various global and regional reanalysis products, gridded records for essential climate variables (ECVs), in which river discharge is included as a key terrestrial ECV, and much more. The CDS requires standardization of data and metadata so that datasets are more useable and discoverable through the CDS metadata pages. The CDS website provides easy access to data through user-friendly download forms. There is also a CDS Python application programming interface (API) to allow programmatic access to data. An innovative feature of the CDS is the Toolbox, which makes it easier to handle large volumes of data by allowing users to make custom applications, filter data by geographical region and date range, and finally present the data using maps and charts directly through the CDS cloud infrastructure.

The GloFAS-ERA5 river discharge reanalysis product is available on the CDS: <https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-glofas-historical?tab=overview> with the following DOI: <https://doi.org/10.24381/cds.a4fdd6b9> (C3S, 2019). The CDS landing page for the GloFAS-ERA5 reanalysis dataset is shown in Fig. 11. Both the long-term consolidated and the near-real-time intermediate reanalysis data are available in two ways. First, through the “Down-

<https://doi.org/10.5194/essd-12-2043-2020>

Earth Syst. Sci. Data, 12, 2043–2060, 2020

load data” tab whereby users can manually select options in a form for which data they would like to download. Second, data can be retrieved through the dedicated Python CDS API; an example API retrieval script is shown in Appendix B. Note that users must register for a CDS account (for free) before gaining access.

6 Conclusions

This paper outlines the production, description, evaluation, and access to the new GloFAS-ERA5 operational global river discharge reanalysis dataset available from 1979 and updated in near real time. This dataset is central to two key steps within GloFAS: (i) the calculation of flood thresholds against which real-time ensemble forecasts are compared to determine the probability of a flood signal and (ii) more consistent hydrometeorological initial conditions for the real-time flood and seasonal forecasts. The evaluation against observations showed that the product is skilful in 86% of catchments according to the modified Kling–Gupta efficiency skill score against a mean flow benchmark. However, skill varies considerably with location, with several regions such as the central United States, Africa, eastern Brazil, and the western coast of South America having large systematic positive biases. The results from the evaluation are comparable with other long-term global river discharge products (e.g. Lin et al., 2019). The attribution of such biases in the GloFAS-ERA5 reanalysis is outside the scope of this data paper, but ongoing investigations such as Zsoter et al. (2019) on the biases introduced by the real-time land data assimilation within the HTESSEL land surface model will help us to better understand existing limitations. GloFAS is an operational system which undergoes constant developments with intensive research on future versions of the model. It is foreseen that a new model version will be made operational in 2021 based on the full LISFLOOD hydrological model and an improved model calibration (Alfieri et al., 2020).

The long-term and operational nature of the GloFAS-ERA5 reanalysis dataset opens avenues for further applications. Forecast evaluation activities within GloFAS now include skill assessment over longer time periods and has allowed a new operational forecast verification suite to be developed whereby the performance of the forecasts can be tracked in near real time for every river in the world. Other applications are envisaged for monitoring the global status of flood and drought conditions, the identification of hydroclimatic variability and change, and as raw input for post-processing and machine learning methods that can add further value.

Appendix A

Standard NetCDF header metadata information for example file

```

$ ncdump -h CEMS_ECMWF_dis24_20190101_glofas_v2.1.nc

netcdf CEMS_ECMWF_dis24_20190101_glofas_v2.1 {
dimensions:
  time = UNLIMITED ; // (1 currently)
  lon = 3600 ;
  lat = 1500 ;
variables:
  double time(time) ;
    time:standard_name = "time" ;
    time:long_name = "time" ;
    time:units = "hours since 1979-01-01 00:00:00" ;
    time:calendar = "standard" ;
    time:axis = "T" ;
  double lon(lon) ;
    lon:standard_name = "longitude" ;
    lon:long_name = "longitude" ;
    lon:units = "degrees_east" ;
    lon:axis = "X" ;
  double lat(lat) ;
    lat:standard_name = "latitude" ;
    lat:long_name = "latitude" ;
    lat:units = "degrees_north" ;
    lat:axis = "Y" ;
  float dis24(time, lat, lon) ;
    dis24:long_name = "mean discharge in the last 24 hours" ;
    dis24:units = "m3/s" ;
    dis24:FillValue = 1.e+20f ;
    dis24:missing_value = 1.e+20f ;

```

The 'ncdump' command-line utility converts NetCDF data to human-readable text form

Summary of number of time and space dimensions. This file is for 1 day

The 'time' variable with units in hours since reference time. In this e.g. it is 350640 h (or 14,610 d)

Longitude ('lon') and latitude ('lat') variables for grid cells in the GloFAS domain

Primary variable: mean daily river discharge ('dis24') with dimensions ('time', 'lat', 'lon'). Units & missing value also shown

Appendix B

```
# Example CDS Python API request script

# Code snippets can be found by clicking 'Show API request' at
# bottom of GloFAS-ERA5 reanalysis download form:
# https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-glofas-historical?tab=form

# Instructions on how to download CDS API can be found here:
# https://cds.climate.copernicus.eu/api-how-to

import cdsapi

c = cdsapi.Client()

# Example download consolidated data (GloFAS) for 31 December 2018 (note: date stamp
# represents end of 24 h averaging period)

c.retrieve(
    'cems-glofas-historical',
    {
        'variable': 'River discharge',
        'dataset': 'Consolidated reanalysis',
        'version': '2.1',
        'year': '2019',
        'month': '01',
        'day': '01',
        'format': 'tgz'
    },
    'download.tar.gz')

# Example download near real time intermediate data (GloFAST) for 12 November 2019 (note:
# date stamp represent end of 24 h averaging period)

c.retrieve(
    'cems-glofas-historical',
    {
        'variable': 'River discharge',
        'dataset': 'Intermediate dataset',
        'version': '2.1',
        'year': '2019',
        'month': '11',
        'day': '13',
        'format': 'tgz'
    },
    'download.tar.gz')
```


Supplement. The supplement related to this article is available online at: <https://doi.org/10.5194/essd-12-2043-2020-supplement>.

Author contributions. SH drafted the paper and performed the evaluation. EZ wrote the suite to produce the dataset. CB adapted the suite to produce the dataset operationally. FW and CB were responsible for the integration of the dataset into the Climate Data Store. LA, CP, PS, HC, and FP helped frame the paper. All co-authors contributed to the editing of the paper and to the discussion and interpretation of results.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We thank colleagues from the Copernicus Climate Change Service (C3S) for helping with ingesting the dataset into the Climate Data Store (CDS) and Cinzia Mazzetti (ECMWF) and Domenico Nappo (JRC) for helpful discussions during the revision regarding LISFLOOD. The providers of observed river discharge observations are greatly thanked, as well as both GloFAS partners and the Global Runoff Data Centre (GRDC), 56068 Koblenz, Germany.

Financial support. This research has been supported by the European Commission Copernicus Emergency Management Service (CEMS) (grant no. 198702).

Review statement. This paper was edited by David Carlson and reviewed by two anonymous referees.

References

- Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger, F.: 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>, 2013.
- Alfieri, L., Lorini, V., Hirpa, F. A., Harrigan, S., Zsoter, E., Prudhomme, C., and Salamon, P.: A global streamflow reanalysis for 1980–2018, *J. Hydrol.*, 6, 100049, <https://doi.org/10.1016/j.jhydroa.2019.100049>, 2020.
- Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., and Betts, A. K.: 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, 623–643, <https://doi.org/10.1175/2008JHM1068.1>, 2009.
- Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Dutra, E., Fink, G., Orth, R., and Schellekens, J.: Global evaluation of runoff from 10 state-of-the-art hydrological models, *Hydrol. Earth Syst. Sci.*, 21, 2881–2903, <https://doi.org/10.5194/hess-21-2881-2017>, 2017.
- Burek, P., van der Knijff, J. M., and de Roo, A. P. J. D.: LISFLOOD – Distributed Water Balance and Flood Simulation Model – Revised User Manual, Publications Office of the European Union, <https://doi.org/10.2788/24719>, 2013.
- C3S: River discharge and related historical data from the Global Flood Awareness System, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), <https://doi.org/10.24381/cds.a4fdd6b9>, 2019.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Källberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. Roy. Meteor. Soc.*, 137, 553–597, <https://doi.org/10.1002/qj.828>, 2011.
- de Rosnay, P., Balsamo, G., Albergel, C., Muñoz-Sabater, J., and Isaksen, L.: Initialisation of Land Surface Variables for Numerical Weather Prediction, *Surv. Geophys.*, 35, 607–621, <https://doi.org/10.1007/s10712-012-9207-x>, 2014.
- Döll, P., Kaspar, F., and Lehner, B.: A global hydrological model for deriving water availability indicators: model tuning and validation, *J. Hydrol.*, 270, 105–134, [https://doi.org/10.1016/S0022-1694\(02\)00283-4](https://doi.org/10.1016/S0022-1694(02)00283-4), 2003.
- 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.: Continental and global scale flood forecasting systems, *WIREs Water*, 3, 391–418, <https://doi.org/10.1002/wat2.1137>, 2016.
- Emerton, R., Zsoter, E., Arnal, L., Cloke, H. L., Muraro, D., Prudhomme, C., Stephens, E. M., Salamon, P., and Pappenberger, F.: Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0, *Geosci. Model Dev.*, 11, 3327–3346, <https://doi.org/10.5194/gmd-11-3327-2018>, 2018.
- Fekete, B. M., Vörösmarty, C. J., and Grabs, W.: High-resolution fields of global runoff combining observed river discharge and simulated water balances, *Global Biogeochem. Cy.*, 16, 15–15-10, <https://doi.org/10.1029/1999GB001254>, 2002.
- Ghiggi, G., Humphrey, V., Seneviratne, S. I., and Gudmundsson, L.: GRUN: an observation-based global gridded runoff dataset from 1902 to 2014, *Earth Syst. Sci. Data*, 11, 1655–1674, <https://doi.org/10.5194/essd-11-1655-2019>, 2019.
- Grill, G., Lehner, B., Thieme, M., Geenen, B., Tickner, D., Antonelli, F., Babu, S., Borrelli, P., Cheng, L., Crochetiere, H., Macedo, H. E., Filgueiras, R., Goichot, M., Higgins, J., Hogan, Z., Lip, B., McClain, M. E., Meng, J., Mulligan, M., Nilsson, C., Olden, J. D., Opperman, J. J., Petry, P., Liermann, C. R., Sáenz, L., Salinas-Rodríguez, S., Schelle, P., Schmitt, R. J. P., Snider, J., Tan, F., Tockner, K., Valdujo, P. H., van Soesbergen, A., and Zarfl, C.: Mapping the world's free-flowing rivers, *Nature*, 569, 215–221, <https://doi.org/10.1038/s41586-019-1111-9>, 2019.
- Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, 377, 80–91, <https://doi.org/10.1016/j.jhydrol.2009.08.003>, 2009.
- Harrigan, S., Murphy, C., Hall, J., Wilby, R. L., and Sweeney, J.: Attribution of detected changes in streamflow using multiple

<https://doi.org/10.5194/essd-12-2043-2020>

Earth Syst. Sci. Data, 12, 2043–2060, 2020

- working hypotheses, *Hydrol. Earth Syst. Sci.*, 18, 1935–1952, <https://doi.org/10.5194/hess-18-1935-2014>, 2014.
- Harrigan, S., Prudhomme, C., Parry, S., Smith, K., and Tanguy, M.: Benchmarking ensemble streamflow prediction skill in the UK, *Hydrol. Earth Syst. Sci.*, 22, 2023–2039, <https://doi.org/10.5194/hess-22-2023-2018>, 2018.
- 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., Chiara, G. D., 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., Rosnay, P. de, Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 Global Reanalysis, *Q. J. Roy. Meteor. Soc.*, 146, 1999–2049, <https://doi.org/10.1002/qj.3803>, 2020.
- Hirpa, F. A., Salamon, P., Beck, H. E., Lorini, V., Alfieri, L., Zsoter, E., and Dadson, S. J.: 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>, 2018.
- Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, *J. Hydrol.*, 424–425, 264–277, <https://doi.org/10.1016/j.jhydrol.2012.01.011>, 2012.
- Knoben, W. J. M., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, *Hydrol. Earth Syst. Sci.*, 23, 4323–4331, <https://doi.org/10.5194/hess-23-4323-2019>, 2019.
- Lavers, D., Harrigan, S., Andersson, E., Richardson, D. S., Prudhomme, C., and Pappenberger, F.: A vision for improving global flood forecasting, *Environ. Res. Lett.*, 14, 121002, <https://doi.org/10.1088/1748-9326/ab52b2>, 2019.
- 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.: Global Reconstruction of Naturalized River Flows at 2.94 Million Reaches, *Water Resour. Res.*, 55, 6499–6516, <https://doi.org/10.1029/2019WR025287>, 2019.
- Magnusson, L., Zsoter, E., Prudhomme, C., Baugh, C., Harrigan, S., Ficchi, A., Emerton, R., Cloke, H., Stephens, L., and Speight, L.: ECMWF works with universities to support response to tropical cyclone Irai, *ECMWF Newsletter*, 160, 2–3, 2019.
- Pappenberger, F., Cloke, H. L., Balsamo, G., Ngo-Duc, T., and Oki, T.: Global runoff routing with the hydrological component of the ECMWF NWP system, *Int. J. Climatol.*, 30, 2155–2174, <https://doi.org/10.1002/joc.2028>, 2010.
- Qian, T., Dai, A., Trenberth, K. E., and Oleson, K. W.: Simulation of Global Land Surface Conditions from 1948 to 2004. Part I: Forcing Data and Evaluations, *J. Hydrometeorol.*, 7, 953–975, <https://doi.org/10.1175/JHM540.1>, 2006.
- Reichle, R. H., Koster, R. D., De Lannoy, G. J. M., Forman, B. A., Liu, Q., Mahanama, S. P. P., and Touré, A.: Assessment and Enhancement of MERRA Land Surface Hydrology Estimates, *J. Climate*, 24, 6322–6338, <https://doi.org/10.1175/JCLI-D-10-05033.1>, 2011.
- Sperna Weiland, F. C., van Beek, L. P. H., Kwadijk, J. C. J., and Bierkens, M. F. P.: The ability of a GCM-forced hydrological model to reproduce global discharge variability, *Hydrol. Earth Syst. Sci.*, 14, 1595–1621, <https://doi.org/10.5194/hess-14-1595-2010>, 2010.
- UNDRR: Sendai Framework for Disaster Risk Reduction 2015–2030, United Nations Office for Disaster Risk Reduction, Geneva, available at: <https://www.unisdr.org/we/inform/publications/43291> (last access: 30 October 2019), 2015.
- van der Knijff, J. M., Younis, J., and de Roo, A. P. J. D.: LIS-FLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation, *Int. J. Geogr. Inf. Sci.*, 24, 189–212, <https://doi.org/10.1080/13658810802549154>, 2010.
- Yamazaki, D., Kanae, S., Kim, H., and Oki, T.: A physically based description of floodplain inundation dynamics in a global river routing model, *Water Resour. Res.*, 47, W04501, <https://doi.org/10.1029/2010WR009726>, 2011.
- Zajac, Z., Revilla-Romero, B., Salamon, P., Burek, P., Hirpa, F. A., and Beck, H.: The impact of lake and reservoir parameterization on global streamflow simulation, *J. Hydrol.*, 548, 552–568, <https://doi.org/10.1016/j.jhydrol.2017.03.022>, 2017.
- Zsoter, E., Cloke, H., Stephens, E., de Rosnay, P., Muñoz-Sabater, J., Prudhomme, C., and Pappenberger, F.: How Well Do Operational Numerical Weather Prediction Configurations Represent Hydrology?, *J. Hydrometeorol.*, 20, 1533–1552, <https://doi.org/10.1175/JHM-D-18-0086.1>, 2019.

A7: Daily ensemble discharge reforecasts and real-time forecasts from the operational Global Flood Awareness System

This paper presents a co-author contribution arising through collaboration during this PhD, and has the following reference:

Harrigan, S., E. Zsoter, H. Cloke, P. Salamon and C. Prudhomme, 2023: 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-27-1-2023>

Hydrol. Earth Syst. Sci., 27, 1–19, 2023
<https://doi.org/10.5194/hess-27-1-2023>
 © Author(s) 2023. This work is distributed under
 the Creative Commons Attribution 4.0 License.



Hydrology and
 Earth System
 Sciences 

Daily ensemble river discharge reforecasts and real-time forecasts from the operational Global Flood Awareness System

Shaun Harrigan¹, Ervin Zsoter^{1,2}, Hannah Cloke^{2,3,4,5}, Peter Salamon⁶, and Christel Prudhomme^{1,7,8}

¹Forecast Department, European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK

²Department of Geography and Environmental Science, University of Reading, Reading, UK

³Department of Meteorology, University of Reading, Reading, UK

⁴Department of Earth Sciences, Uppsala University, Uppsala, Sweden

⁵Centre of Natural Hazards and Disaster Science, CNDS, Uppsala, Sweden

⁶European Commission, Joint Research Centre (JRC), Ispra, Italy

⁷Centre for Ecology and Hydrology (CEH), Wallingford, UK

⁸Department of Geography and Environment, University of Loughborough, Loughborough, UK

Correspondence: Shaun Harrigan (shaun.harrigan@ecmwf.int)

Received: 13 October 2020 – Discussion started: 16 October 2020

Revised: 26 March 2022 – Accepted: 1 July 2022 – Published: 2 January 2023

Abstract. Operational global-scale hydrological forecasting systems are used to help manage hydrological extremes such as floods and droughts. The vast amounts of raw data that underpin forecast systems and the ability to generate information on forecast skill have, until now, not been publicly available. As part of the Global Flood Awareness System (GloFAS; <https://www.globalfloods.eu/>, last access: 3 December 2022) service evolution, in this paper daily ensemble river discharge reforecasts and real-time forecast datasets are made free and openly available through the Copernicus Climate Change Service (C3S) Climate Data Store (CDS). They include real-time forecast data starting on 1 January 2020 updated operationally every day and a 20-year set of reforecasts and associated metadata. This paper describes the model components and configuration used to generate the real-time river discharge forecasts and the reforecasts. An evaluation of ensemble forecast skill using the continuous ranked probability skill score (CRPSS) was also undertaken for river points around the globe. Results show that GloFAS is skilful in over 93 % of catchments in the short (1 to 3 d) and medium range (5 to 15 d) 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. However, the strength of skill varies considerably by location with GloFAS found to have no or negative skill at longer lead times in broad hydroclimatic regions in tropi-

cal Africa, western coast of South America, and catchments dominated by snow and ice in high northern latitudes. Forecast skill is summarised as a new headline skill score available as a new layer on the GloFAS forecast Web Map Viewer to aid user interpretation and understanding of forecast quality.

1 Introduction

Hydrological extremes, such as floods and droughts, have severe negative socio-economic impacts, and climate change is expected to alter their timing and magnitude (Blöschl et al., 2017, 2019; Ward et al., 2020). Since 1990, reported disasters have led to over 94 million people being affected by flooding each year, and economic losses are estimated at around USD 260–310 billion per year (UNDRR, 2015a). The need to reduce this risk has been identified under the Sendai Framework for Disaster Risk Reduction (UNDRR, 2015b). One of the primary methods of achieving DRR and building resilience in society is through early warning of extreme events. There are now several centres producing global- and continental-scale hydrological forecasts operationally which are working to support national forecasting and decision-making in the water sector (Emerton et al., 2016). In Europe, there is the European Flood Awareness

Published by Copernicus Publications on behalf of the European Geosciences Union.

System (EFAS; <https://www.efas.eu/en>, last access: 3 December 2022; Thielen et al., 2009) and the European Hydrological Predictions for the Environment (E-HYPE; <https://hypeweb.smhi.se/>, last access: 3 December 2022; Donnelly et al., 2016), in the US the Hydrologic Ensemble Forecast Service (HEFS; <https://water.weather.gov/ahps/>, last access: 3 December 2022; Demargne et al., 2014), and in Australia the Flood Forecasting and Warning Service (FFWS; <http://www.bom.gov.au/water/>, last access: 3 December 2022). For global-scale systems there is the Global Flood Awareness System (GloFAS; <https://www.globalfloods.eu/>, last access: 3 December 2022; Alfieri et al., 2013) and World-Wide HYPE (WWH; <https://hypeweb.smhi.se/>, last access: 3 December 2022; Arheimer et al., 2020).

GloFAS is the global flood service of the European Commission's Copernicus Emergency Management Service (CEMS), an operational system for monitoring and forecasting floods across the world with over 6000 registered users in March 2021. The service and data are available through a free and open license and the system is designed to help decision makers and forecasters in sectors such as national and international water authorities, water resources managers, hydropower companies, civil protection authorities, and international humanitarian aid organisations. GloFAS is not designed to be a replacement for local operational hydrological forecasting systems; in many parts of the world, however, a local or national system for operational forecasts of river discharge does not yet exist so it might be the only information available. GloFAS covers all river basins out to medium- and extended-range lead times (30 d ahead) and updated daily, with GloFAS-Seasonal (Emerton et al., 2018) updated monthly out to a 16-week lead time. Therefore, it has been used to complement local forecast systems by allowing forecasters to gain information on surrounding and upstream basins, monitoring for potential flood signals where advanced warning is needed.

GloFAS can be used for providing daily assessments of potential upcoming flood events for the whole globe, such a spatio-temporal consistent overview is required by several users. For example, GloFAS is used daily as the main information source to monitor existing and upcoming river flood events and report back potential risks of flood impacts to the Emergency Response Coordination Centre (ERCC) of the European Commission, as part of the Aristotle-ENHSP project (European Natural Hazard Scientific Partnership, <http://aristotle.ingv.it/tiki-index.php>, last access: 10 September 2020). Example real-world use cases of GloFAS include supporting the humanitarian response to the devastating floods that affected large parts of Mozambique, Malawi, and Zimbabwe in the wake of tropical cyclones Idai in March 2019 following a request from the Department for International Development of the UK government (Magnusson et al., 2019; Emerton et al., 2020) and during the 2020 monsoon season by the Bangladesh Flood Forecasting and Warning Centre (FFWC) (Hossain et al., 2020).

GloFAS has been developed together by the Joint Research Centre (JRC) of the European Commission, the University of Reading, and the European Centre for Medium-Range Weather Forecasts (ECMWF), and was originally designed for large river basins and transboundary rivers. The system went pre-operational in July 2011 (Alfieri et al., 2013), becoming a fully operational 24/7 supported service in April 2018 (version 1.0, upgraded to version 2.0 in November 2018). GloFAS version 2.1 was released on 5 November 2019 (GloFAS user wiki: <https://confluence.ecmwf.int/display/CEMS/GloFAS+versioning+system>, last access: 16 December 2022).

There are two major gaps in data service delivery of the current generation of global hydrological forecasting, including GloFAS: Firstly, forecasts are generally issued as post-processed information (e.g. focusing on river discharge exceeding pre-defined flood thresholds) shown as maps and graphics on a dedicated web interface, but the raw data are not readily available to users. Having fast access to post-processed information has the advantage of providing an overview of the forecast output as an active flood event unfolds. However, not also having direct access to the raw data precludes the use in further downstream applications (e.g. impact modelling, multi-model forecast systems, production of value-added products for specific sectors such as river transport and hydropower industries, and advancement in techniques requiring large-scale datasets such as machine learning). Secondly, "reforecasts" (i.e. forecasts for a set of past dates, also known as hindcasts) as consistent as possible with the real-time forecasting system, ideally updated for each major model cycle upgrade, have not been made publicly available, limiting both global and user-specific local evaluation of forecast skill.

As part of the continued evolution of GloFAS in light of the aforementioned service gaps, the GloFAS real-time forecasts and a long-term and large-sample set of reforecasts was made available to users as part of the release of GloFAS version 2.2 on 2 December 2020 (<https://confluence.ecmwf.int/display/CEMS/GloFAS+v2.2>, last access: 16 December 2022). This paper describes how the GloFAS forecast datasets (real-time and reforecasts) are generated, the methodology implemented for the forecast skill evaluation, and provides a global overview of the forecast skill assessment results that form the scientific basis for a new headline forecast skill layer on the GloFAS Web Map Viewer.

2 GloFAS components, configuration and data

The GloFAS hydrological forecasting system couples global numerical weather prediction (NWP) with hydrological modelling to produce ensemble forecasts of river discharge operationally each day across the world. The key model components of GloFAS version 2.1 (identical to version 2.2, the latter being a minor service-only upgrade with increased data

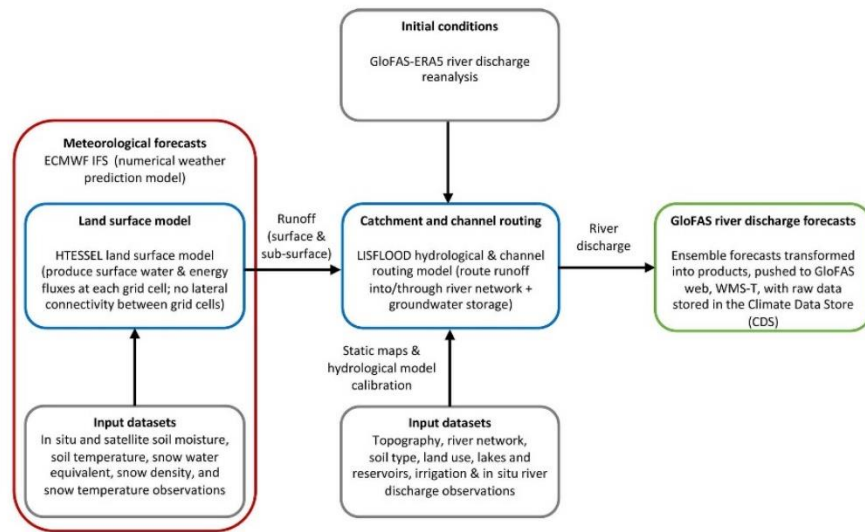


Figure 1. Key components of GloFAS version 2.1/2.2.

availability and new information layers added to the service) are shown in Fig. 1 with a summary of real-time forecast (Sect. 2.1), reforecast (Sect. 2.2), and reanalysis (Sect. 2.3) configurations in Table 1. Individual GloFAS model components have already been published in the scientific literature and hence are not described in detail here (Table A1).

2.1 GloFAS real-time forecasts

GloFAS is driven by the NWP model of the European Centre for Medium-Range Weather Forecasts (ECMWF), known as the Integrated Forecasting System (IFS). The current operational IFS model cycle is 47r3, implemented on 12 October 2021 (<https://www.ecmwf.int/en/forecasts/about-our-forecasts/evolution-ifs/cycles/summary-cycle-47r3>, last access: 25 March 2022). Because the atmosphere is a chaotic system, ECMWF ensemble forecasts (ENS) are used to account for the inherent uncertainty and provide probabilistic forecasts in GloFAS (Fig. 2). ECMWF ENS (~ 18 km horizontal resolution) produces 51 ensemble members operationally out to a lead time of 15 d twice per day at 00:00 and 12:00 UTC. Ensemble members are comprised of a single “control” (CTL) member which is generated from the most accurate estimate of current conditions and the remaining 50 members which have their initial conditions perturbed to provide a range of possible future weather states. Twice per week (on Monday and Thursday) ECMWF ENS is extended to run to 46 d ahead at a coarser resolution (~ 36 km horizontal resolution), although in GloFAS only days 16 to 30 are used. The ECMWF ENS is run at a 6-

hourly forecast time step and for ingestion into the GloFAS hydrological modelling chain, data from the 00:00 UTC run are extracted and aggregated to 24-hourly time step.

The hydrological modelling components of GloFAS (Fig. 1) comprises the land surface model of ECMWF IFS, HTESSSEL (Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land; Balsamo et al., 2009), and LISFLOOD, a spatially distributed grid-based hydrological and channel routing model (van der Knijff et al., 2010). Precipitation is transformed to surface and subsurface runoff in HTESSSEL, with groundwater and channel routing processes simulated in LISFLOOD. In HTESSSEL, excess precipitation and snowmelt are partitioned as surface runoff or infiltrated into a four-layer soil column (7 cm depth for top layer and then 21, 72, and 189 cm) at each IFS grid cell, before draining from the bottom of the soil column as subsurface runoff.

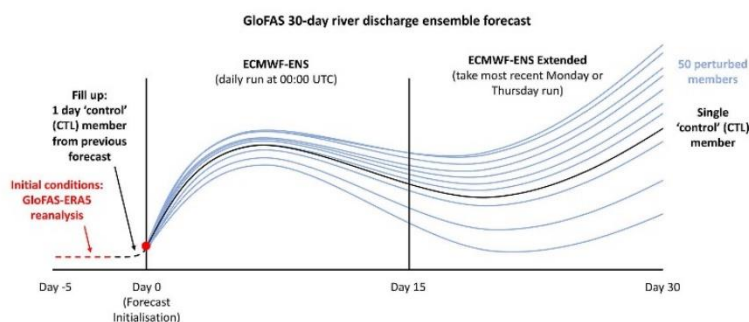
Output from HTESSSEL is downscaled to the GloFAS 0.1° (~ 11 km) gridded river network using the nearest neighbour method before being input to LISFLOOD. Surface runoff is then routed through the river network using the kinematic wave approach. Subsurface runoff is used as input to the LISFLOOD groundwater module representing both base flow and faster groundwater pathways; it consists of two parallel linear reservoirs (upper zone for quick and lower zone for slower groundwater flow) that store and subsequently transport water to the river channel with a time delay. Groundwater and river routing parameters were calibrated against river discharge observations for 1287 catchments globally by Hirpa et al. (2018). A total of 463 of the largest lakes (surface area > 100 km²) and 667 largest reservoirs have been incor-

Table 1. Real-time forecast and reforecast configurations for GloFAS version 2.1/2.2.

	GloFAS real-time forecasts	GloFAS reforecasts
GloFAS version ^a	2.1/2.2	2.1/2.2
ECMWF IFS version ^b (including HTESSEL)	46r1 (5 November 2019 to 29 June 2020) 47r1 (30 June 2020) 47r2 (11 May 2021) 47r3 (12 October 2021)	45r1 (1 January to 10 June 2019) 46r1 (11 June to 31 December 2019)
LISFLOOD version and calibration	Hirpa et al. (2018)	Hirpa et al. (2018)
Hydrological forecast initialisation	Latest GloFAS-ERA5 reanalysis with any temporal gap until real-time (i.e. “fill-up”) with control (CTL) member of ENS	GloFAS-ERA5
Variable	River discharge ($\text{m}^3 \text{s}^{-1}$) in the last 24 h	River discharge ($\text{m}^3 \text{s}^{-1}$) in the last 24 h
Time step	24 h	24 h
Horizontal resolution	0.1° (~ 11 km at equator)	0.1° (~ 11 km at equator)
Lead time	1 to 30 d	1 to 30 d
Number of ensemble members	51	11
(Re)forecast frequency	Daily at 00:00 UTC dates per year)	Twice per week (Mondays and Thursdays) at 00:00 UTC (104 start dates per year)
(Re)forecast period	5 November 2019 to present	January 1999 to December 2018

^a <https://confluence.ecmwf.int/display/CEMS/GloFAS+versioning+system> (last access: 3 December 2022);

^b <https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model> (last access: 3 December 2022).

**Figure 2.** Schematic of a single GloFAS 30 d real-time river discharge ensemble forecast initialised at day 0 for GloFAS version 2.1/2.2.

porated into the GloFAS river network (Zajac et al., 2017). Reservoir outflow is calculated with a set of four rules depending on the current reservoir filling level (see Burek et al., 2013).

GloFAS real-time river discharge forecasts are produced operationally once per day using the ECMWF ENS initialised at 00:00 UTC (Fig. 2). Initial hydrometeorological conditions are provided by the latest near-real-time GloFAS-ERA5 river discharge reanalysis (Harrigan et al., 2020a and

Sect. 2.3), a product publicly available to users 2 to 5 d behind real time through the CDS. To fill this 2 to 5 d gap between the latest available GloFAS-ERA5 data and real-time initialisation of the GloFAS forecast, the first 24 h period from the single ECMWF ENS CTL member from the preceding day's forecast is used as “fill-up” (see Fig. 2).

The final stage in the real-time forecast production is to generate plots and maps from the raw data highlighting possible upcoming flood events (see <https://confluence.ecmwf>.

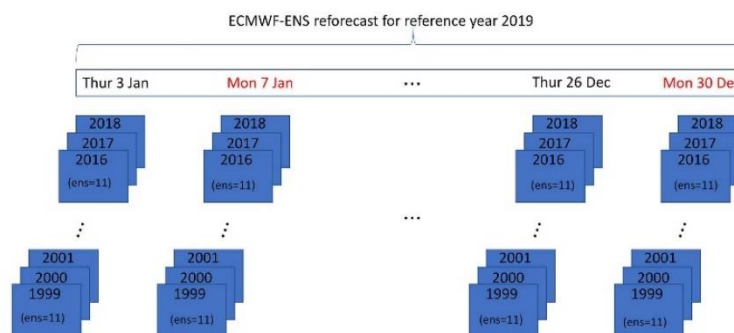


Figure 3. ECMWF-ENS reforecast schematic for the reference period January to December 2019 with 11 ensemble members (ens = 11).

int/display/CEMS/Overall+GloFAS+products+summary, last access: 25 March 2022) for a complete description of all GloFAS products. These products are pushed each day to the GloFAS Web Map Viewer and are freely available to users (<https://www.globalfloods.eu/>, last access: 3 December 2022 and available as Web Map Service with temporal requests each day, WMS-T; <https://confluence.ecmwf.int/display/CEMS/Web+Services>, last access: 25 March 2022). The raw real-time forecast GloFAS river discharge data, together with corresponding metadata, are then stored in the user-friendly data repository, the Copernicus Climate Data Store (CDS; <https://cds.climate.copernicus.eu#!/home>, last access: 25 March 2022) for use in downstream applications. Full details on data access can be found in Sect. 4.3.

2.2 GloFAS reforecasts

The quality of any forecast system can be evaluated by comparing a set of past forecasts with their corresponding observations (Jolliffe and Stephenson, 2012; Wilks, 2011). The set of past forecasts can be forecasts from the operational forecast system or a dedicated set of “reforecasts” (also known as hindcasts) that are computed retrospectively using the same (or as close as possible) model as the real-time forecast for a number of past dates.

The set of past forecasts used to evaluate the skill of GloFAS (Sect. 3) were generated from the ECMWF 20-year operational reforecasts. Compared with archived forecasts, using reforecasts has the advantage of being generated from the latest NWP configuration, which is generally more stable than archived forecasts produced from different model cycles. Typically, there is a new ECMWF IFS cycle release every 6 to 18 months (Table 1). In addition, changes can be made to the IFS or GloFAS modelling system components independently from the full ECMWF IFS cycle release (see <https://confluence.ecmwf.int/display/CEMS/GloFAS+versioning+system>, last access: 3 December 2022, for a description of GloFAS release cycles since its opera-

tional launch). The last 10 years of GloFAS archived forecasts contain at least 19 different ECMWF IFS model evolutions. Whilst not all IFS model changes impact the terrestrial water cycle, it is likely that there are significant changes in forecast errors between each model evolution, making the evaluation inconsistent through time. In contrast, the use of reforecasts has a number of advantages compared to using archived forecasts for forecast performance evaluation: (1) being run off-line, the latest hydrological routing component and simulation configuration (e.g. initial conditions) can be used, providing a stable simulation of river discharge processes; and (2) a large sample of ECMWF-ENS reforecasts are available (20 years long), albeit with a smaller-sized ensemble than the real-time simulation (11 members instead of 51), allowing for robust evaluation of forecast skill.

ECMWF uses an “on-the-fly” configuration to generate a continuous large reforecast sample, while balancing the computational resources needed to run the operational global NWP. A reforecast task is run twice per week (on Mondays and Thursdays) in parallel to the real-time forecast, using ERA5 atmospheric reanalysis (Hersbach et al., 2020) for initial conditions of past dates. A reforecast of the corresponding date for the previous 20 years is produced with a reduced number of 11 ensemble members but using the same model version as real-time (Vitart, 2014). For example, on Thursday 3 January 2019 a real time forecast based on IFS cycle 45r1 as well as a retrospective reforecast for 3 January for 20 years in the past (i.e. 3 January 1999 to 3 January 2018) was produced and archived. On Monday 7 January 2019, the process was repeated with reforecasts run for 7 January 1999 to 7 January 2018, and so on each Monday and Thursday operationally (Fig. 3).

The GloFAS reforecast used here and made available was generated during the full calendar year of 2019 (i.e. Thursday 3 January to Monday 30 December). It is an ensemble containing forecast simulations of 104 start dates per year for the previous 20 years 1999 to 2018 (2080 start dates in total) composed of 11 members and running for lead times 1 to 30 d

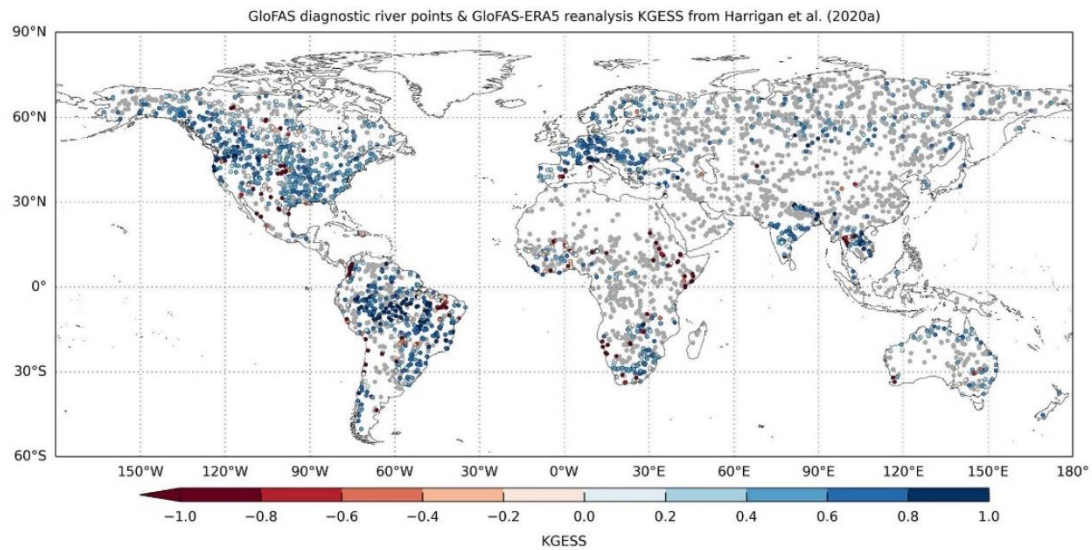


Figure 4. GloFAS diagnostic river points ($n = 5997$) are highlighted by grey dots. Coloured dots show hydrological performance of GloFAS-ERA5 river discharge reanalysis against a subset of GloFAS diagnostic river points with observations ($n = 1801$) from Harrigan et al. (2020a) using the modified Kling–Gupta efficiency skill score (KGESS). Optimum value of KGESS is 1. Blue (red) dots show catchments with positive (negative) hydrological skill.

at a 24 h time step (Table 1). The river discharge reforecast was initialised from GloFAS-ERA5 (Sect. 2.3) and forced by ECMWF-ENS reforecast runoff from the twice weekly, 11-member, 20-year ECMWF meteorological ensemble reforecasts.

2.3 GloFAS-ERA5 river discharge reanalysis

The GloFAS-ERA5 reanalysis dataset (Harrigan et al., 2020a) provides a spatio-temporally consistent estimate of daily historic river discharge. It is produced for every 0.1° river cell globally from 1979 to the present. It is updated operationally with a latency of 2 to 5 d behind real time, following the release of ERA5 atmospheric reanalysis (Hersbach et al., 2020). In GloFAS operational forecasts, GloFAS-ERA5 is used as initial conditions for the real-time forecasts (Figs. 1 and 2), and for calculating flood thresholds against which real time ensemble forecasts are compared to determine the probability of a flood signal (Zsoter et al., 2020a). For the forecast evaluation undertaken here, GloFAS-ERA5 is used as initial conditions for reforecasts, to generate benchmark forecasts and as proxy observations to evaluate forecast skill.

The hydrological performance of GloFAS-ERA5 will have implications for the forecasts and reforecasts here. If for example GloFAS-ERA5 has poor hydrological skill in resolving hydrological dynamics, particularly the timing of river discharge, this would contribute to poorer forecasts. An evaluation of GloFAS-ERA5 against a global

network of 1801 in-situ river discharge observation stations was undertaken by Harrigan et al. (2020a) and shown here in Fig. 4 for context. They found the reanalysis is skilful in 86 % of catchments according to the modified Kling–Gupta Efficiency Skill Score (KGESS) against a mean flow benchmark (see Fig. 4). The global median Pearson correlation coefficient is 0.61 with an interquartile range of 0.44 to 0.74. However, skill varies considerably with location with several regions such as central US, Africa, eastern Brazil, and western coast of South America having large systematic positive biases. For the evaluation presented here, GloFAS-ERA5 v2.1 data from 1979 to 2019 are used as downloaded from the Copernicus Climate Data Store (CDS): <https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-glofas-historical?tab=overview> (last access: 25 March 2022) (Harrigan et al., 2019).

3 Global forecast skill evaluation method

A first systematic evaluation of GloFAS hydrological forecast skill was carried out using the operational version 2.1/2.2 at the global scale, across lead times from 1 to 30 d, based on the comprehensive set of 20-year reforecasts described in Sect. 2.2. The forecast evaluation methodology is set out below with the aim of being applied routinely to all future major releases of GloFAS, with the forecast skill statistics provided as a new forecast skill layer on the GloFAS Web Map Viewer

as well as metadata information associated with the raw data provided on the Copernicus Climate Data Store (CDS). This aims to help users make better informed decisions on how, when, and where GloFAS forecasts might be appropriate for their needs.

3.1 Data sample

There are 5.4 M GloFAS 0.1° river network cells covering the global land area, so to avoid excessive redundancy, forecast skill is calculated for a subset called the GloFAS diagnostic river points. There are 5997 of these diagnostic points in total used across the GloFAS project by both model developers and users for a range of purposes, such as displaying forecast hydrographs and associated detailed metadata at each point on the Web Map Viewer (known as GloFAS web points), diagnosing reanalysis and forecast errors, and tracking improvement between model upgrades. These river points drain catchment areas ranging from 1068 to 5 359 150 km² with a median area of 29 051 km² (Fig. 4) and more information on each point can be found in Table S1 in the Supplement.

3.2 Benchmark forecasts

Forecast skill refers to the relative accuracy of a set of forecasts with respect to a set of standard reference or *benchmark forecasts* (Wilks, 2011). When designing a forecast evaluation experiment, a critical consideration is the selection of a benchmark forecast that has sufficient skill discrimination, i.e. is not too simple and represents as closely as possible the observations (Pappenberger et al., 2015).

Following Pappenberger et al. (2015) and because GloFAS produces seamless forecasts across short-, medium- and extended-range lead times (day 1 to 30), two benchmarks are considered here, each calculated for all GloFAS diagnostic river points: *persistence*, typically used for short-range lead times where the forecast signal is dominated by serial correlation of river discharge, and *climatology*, typically used for longer lead times where the forecast signal is dominated by the seasonality of river discharge defined as follows.

- Persistence benchmark forecast is defined as the single GloFAS-ERA5 daily river discharge of the day preceding the reforecast start date. The same river discharge value is used for all lead times. For example, for a forecast issued on 3 January at 00:00 UTC, the persistence benchmark forecast is the average river discharge over the 24 h time step from 2 January 00:00 UTC to 3 January 00:00 UTC, and the same value is used as benchmark for all 30 lead times (i.e., 4 January to 2 February).
- Climatology benchmark forecast is based on a 40-year climatological sample (1979–2018) of moving 31 d windows of GloFAS-ERA5 river discharge reanalysis values, centred on the date being evaluated (± 15 d). From each 1240-valued climatological sample

(i.e. 40 years \times 31 d window), 11 fixed quantiles (Q_n) at 10 % intervals were extracted ($Q_0, Q_{10}, Q_{20}, \dots, Q_{80}, Q_{90}, Q_{100}$). The fixed quantile climate distribution used therefore varies by lead time, capturing the temporal variability in local river discharge climatology.

3.3 Skill score

The ensemble forecast performance is evaluated using the continuous ranked probability score (CRPS) (Hersbach, 2000), one of the most widely used headline scores for operational probabilistic forecasts (Pappenberger et al., 2015; Alfieri et al., 2014). The CRPS compares the continuous cumulative distribution of an ensemble forecast with the distribution of the observations. It has an optimum value of 0 and measures the error in the same units as the variable of interest (here river discharge in m³ s⁻¹). It collapses to the mean absolute error (MAE) for deterministic forecasts, which is important here as the persistence benchmark forecast we use is deterministic. The CRPS is expressed as a skill score, CRPSS, to calculate forecast skill which measures the improvement in GloFAS over the benchmark forecast and is given in Eq. (1):

$$\text{CRPSS} = 1 - \frac{\text{CRPS}_{\text{fc}}}{\text{CRPS}_{\text{bench}}} \quad (1)$$

where CRPS_{fc} is the CRPS of the forecast against observations and $\text{CRPS}_{\text{bench}}$ is the CRPS of the benchmark forecast against observations. A CRPSS value of 1 indicates a perfect forecast, $\text{CRPSS} > 0$ shows forecasts are more skillful than the benchmark, $\text{CRPSS} = 0$ shows forecasts are only as accurate as the benchmark, and $\text{CRPSS} < 0$ means that forecasts are less skillful than the benchmark forecast.

The CRPSS was calculated using GloFAS reforecasts over 1999 to 2018 generated in Sect. 2.2 using both persistence and climatology benchmark forecasts (Sect. 3.2) and verified against GloFAS-ERA5 river discharge reanalysis used as proxy observations (following Alfieri et al., 2014) at each of the 5997 GloFAS diagnostic river points. Calculating forecast skill against proxy observations such as reanalysis is common in hydrological forecasting as it has the advantage of providing a spatio-temporally complete picture of forecast skill, currently not possible based on availability of the current global in situ observed river network (Lavers et al., 2019). It also allows the forecast predictability range to be isolated in the absence of systematic hydrological model errors. There is a disadvantage of forecast evaluation against proxy observations for catchments that represent hydrological dynamics poorly. While Harrigan et al. (2020a) demonstrate the performance of GloFAS-ERA5 reanalysis is largely hydrologically skillful, readers should be aware that there are areas where performance is poor and that there are large parts of the world where the performance is unknown due to the lack of in situ observations to evaluate against (Fig. 4).

4 Results and discussion

4.1 GloFAS forecast skill

4.1.1 GloFAS skill by lead time

Overall, GloFAS version 2.1/2.2 is skilful at the global scale for the majority of catchments across all lead times analysed (Fig. 5a). The CRPSS against persistence decays exponentially as a function of lead time out to around 15 d lead time, then stabilises reaching a minimum of 0.48 by day 22. The CRPSS at day 1 is 0.96 (interquartile range of 0.88 to 0.99), day 3 = 0.82 (0.62, 0.94), day 5 = 0.70 (0.50, 0.87), day 10 = 0.55 (0.37, 0.72). The CRPSS against climatology begins higher than persistence and decays continuously towards day 30. At day 15 the CRPSS is 0.47 (0.27, 0.68), day 20 = 0.37 (0.17, 0.58), day 25 = 0.30 (0.11, 0.49), day 30 = 0.25 (0.06, 0.43).

As the global median CRPSS against climatology becomes lower than against persistence (0.49 versus 0.50, respectively) from day 14, we present and discuss all forecast skill from short- (1 to 3 d) to medium-range (5 to 10 d) lead times calculated against the persistence benchmark forecast, and from extended lead times (15 to 30 d) calculated against the climatology benchmark forecast. To aid the readers interpretation of the CRPSS, individual CRPS components used in Eq. (1) (i.e. $CRPS_{fc_GloFAS}$, $CRPS_{bench_persistence}$, and $CRPS_{bench_climatology}$) are also shown in Fig. 5b expressed as a global median across all lead times.

4.1.2 Spatial distribution of GloFAS skill

At short-range lead times (1 and 3 d), GloFAS is skilful compared to the persistence benchmark forecast in over 96 % of catchments (Fig. 6). In the medium range at day 5, GloFAS remains skilful for 93 % of catchments. Regions with the highest skill ($CRPSS \geq 0.8$) include South America, especially the Amazon basin, the US, southern Africa, central Asia, and eastern Australia. There are notable clusters of catchments with negative skill (i.e. $CRPSS < 0$) mainly located in northern polar latitudes above $60^\circ N$ as well as in the Congo River Basin. The global median CRPSS at day 5 for catchments located in the northern polar climate region is 0.63 compared to 0.70 and 0.73 for extratropics and tropics, respectively (Fig. 8a). By day 10, the strength of skill has decreased, but 89 % of catchments remain skilful (i.e. $CRPSS > 0$).

For extended-range lead times shown in Fig. 7, GloFAS is skilful compared to the climatology benchmark forecast in 89 % of catchments by day 15, reducing to 86 %, 83 % and 80 % by lead times 20, 25 and 30 d, respectively. The regions of highest skill are similar to those for the short and medium range, with areas of negative skill expanding to tropical Africa, a large region in central and northern Asia, and western coast of South America. The global median CRPSS

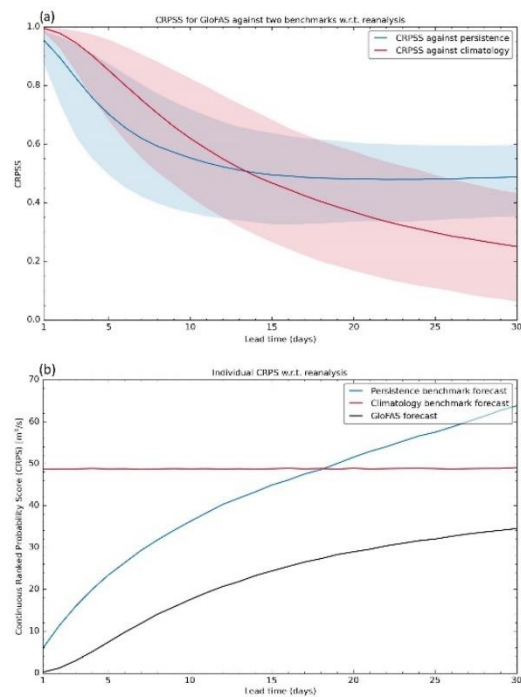


Figure 5. Skill of GloFAS 2.1/2.2 with global median continuous ranked probability skill score (CRPSS) for reforecasts against persistence (red line) and climatology (blue line) benchmarks from 1 to 30 d lead times with respect to GloFAS-ERA5 river discharge reanalysis at 5997 diagnostic river points (a). The interquartile range of CRPSS values at each lead time are shown by semi-transparent bands. Corresponding individual CRPS components used in Eq. (1) for GloFAS forecasts (black line), persistence benchmark (blue line) and climatology (red line) shown as a median across the diagnostic river points (b).

at day 20 for catchments located in the broader tropics (latitudes $23^\circ S$ to $23^\circ N$) is 0.31 compared to 0.40 and 0.37 for the extratropics and polar climate regions, respectively (Fig. 8a).

The choice of benchmark forecast used for the short- to medium-range (i.e., Fig. 6) and extended-range (i.e., Fig. 7) maps was based on the global median of all stations in Fig. 5a. However, there is spatial variability in the choice of best benchmark according to unique hydroclimate properties. For example, in northern latitudes around Russia and northern Scandinavia GloFAS was shown to be negatively skilful against persistence at a 10 d lead time (Fig. 6d). However, GloFAS is shown to be skilful against a climatology benchmark in the same region at lead time 15 d. This shows that persistence is a much tougher benchmark to beat in these

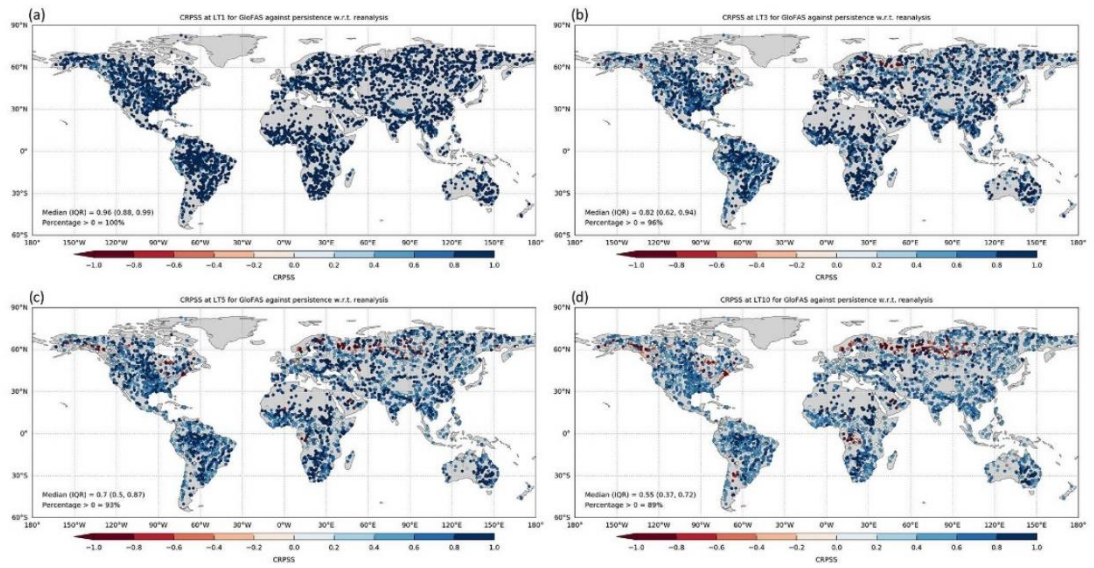


Figure 6. GloFAS 2.1/2.2 with continuous ranked probability skill score (CRPSS) for reforecasts against the persistence benchmark for short- to medium-range lead times – 1 d (a), 3 d (b), 5 d (c), and 10 d (d) – with respect to GloFAS-ERA5 river discharge reanalysis at 5997 diagnostic river points. Optimum value of CRPSS is 1. Blue (red) dots show catchments with positive (negative) skill.

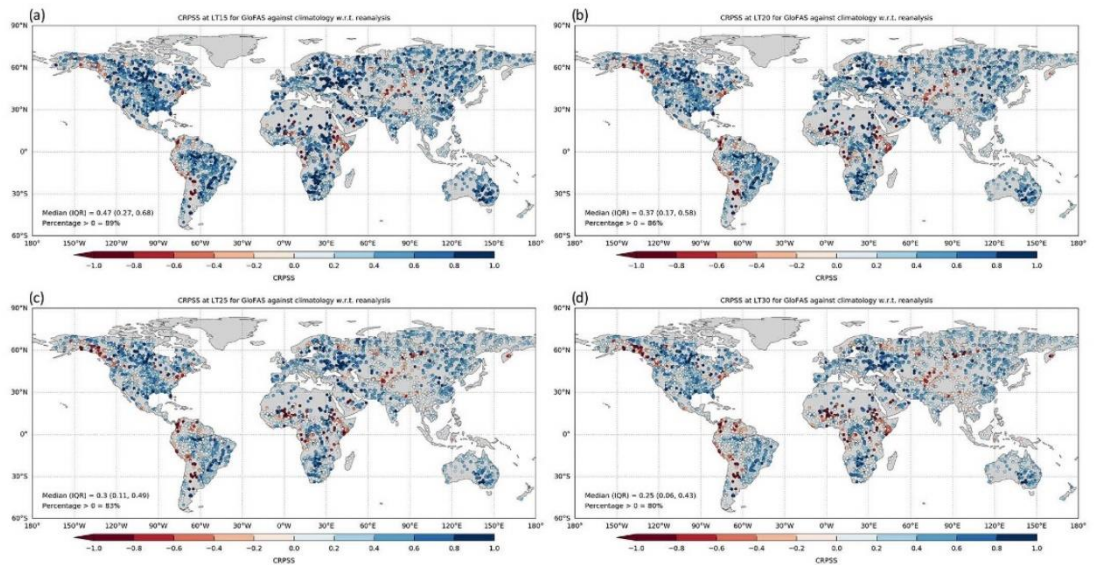


Figure 7. GloFAS 2.1/2.2 continuous ranked probability skill score (CRPSS) for reforecasts against the climatology benchmark for extended lead times – 15 d (a), 20 d (b), 25 d (c), and 30 d (d) – with respect to GloFAS-ERA5 river discharge reanalysis at 5997 diagnostic river points. Optimum value of CRPSS is 1. Blue (red) dots show catchments with positive (negative) skill.

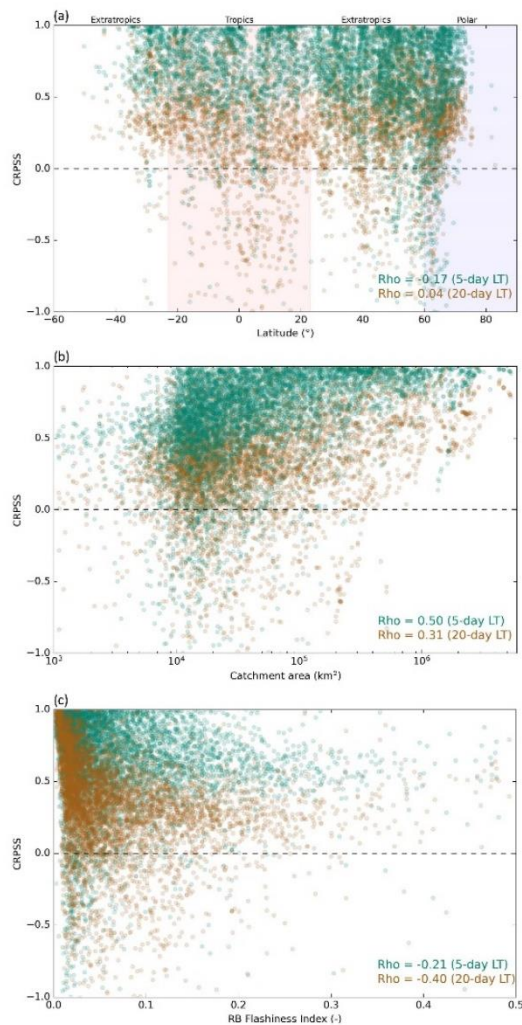


Figure 8. GloFAS 2.1/2.2 continuous ranked probability skill Score (CRPSS) for 5 d (green dots; against persistence benchmark) and 20 d (brown dots; against climatology benchmark) lead times at 5997 diagnostic river points by degree latitude of the river point with polar (tropics) climate region shaded in blue (red) (a), catchment area (b), and RB flashiness index (c). Spearman rank correlation coefficients (ρ) for each combination given in text in the bottom right.

catchments compared to climatology, likely due to the high degree of serial correlation from snow processes.

4.1.3 GloFAS skill by catchment area and hydrological flashiness

GloFAS skill using CRPSS for representative medium-range (using 5 d) and extended-range (using 20 d) lead times is correlated against catchment area in Fig. 8b and the Richards–Baker flashiness index (RB index; Baker et al., 2004) in Fig. 8c using the Spearman rank correlation coefficient (ρ). Forecast skill is moderately positively correlated with catchment area ($\rho = 0.50$ (0.31) for 5 (20)d lead times); catchments with larger areas have higher skill. This is consistent with findings in Ireland (Donegan et al., 2021; Quinn et al., 2021), and at the European scale from EFAS (Alfieri et al., 2014). While catchments with no skill (CRPSS ≤ 0) tend to be smaller, the majority of catchments with areas ranging between 1000 to 10 000 km² are skillful.

The RB index is calculated by dividing the pathlength of day-to-day river discharge changes by total river discharge for a given time interval. For each catchment, the RB index was extracted from GloFAS-ERAS over the time interval 1979 to 2019. The index provides a useful summary of hydrological functioning of a catchment. Catchments with a high RB index tend to have flashy hydrological response and are characterised as smaller upland catchments with increased frequency and magnitude of storm events, whereas catchments with a low RB index tend to be slower responding larger catchments with higher baseflow components (Baker et al., 2004). Forecast skill is weakly to moderately negatively correlated with RB index ($\rho = -0.21$ (–0.40) for 5 (20)d lead times); catchments with higher hydrological flashiness have lower skill. The link between higher catchment responsiveness and lower forecast skill has also been found in Ireland (Donegan et al., 2021; Quinn et al., 2021), the UK (Harrigan et al., 2018), Sweden (Girons Lopez et al., 2021), and at the European scale from EFAS (Pappenberger et al., 2015).

Formal attribution of the drivers of high and low hydrological forecast skill is outside the scope of this study but results point to several areas to prioritise research and development into model improvements. First, improving GloFAS forecast performance in smaller catchments with more flashy hydrological response should be a priority. This finding is expected given the relatively coarse horizontal (~ 11 km) and time (daily) resolution of a global-scale system such as GloFAS. A “hyperresolution” target in the order 1 km globally is required for hydrological prediction to be useful at local scales (Wood et al., 2011), but will bring computational, data, and hydrological science challenges (Harrigan et al., 2020b). Second, hydrological forecast skill is inherently dependent on global NWP model skill. Prediction of convective rainfall, dominant in the tropics, remains a challenge in the current generation of NWP, including the ECMWF IFS (~ 18 km

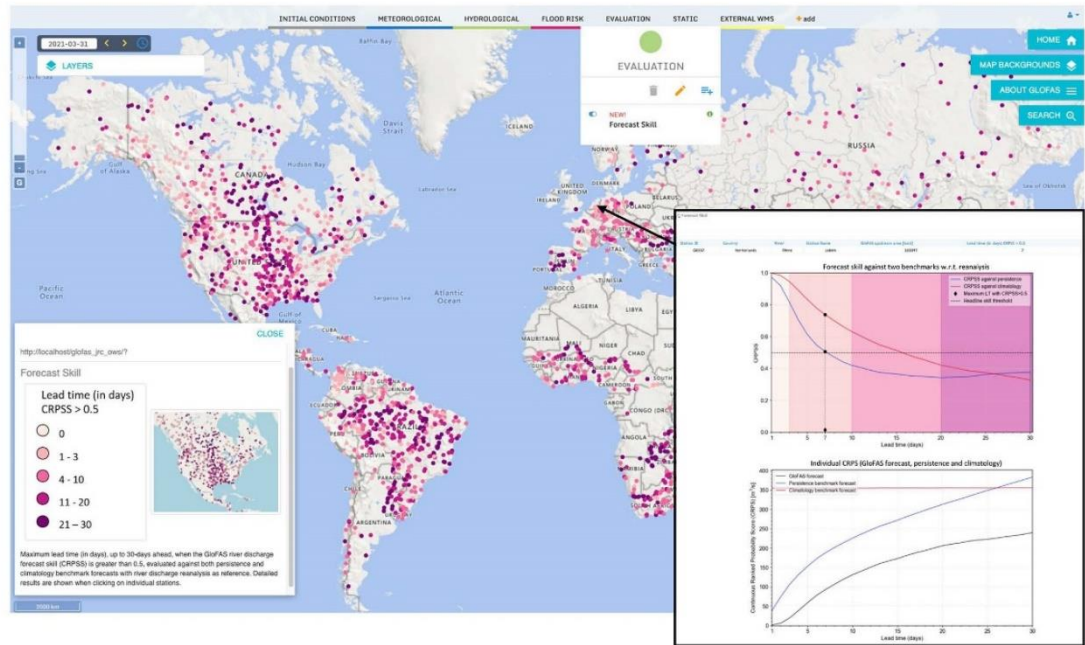


Figure 9. GloFAS 30 d forecast skill layer for the headline score available on the GloFAS Web Map Viewer. The headline score is the maximum lead time (in days) the continuous ranked probability skill scores (CRPSS) is greater than a value of 0.5, evaluated against a persistence or climatology benchmark forecast. Clicking on each GloFAS reporting point, a “pop-out” window shows the detailed CRPSS and CRPS across all lead times. An example for the Rhine at Lobith (Netherlands, G0337) is shown in the inset.

horizontal resolution for ENS) used to force GloFAS forecasts (Haiden et al., 2021; Lavers et al., 2021). Progress is, however, already underway. Recent increases in supercomputer power has allowed ground-breaking kilometre-scale NWP to be tested with promising results showing that deep convection can be explicitly simulated rather than parameterised as it is currently, thus providing better representation of convective storm activity (Wedi et al., 2020). Assessing the hydrological impact of any new precipitation improvement needs to be prioritised. Thirdly, hydrological prediction in regions with more challenging hydroclimate conditions needs further investigation, particularly snowy and icy catchments in polar regions where simplified snow accumulation and melt processes as well as rain-on-snow events are known to be highly sensitive to error (Fehlmann et al., 2019). From this first order assessment GloFAS forecast performance can drop considerably for many catchments in these regions. Therefore, more work is needed to investigate how existing and new representations of snow processes can deliver more skillful river discharge forecasts.

4.2 New GloFAS headline forecast skill layer on Web Map Viewer

To help the interpretation and understanding of the quality of GloFAS 30 d forecasts, the forecast skill scores produced in this paper are presented as new layer on the GloFAS Web Map Viewer since the release of GloFAS version 2.2 on 9 December 2020. Figure 9 shows a screenshot of the “Forecast skill” layer on the website. The new headline forecast skill score is defined as the maximum lead time (in days), up to 30 d ahead, in which the CRPSS is greater than a value of 0.5, when compared to a persistence or climatology benchmark forecast using GloFAS-ERA5 as proxy observations. A threshold of CRPSS = 0.5 is chosen for the summary layer to distinguish the lead time in which a station is “highly skillful” and is interpreted practically as the threshold at which the GloFAS forecast are 50 % more accurate than the respective benchmark forecast. The headline score is shown for the GloFAS reporting points. An example of the detailed skill information available for individual stations is shown for the Rhine at Lobith (Netherlands; G0337) in the inset of Fig. 9. The headline score for this station is at day 7, when the CRPSS against persistence drops below the 0.5 threshold.

<https://doi.org/10.5194/hess-27-1-2023>

Hydrol. Earth Syst. Sci., 27, 1–19, 2023

When the station is clicked on the web interface, a “pop-out” window appears and includes two plots, the CRPSS across the 30 d lead time and corresponding individual CRPS components. This will provide vital information for forecasters when conducting forecasting assessment during emergency situations.

4.3 Operational delivery of GloFAS data and metadata

The GloFAS global river discharge forecasts (real-time and reforecasts) and associated skill assessment analysis are provided free and openly by the European Commission Copernicus Emergency Management Service (CEMS). It follows the Copernicus open data policy that users shall have free, full, and open access to Copernicus Service Information. Users should, however, adhere to its terms and conditions available at <https://www.globalfloods.eu/terms-of-service/> (last access: 3 December 2022).

The Copernicus Climate Change Service (C3S) Climate Data Store (CDS; <https://cds.climate.copernicus.eu/cdsapp#!/home>, last access: 3 December 2022) hosts numerous global and regional reanalysis and forecast products, generally in the form of gridded records for essential climate variables (ECVs), including river discharge data as a key terrestrial ECV. The CDS requires standardisation of data and metadata so that datasets are more useable and discoverable through the CDS metadata pages. Its website provides easy access to data through user-friendly download forms, as well as a CDS Python Application Programming Interface (API) to allow programmatic access to data. An innovative feature of the CDS is its “Toolbox”, which makes it easier to handle large volumes of data by allowing users to make custom applications, filter data by geographical region and date range, and finally present the data using maps and charts directly through the CDS cloud infrastructure.

The GloFAS real-time river discharge forecasts from 5 November 2019 until present are available on the CDS and updated operationally every day: <https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-glofas-forecast?tab=overview> (last access: 25 March 2022) (Zsoter et al., 2019). The GloFAS river discharge reforecasts for the period 1999 to 2018 are also available on the CDS and update ahead of each major model cycle release since version 2.2: <https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-glofas-reforecast?tab=overview> (last accessed: 25 March 2022) (Zsoter et al., 2020b). The CDS landing page for the GloFAS forecast dataset is shown in Fig. 10. The forecast data are available in two ways. The first is through the “Data Download” tab whereby users can manually select options in a form for which data they would like to download in either GRIB or NetCDF file format. Second, data can be retrieved through the dedicated Python CDS API; an example API retrieval script is shown in Fig. B1 for the forecast start date of 1 January 2022 for both the single control (CTL) forecast and 50 ensemble perturbed members

out to a lead time of 30 d at 24 h steps and downloaded in NetCDF format. Note that users must register for a CDS account (for free) before gaining access. This landing page always provide access to the latest operational system, with the possibility to go through earlier versions in the archive when searching through past dates. For users interested in the raw forecast skill scores calculated in this paper, they are provided for all GloFAS diagnostic river points through the “Documentation” tab on the CDS as well as in Table S1. See Fig. 11 for an extract of the skill score information provided.

While producing large sets of reforecasts and providing data free and open to the community has many benefits, it comes with challenges and key considerations. One of the main considerations is the data storage and delivery infrastructure. A full set of 20-year GloFAS reforecasts is ~ 23 TB in size. For each new major model upgrade a new set of reforecasts are generated, together with ~ 35 GB of raw data generated every day for the real-time forecast stream. It is clear that the size of data is a barrier for many users to use. Most users do not require the full temporal range of data and are usually interested in a sub-domain, for example their study region or country. It is simply not practical for every user to download ~ 23 TB of global data to their computing infrastructure if they only want data for their individual catchment, not to mention if a standard laptop is the only computer available to them. Our solution was to store GloFAS data on the ECMWF Meteorological Archival and Retrieval System (MARS; <https://confluence.ecmwf.int/display/UDOC/MARS+user+documentation>, last access: 25 March 2022) – MARS offers the functionality for users to choose temporal and/or spatial subsets (among others) and the heavy data handling and computation happens on ECMWF infrastructure so the user can download a smaller and more manageable subset of data. The CDS is the public facing front end for users to access GloFAS data and metadata, and communicates with MARS in the backend. A further consideration is producing sufficient documentation for users to interact with the data and provision of a support service whereby users can get in contact with GloFAS data and domain experts for queries: <https://confluence.ecmwf.int/site/support> (last access: 25 March 2022).

4.4 Future directions

While this paper sets out the components, operational configuration, and a global forecast evaluation of GloFAS 2.1/2.2, the raw real-time forecast and reforecast data have been made openly available to encourage users to use the data for downstream value-added applications and to perform user-specific evaluation of forecast quality. Additionally, GloFAS forecasts and reforecasts have not been post-processed, therefore there is room for users to increase further forecast quality by applying post-processing with their local observations data to correct forecast bias or timing errors, for example. The evaluation carried out here looks at the overall quality of

River discharge and related forecasted data by the Global Flood Awareness System

Overview | Download data | Documentation

This dataset contains global modelled daily data of **river discharge** forced with meteorological forecasts. The data was produced by the Global Flood Awareness System (GloFAS), which is part of the Copernicus Emergency Management Service (CEMS). River discharge, or river flow as it is also known, is defined as the amount of water that flows through a river section at a given time.

This dataset is simulated by forcing the hydrological river routing model with modelled gridded runoff data from ECMWF ensemble forecast combined with the ECMWF extended-range ensemble forecast up to 30 days. Data availability for the GloFAS forecast is from 2019-11-05 up to near real time.

The runoff is produced by the land surface model HTESEL, and the river routing model component is LISFLOOD, run with a 0.1 x 0.1 degree lat-lon resolution.

Continuous ranked probability skill score for GloFAS forecast at lead time 5 days against reanalysis

DATA DESCRIPTION	
Data type	Gridded
Projection	Regular latitude-longitude grid
Horizontal coverage	Global except for Antarctica (90N-60S, 180W-180E)
Horizontal resolution	0.1° x 0.1°
Temporal coverage	5 November 2019 to near real time
Temporal resolution	Daily data
File format	GRIB2 and NetCDF-4
Conventions	WMO standards for GRIB2. The NetCDF-4 files inherit the WMO GRIB2 conventions
Versions	Current version - GloFAS v2.1 released 2020-05-19. For more information on versions we refer to the documentation
Update frequency	Updated daily. For more information on the model versions, please refer to the documentation

MAIN VARIABLES		
Name	Units	Description
River discharge in the last 24 hours	m ³ s ⁻¹	Volume rate of water flow, including sediments, chemical and biological material, in the river channel averaged over a time step through a cross-section. The value is an average over a 24-hour period.

RELATED VARIABLES		
Name	Units	Description
Upstream area	m ²	Static file - upArea.nc, upstream area for the point in the river network

Record updated 2021-04-05 10:50:07 UTC

Contact
copernicus-support@ecmwf.int

Licence
CEMS-FLOODS datasets licence

Publication date
2020-05-19

References
Citation
DOI: 10.24381/cds.ff1ae777

Related data
Reforecasts of river discharge and related data by the Global Flood Awareness System
River discharge and related historical data from the Global Flood Awareness System
Seasonal forecasts of river discharge and related data by the Global Flood Awareness System
Seasonal reforecasts of river discharge and related data from the Global Flood Awareness System

About C3S | Contact us | Cookies | Disclaimer | Privacy

Figure 10. The GloFAS river discharge forecast landing page on the C3S Climate Data Store (CDS: <https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-glofas-forecast?tab=overview>, last access: 3 December 2022).

forecasts only. Future work should assess other aspects of forecast quality such as reliability (Robertson et al., 2013), value (Clope et al., 2017) or performance during extreme events (Bischniotis et al., 2019). The robust and comprehensive reforecast strategy established for this first evaluation will serve as benchmark against which any new major GloFAS model upgrades can be compared. The GloFAS release strategy now includes public availability and easy access to the river discharge reanalysis, real-time forecasts and reforecasts together with a first assessment of global forecast skill and will continue for all future major GloFAS launches. We recommend a similar strategy for all global- and continental-scale hydrological forecasting systems as release of data has traditionally been limited to historical data used for specific inter-comparisons in hydrological performance (e.g. Beck et al., 2017; Towner et al., 2019) rather than a comprehensive set of reforecasts or real-time forecasts. This will pave the way for multi-model forecast skill comparisons, such


as those carried out routinely in the NWP field (for example, WMO Lead Centre for Deterministic NWP verification: <https://apps.ecmwf.int/wmolcdnv/>, last access: 13 October 2020).

5 Conclusion

It is now technically and computationally feasible to produce operational hydrological forecasting at the global scale. This offers enormous potential in aiding decision-making and humanitarian action in the face of large-scale and often trans-boundary flood events, as demonstrated by the application of GloFAS in recent floods such as those in Mozambique and Bangladesh. Nevertheless, up until now there have been limited information on hydrological forecast skill, both published in the scientific literature and available to users within the forecast web interface. This paper sets out the model components and operational configuration used in the pro-

<https://doi.org/10.5194/hess-27-1-2023>

Hydrol. Earth Syst. Sci., 27, 1–19, 2023

Copernicus Emergency Management Service 

Country	GloFAS ID	Station filter	LT1	LT2	LT3	LT4	LT5	LT6	LT7	LT8	LT9	LT10	LT12	LT14	LT16	LT18	LT20	LT25	L30
France	G1589	1	0.98	0.90	0.79	0.70	0.63	0.58	0.54	0.52	0.49	0.47	0.44	0.42	0.41	0.42	0.43	0.47	0.48
France	G1590	1	0.83	0.74	0.69	0.65	0.63	0.59	0.55	0.51	0.48	0.45	0.43	0.40	0.36	0.33	0.31	0.30	0.31
France	G1591	1	0.74	0.58	0.53	0.50	0.48	0.47	0.44	0.42	0.39	0.38	0.36	0.35	0.35	0.34	0.33	0.34	0.35
France	G1592	1	0.95	0.75	0.31	-0.11	-0.35	-0.45	-0.49	-0.49	-0.51	-0.50	-0.46	-0.42	-0.37	-0.19	-0.03	0.11	0.16
France	G1593	1	0.91	0.72	0.50	0.30	0.17	0.10	0.05	0.01	-0.03	-0.04	-0.04	-0.03	0.00	0.05	0.10	0.21	0.27
France	G1594	1	0.73	0.64	0.57	0.52	0.49	0.45	0.41	0.36	0.34	0.31	0.29	0.27	0.25	0.19	0.16	0.15	0.17
France	G1595	1	0.91	0.87	0.82	0.77	0.71	0.65	0.55	0.47	0.40	0.36	0.32	0.30	0.29	0.29	0.31	0.38	0.41
France	G1596	1	0.94	0.91	0.88	0.82	0.73	0.63	0.50	0.39	0.31	0.27	0.23	0.22	0.22	0.24	0.27	0.37	0.41
France	G1597	1	0.99	0.98	0.95	0.86	0.72	0.51	0.30	0.15	0.05	0.00	-0.06	-0.06	-0.05	-0.01	0.06	0.22	0.29
France	G1598	1	0.86	0.76	0.69	0.62	0.58	0.56	0.53	0.49	0.47	0.46	0.42	0.40	0.36	0.33	0.31	0.29	0.29
France	G1599	1	0.89	0.79	0.73	0.70	0.66	0.64	0.61	0.59	0.56	0.54	0.49	0.47	0.43	0.43	0.42	0.46	0.46
France	G1600	1	0.99	0.96	0.92	0.86	0.81	0.75	0.70	0.66	0.63	0.60	0.55	0.50	0.48	0.45	0.44	0.44	0.44
France	G1601	1	0.97	0.90	0.78	0.68	0.61	0.57	0.54	0.52	0.51	0.50	0.48	0.46	0.46	0.45	0.45	0.47	0.47
France	G3694	1	0.79	0.71	0.66	0.60	0.57	0.56	0.53	0.52	0.50	0.51	0.47	0.44	0.42	0.40	0.40	0.42	0.42
France	G3695	1	0.93	0.91	0.87	0.82	0.74	0.64	0.52	0.41	0.34	0.29	0.25	0.23	0.23	0.25	0.27	0.38	0.42
France	G3699	1	0.75	0.62	0.49	0.41	0.37	0.37	0.36	0.35	0.35	0.35	0.33	0.34	0.34	0.35	0.35	0.38	0.39
France	G3716	1	0.98	0.94	0.90	0.86	0.82	0.78	0.76	0.73	0.70	0.68	0.64	0.60	0.57	0.55	0.54	0.53	0.53
France	G3851	1	0.86	0.78	0.67	0.63	0.61	0.57	0.55	0.52	0.52	0.50	0.47	0.45	0.44	0.43	0.44	0.45	0.47
Germany	G0292	1	0.97	0.93	0.84	0.73	0.59	0.45	0.35	0.28	0.22	0.19	0.16	0.17	0.19	0.22	0.25	0.32	0.38
Germany	G0297	1	0.98	0.91	0.82	0.69	0.54	0.41	0.31	0.24	0.19	0.16	0.15	0.16	0.18	0.22	0.25	0.32	0.38
Germany	G0298	1	0.93	0.72	0.38	0.19	0.11	0.09	0.10	0.13	0.16	0.19	0.24	0.29	0.33	0.35	0.37	0.44	0.49
Germany	G0301	1	0.99	0.93	0.74	0.49	0.30	0.17	0.09	0.03	0.01	0.00	0.02	0.06	0.10	0.13	0.17	0.24	0.30
Germany	G0304	1	0.91	0.64	0.25	0.05	-0.02	-0.03	-0.01	0.02	0.05	0.08	0.15	0.22	0.27	0.31	0.35	0.43	0.49
Germany	G0306	1	0.98	0.97	0.95	0.92	0.86	0.79	0.72	0.65	0.59	0.53	0.45	0.39	0.35	0.33	0.32	0.32	0.34
Germany	G0309	1	0.97	0.95	0.91	0.86	0.81	0.77	0.73	0.69	0.66	0.64	0.59	0.54	0.50	0.48	0.47	0.46	0.46
Germany	G0310	1	0.97	0.95	0.91	0.86	0.81	0.77	0.73	0.69	0.66	0.64	0.59	0.54	0.50	0.48	0.47	0.46	0.46
Germany	G0311	1	0.87	0.54	0.15	-0.02	-0.07	-0.07	-0.05	-0.02	0.02	0.05	0.13	0.19	0.24	0.28	0.32	0.40	0.46
Germany	G0313	1	0.98	0.94	0.86	0.71	0.55	0.41	0.30	0.23	0.18	0.15	0.14	0.15	0.18	0.21	0.24	0.32	0.38
Germany	G0316	1	0.94	0.87	0.81	0.76	0.72	0.69	0.66	0.63	0.59	0.55	0.50	0.46	0.44	0.44	0.44	0.44	0.44
Germany	G0320	1	0.86	0.51	0.13	-0.03	-0.09	-0.08	-0.06	-0.03	0.01	0.04	0.12	0.19	0.24	0.28	0.31	0.40	0.46
Germany	G0321	1	0.84	0.47	0.07	-0.09	-0.14	-0.13	-0.11	-0.07	-0.03	0.00	0.08	0.16	0.21	0.25	0.28	0.38	0.44
Germany	G0323	1	0.83	0.40	0.07	-0.08	-0.17	-0.20	-0.19	-0.16	-0.15	-0.12	-0.06	-0.01	0.04	0.08	0.12	0.22	0.30
Germany	G0325	1	0.99	0.95	0.85	0.69	0.52	0.38	0.28	0.21	0.16	0.14	0.13	0.15	0.17	0.20	0.24	0.32	0.38
Germany	G0329	1	0.99	0.95	0.85	0.69	0.52	0.38	0.28	0.21	0.16	0.14	0.13	0.15	0.17	0.20	0.24	0.32	0.38
Germany	G0330	1	0.75	0.23	-0.10	-0.22	-0.23	-0.21	-0.17	-0.13	-0.08	-0.03	0.06	0.13	0.19	0.23	0.27	0.36	0.43
Germany	G0331	1	0.99	0.97	0.92	0.85	0.75	0.64	0.52	0.42	0.33	0.27	0.19	0.17	0.19	0.21	0.24	0.31	0.37
Germany	G0332	1	0.99	0.93	0.81	0.62	0.43	0.28	0.18	0.12	0.08	0.07	0.08	0.11	0.14	0.18	0.21	0.30	0.36
Germany	G0333	1	0.98	0.92	0.79	0.58	0.39	0.26	0.18	0.13	0.11	0.10	0.11	0.13	0.16	0.20	0.24	0.32	0.38

GloFAS v2.2 Metadata and Skill Scores | GloFAS Forecast Skill Scores against Persistence 432

Figure 11. Extract of GloFAS 2.1/2.2 river discharge forecast skill scores for CRPSS against persistence provided for a selection of lead times (LT) out to 30 d ahead as metadata information available through the Climate Data Store (CDS) documentation tab for each of the GloFAS diagnostic points as well as in Table S1.

duction of GloFAS real-time forecasts and in the generation of the corresponding large-sample set of 20-year reforecasts. A comprehensive global ensemble forecast evaluation strategy was developed that included a sensitivity assessment on both persistence and climatology benchmark forecasts given the 30 d range of GloFAS (re)forecasts. Global forecast skill results show that GloFAS is skilful in over 93 % of catchments in the short (1 to 3 d) and medium range (5 to 15 d) against a persistence benchmark forecast and skilful in over 80 % of catchments out to the extended range (16 to 30 d) against a climatology benchmark forecast. However, the strength of skill varies considerably by location with GloFAS found to have no or negative skill at longer lead times in broad hydroclimatic regions in tropical Africa, western coast of South America, and catchments dominated by snow and ice in high northern latitudes. These results highlight to users where and when GloFAS is skilful and is a crucial piece of information in the forecast decision-making process and has been made available to forecasters as a new layer in the GloFAS Web Map Viewer since the service-only upgrade to version 2.2 as of 9 December 2020. The results are also useful for model development so that areas where GloFAS performs

poorly can be further investigated and new model components designed and tested for improvements. An innovative feature of the GloFAS service development is providing the raw real-time forecast and reforecast data openly to encourage users to explore the data for downstream value-added applications and to perform user-specific and local evaluation of forecast quality.

Appendix A

Table A1. Scientific papers and model documentation for the key GloFAS model components. Adapted from Harrigan et al. (2020a).

GloFAS component	Description	Reference
ECMWF IFS	ECMWF Integrated Forecast System (IFS) model documentation for current operational model cycle 47r1	ECMWF (2021)
ERA5	Global reanalysis dataset using ECMWF IFS model cycle 41r2 from 1979 to present	Hersbach et al. (2020)
ECMWF IFS/ERA5 runoff	Surface and subsurface runoff within ECMWF IFS/ERA5 generated using the HTESSEL land surface model	Balsamo et al. (2009)
GloFAS-ERA5	GloFAS-ERA5 operational global river discharge reanalysis 1979–present	Harrigan et al. (2020a)
LISFLOOD river discharge	River discharge generated using LISFLOOD hydrological and channel routing model to route runoff into and through the river network and provide groundwater storage. LISFLOOD includes lake, reservoir and human water use routines	Burek et al. (2013)
Lakes and reservoirs used in GloFAS	Incorporated 463 lakes and 667 reservoirs into the GloFAS river network	Zajac et al. (2017)
Calibration of LISFLOOD used in GloFAS	LISFLOOD was calibrated against daily river discharge from 1287 observation stations worldwide	Hirpa et al. (2018)

Appendix B

```

# Example CDS Python API request script

# Code snippets can be found by clicking 'Show API request' at
# bottom of the download form:
# https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-glofas-forecast?tab=form

# Instructions on how to download CDS API can be found here:
# https://cds.climate.copernicus.eu/api-how-to

import cdsapi

c = cdsapi.Client()

c.retrieve(
  'cems-glofas-forecast',
  {
    'system_version': 'version_2_1',
    'hydrological_model': 'htesse_lisflood',
    'product_type': [
      'control_forecast', 'ensemble_perturbed_forecasts',
    ],
    'variable': 'river_discharge_in_the_last_24_hours',
    'year': '2022',
    'month': '01',
    'day': '01',
    'leadtime_hour': [
      '24', '48', '72',
      '96', '120', '144',
      '168', '192', '216',
      '240', '264', '288',
      '312', '336', '360',
      '384', '408', '432',
      '456', '480', '504',
      '528', '552', '576',
      '600', '624', '648',
      '672', '696', '720',
    ],
    'format': 'netcdf',
  },
  'download.nc')

```

Figure B1. Example Climate Data Store (CDS) Python API retrieval script for the GloFAS v2.1 daily river discharge forecast on 1 January 2022 for both the single control (CTL) forecast and 50 ensemble perturbed members out to a lead time of 30 d to be downloaded in NetCDF format.

Code availability. The underlying code is available upon request from the corresponding author.

Supplement. The supplement related to this article is available online at: <https://doi.org/10.5194/hess-27-1-2023-supplement>.

Data availability. GloFAS-ERA5 v2.1 river discharge reanalysis data can be downloaded from the Copernicus Climate Data Store (CDS): <https://doi.org/10.24381/cds.a4fdd6b9> (Harrigan et al., 2019). GloFAS v2.1 river discharge real-time forecast data can be downloaded from the Copernicus Climate Data Store (CDS): <https://doi.org/10.24381/cds.ff1aef77> (Zsoter et al., 2019). GloFAS v2.2 river discharge reforecast data can be downloaded from the Copernicus Climate Data Store (CDS): <https://doi.org/10.24381/cds.2d78664e> (Zsoter et al., 2020b).

Author contributions. SH and CP designed the study. SH drafted the manuscript and performed the forecast evaluation. EZ developed the GloFAS suites to produce the reforecasts and real-time forecasts. HC and PS helped frame the paper. All co-authors contributed to the editing of the manuscript and to the discussion and interpretation of results.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. We thank Francesca Moschini and Karen O'Regan (both ECMWF) for their help with production of the forecast skill layer and metadata for the GloFAS Web Map Viewer and CDS, respectively; Christopher Barnard (ECMWF) for his help in integrating the real-time forecast and reforecast suits into operations; Corentin Carton De Wiart (ECMWF) for his help in implementing the forecast evaluation suite; and Iacopo Ferrario, Christopher Barnard, Fredrik Wetterhall, Sebastien Villaume and the wider CDS/MARS team at ECMWF for their work ingesting the GloFAS data into the CDS and MARS. We thank the water@reading team from the University of Reading who collaborate in the development and use of GloFAS in which many conversations and engagement with users have shaped the service evolution. Finally, we thank Berit Arheimer and four anonymous referees for their constructive feedback that has greatly improved this paper.

Financial support. This research has been supported by the European Commission Copernicus Emergency Management Service (CEMS) (grant no. 198702).

Review statement. This paper was edited by Jim Freer and reviewed by Berit Arheimer and four anonymous referees.

References

- Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger, F.: 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>, 2013.
- Alfieri, L., Pappenberger, F., Wetterhall, F., Haiden, T., Richardson, D., and Salamon, P.: Evaluation of ensemble streamflow predictions in Europe, *J. Hydrol.*, 517, 913–922, <https://doi.org/10.1016/j.jhydrol.2014.06.035>, 2014.
- Arheimer, B., Pimentel, R., Isberg, K., Crochemore, L., Andersson, J. C. M., Hasan, A., and Pineda, L.: Global catchment modelling using World-Wide HYPE (WWH), open data, and stepwise parameter estimation, *Hydrol. Earth Syst. Sci.*, 24, 535–559, <https://doi.org/10.5194/hess-24-535-2020>, 2020.
- Baker, D. B., Richards, R. P., Loftus, T. T., and Kramer, J. W.: A New Flashiness Index: Characteristics and Applications to Midwestern Rivers and Streams, *J. Am. Water Resour. Assoc.*, 40, 503–522, <https://doi.org/10.1111/j.1752-1688.2004.tb01046.x>, 2004.
- Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., and Betts, A. K.: 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, 623–643, <https://doi.org/10.1175/2008JHM1068.1>, 2009.
- Beck, H. E., Dijk, A. I. J. M. van, Roo, A. de, Dutra, E., Fink, G., Orth, R., and Schellekens, J.: Global evaluation of runoff from 10 state-of-the-art hydrological models, *Hydrol. Earth Syst. Sci.*, 21, 2881–2903, <https://doi.org/10.5194/hess-21-2881-2017>, 2017.
- Bischiotti, K., van den Hurk, B., Zsoter, E., de Perez, E. C., Grilakis, M., and Aerts, J. C. J. H.: Evaluation of a global ensemble flood prediction system in Peru, *Hydrolog. Sci. J.*, 64, 1171–1189, <https://doi.org/10.1080/02626667.2019.1617868>, 2019.
- Blöschl, G., Hall, J., Parajka, J., Perdígão, R. A. P., Merz, B., Arheimer, B., Aronica, G. T., Bilbashi, A., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Fiala, K., Frolova, N., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T. R., Kohnová, S., Koskela, J. J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Rogger, M., Salinas, J. L., Sauquet, E., Šraj, M., Szolgay, J., Viglione, A., Volpi, E., Wilson, D., Zaimi, K., and Živković, N.: Changing climate shifts timing of European floods, *Science*, 357, 588–590, <https://doi.org/10.1126/science.aan2506>, 2017.
- Blöschl, G., Hall, J., Viglione, A., Perdígão, R. A. P., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G. T., Bilbashi, A., Boháč, M., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Frolova, N., Ganora, D., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T. R., Kohnová, S., Koskela, J. J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Salinas, J. L., Sauquet, E., Šraj, M., Szolgay, J., Volpi, E., Wilson, D., Zaimi, K., and Živković, N.: Changing climate both increases and decreases European river floods, *Nature*, 573, 108–111, <https://doi.org/10.1038/s41586-019-1495-6>, 2019.
- Burek, P., van der Knijff, J. M., and de Roo, A. P. J. D.: LISFLOOD – Distributed Water Balance and Flood Simulation Model – Revised User Manual, Publications Office of the European Union, <https://doi.org/10.2788/24719>, 2013.
- Cloke, H. L., Pappenberger, F., Smith, P. J., and Wetterhall, F.: How do I know if I've improved my continental scale flood early warning system?, *Environ. Res. Lett.*, 12, 044006, <https://doi.org/10.1088/1748-9326/aa625a>, 2017.
- Demargne, J., Wu, L., Regonda, S. K., Brown, J. D., Lee, H., He, M., Seo, D.-J., Hartman, R., Herr, H. D., Fresch, M., Schaake, J., and Zhu, Y.: The Science of NOAA's Operational Hydrologic Ensemble Forecast Service, *B. Am. Meteorol. Soc.*, 95, 79–98, <https://doi.org/10.1175/BAMS-D-12-00081.1>, 2014.
- Donegan, S., Murphy, C., Harrigan, S., Broderick, C., Foran Quinn, D., Golian, S., Knight, J., Matthews, T., Prudhomme, C., Scaife, A. A., Stringer, N., and Wilby, R. L.: Conditioning ensemble streamflow prediction with the North Atlantic Oscillation improves skill at longer lead times, *Hydrol. Earth Syst. Sci.*, 25, 4159–4183, <https://doi.org/10.5194/hess-25-4159-2021>, 2021.

<https://doi.org/10.5194/hess-27-1-2023>

Hydrol. Earth Syst. Sci., 27, 1–19, 2023

- Donnelly, C., Andersson, J. C. M., and Arheimer, B.: Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe, *Hydrolog. Sci. J.*, 61, 255–273, <https://doi.org/10.1080/02626667.2015.1027710>, 2016.
- ECMWF: IFS Documentation CY47R3, <https://doi.org/10.21957/eyrpir4vj>, 2021.
- Emerton, R., Zsoter, E., Arnal, L., Cloke, H. L., Muraro, D., Prudhomme, C., Stephens, E. M., Salamon, P., and Pappenberger, F.: Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0, *Geosci. Model Dev.*, 11, 3327–3346, <https://doi.org/10.5194/gmd-11-3327-2018>, 2018.
- 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., Tittley, H., Magnusson, L., Pappenberger, F., Klingaman, N., and Stephens, E.: 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. Disast. Risk Reduct.*, 50, 101811, <https://doi.org/10.1016/j.ijdr.2020.101811>, 2020.
- 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.: Continental and global scale flood forecasting systems, *Wiley Interdiscip. Rev. Water*, 3, 391–418, <https://doi.org/10.1002/wat2.1137>, 2016.
- Fehlmann, M., Gascón, E., Rohrer, M., Schwarb, M., and Stoffel, M.: Improving Medium-Range Forecasts of Rain-on-Snow Events in Prealpine Areas, *Water Resour. Res.*, 55, 7638–7661, <https://doi.org/10.1029/2018WR024644>, 2019.
- Girons Lopez, M., Crochemore, L., and Pechlivanidis, I. G.: Benchmarking an operational hydrological model for providing seasonal forecasts in Sweden, *Hydrol. Earth Syst. Sci.*, 25, 1189–1209, <https://doi.org/10.5194/hess-25-1189-2021>, 2021.
- Haiden, T., Janousek, M., Vitart, F., Ben-Bouallegue, Z., Ferranti, L., Prates, C., and Richardson, D.: Evaluation of ECMWF forecasts, including the 2020 upgrade, ECMWF, <https://doi.org/10.21957/6njp8byz4>, 2021.
- Harrigan, S., Prudhomme, C., Parry, S., Smith, K., and Tanguy, M.: Benchmarking ensemble streamflow prediction skill in the UK, *Hydrol. Earth Syst. Sci.*, 22, 2023–2039, <https://doi.org/10.5194/hess-22-2023-2018>, 2018.
- Harrigan, S., Zsoter, E., Barnard, C., Wetterhall, F., Salamon, P., and Prudhomme, C.: River discharge and related historical data from the Global Flood Awareness System, v2.1, Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [data set], <https://doi.org/10.24381/cds.a4fdd6b9>, 2019.
- Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., Barnard, C., Cloke, H., and Pappenberger, F.: 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>, 2020a.
- Harrigan, S., Cloke, H., and Pappenberger, F.: Innovating global hydrological prediction through an Earth system approach, *WMO Bull.*, <https://public.wmo.int/en/resources/bulletin/innovating-global-hydrological-prediction-through-earth-system> (last access: 3 December 2022), 2020b.
- Hersbach, H.: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems, *Weather Forecast.*, 15, 559–570, [https://doi.org/10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2), 2000.
- 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., Chiara, G. D., 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.: The ERA5 Global Reanalysis, *Q. J. Roy. Meteorol. Soc.*, 146, 1999–2049, <https://doi.org/10.1002/qj.3803>, 2020.
- Hirpa, F. A., Salamon, P., Beck, H. E., Lorini, V., Alfieri, L., Zsoter, E., and Dadson, S. J.: 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>, 2018.
- Hossain, S., Prudhomme, C., Cloke, H., and Stephens, L.: GloFAS helps Bangladesh flood forecasters to predict monsoon flood, *ECMWF Newsl.*, 165, 6–7, 2020.
- Jolliffe, I. T. and Stephenson, D. B.: *Forecast verification: a practitioner's guide in atmospheric science*, John Wiley & Sons, ISBN 978-0-470-66071-3, 2012.
- Lavers, D., Harrigan, S., Andersson, E., Richardson, D. S., Prudhomme, C., and Pappenberger, F.: A vision for improving global flood forecasting, *Environ. Res. Lett.*, 14, 121002, <https://doi.org/10.1088/1748-9326/ab52b2>, 2019.
- Lavers, D. A., Harrigan, S., and Prudhomme, C.: Precipitation Biases in the ECMWF Integrated Forecasting System, *J. Hydrometeorol.*, 22, 1187–1198, <https://doi.org/10.1175/JHM-D-20-0308.1>, 2021.
- Magnusson, L., Zsoter, E., Prudhomme, C., Baugh, C., Harrigan, S., Ficchi, A., Emerton, R., Cloke, H., Stephens, L. and Speight, L.: ECMWF works with universities to support response to tropical cyclone Idai, *ECMWF Newsl.*, 160, 2–3, 2019.
- Pappenberger, F., Ramos, M. H., Cloke, H. L., Wetterhall, F., Alfieri, L., Bogner, K., Mueller, A., and Salamon, P.: How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction, *J. Hydrol.*, 522, 697–713, <https://doi.org/10.1016/j.jhydrol.2015.01.024>, 2015.
- Quinn, D. F., Murphy, C., Wilby, R. L., Matthews, T., Broderick, C., Golian, S., Donegan, S., and Harrigan, S.: Benchmarking seasonal forecasting skill using river flow persistence in Irish catchments, *Hydrolog. Sci. J.*, 66, 672–688, <https://doi.org/10.1080/02626667.2021.1874612>, 2021.
- Robertson, D. E., Pokhrel, P., and Wang, Q. J.: Improving statistical forecasts of seasonal streamflows using hydrological model output, *Hydrol. Earth Syst. Sci.*, 17, 579–593, <https://doi.org/10.5194/hess-17-579-2013>, 2013.
- Thielen, J., Bartholmes, J., Ramos, M.-H., and de Roo, A.: The European Flood Alert System – Part 1: Concept and development, *Hydrol. Earth Syst. Sci.*, 13, 125–140, <https://doi.org/10.5194/hess-13-125-2009>, 2009.
- Towner, J., Cloke, H. L., Zsoter, E., Flamiig, Z., Hoch, J. M., Bazo, J., Coughlan de Perez, E., and Stephens, E. M.: Assessing the performance of global hydrological models for capturing peak river flows in the Amazon basin, *Hydrol. Earth Syst. Sci.*, 23, 3057–3080, <https://doi.org/10.5194/hess-23-3057-2019>, 2019.

- UNDRR: Global Assessment Report 2015, Making Development Sustainable: The Future of Disaster Risk Management, UN Office for Disaster Risk Reduction, Geneva, ISBN 978921132042, 2015a.
- UNDRR: Sendai Framework for Disaster Risk Reduction 2015–2030, United Nations Office for Disaster Risk Reduction, Geneva, Switzerland, <https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030> (last access: 3 December 2022), 2015b.
- van der Knijff, J. M., Younis, J., and de Roo, A. P. J. D.: LIS-FLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation, *Int. J. Geogr. Inf. Sci.*, 24, 189–212, <https://doi.org/10.1080/13658810802549154>, 2010.
- Vitart, F.: Evolution of ECMWF sub-seasonal forecast skill scores, *Q. J. Roy. Meteorol. Soc.*, 140, 1889–1899, <https://doi.org/10.1002/qj.2256>, 2014.
- Ward, P. J., Blauhut, V., Bloemendaal, N., Daniell, J. E., de Ruiter, M. C., Duncan, M. J., Emberson, R., Jenkins, S. F., Kirschbaum, D., Kunz, M., Mohr, S., Muis, S., Riddell, G. A., Schäfer, A., Stanley, T., Veldkamp, T. I. E., and Winsemius, H. C.: Review article: Natural hazard risk assessments at the global scale, *Nat. Hazards Earth Syst. Sci.*, 20, 1069–1096, <https://doi.org/10.5194/nhess-20-1069-2020>, 2020.
- Wedi, N. P., Polichtchouk, I., Dueben, P., Anantharaj, V. G., Bauer, P., Boussetta, S., Browne, P., Deconinck, W., Gaudin, W., Hadade, I., Hatfield, S., Iffrig, O., Lopez, P., Maciel, P., Mueller, A., Saarinen, S., Sandu, I., Quintino, T., and Vitart, F.: A Baseline for Global Weather and Climate Simulations at 1 km Resolution, *J. Adv. Model. Earth Syst.*, 12, e2020MS002192, <https://doi.org/10.1029/2020MS002192>, 2020.
- Wilks, D. S.: Statistical methods in the atmospheric sciences, Academic Press, ISBN 9780123850232, 2011.
- Wood, E. F., Roundy, J. K., Troy, T. J., Beek, L. P. H. van, Bierkens, M. F. P., Blyth, E., Roo, A. de, Döll, P., Ek, M., Famiglietti, J., Gochis, D., Giesen, N. van de, Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivalapan, M., Sheffield, J., Wade, A., and Whitehead, P.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, *Water Resour. Res.*, 47, W05301, <https://doi.org/10.1029/2010WR010090>, 2011.
- Zajac, Z., Revilla-Romero, B., Salamon, P., Burek, P., Hirpa, F. A., and Beck, H.: The impact of lake and reservoir parameterization on global streamflow simulation, *J. Hydrol.*, 548, 552–568, <https://doi.org/10.1016/j.jhydrol.2017.03.022>, 2017.
- Zsoter, E., Harrigan, S., Barnard, C., Wetterhall F., Salamon, P., and Prudhomme, C.: River discharge and related forecasted data from the Global Flood Awareness System, v2.1, Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [data set], <https://doi.org/10.24381/cds.fl1aef77>, 2019.
- Zsoter, E., Prudhomme, C., Stephens, E., Pappenberger, F., and Cloke, H.: Using ensemble reforecasts to generate flood thresholds for improved global flood forecasting, *J. Flood Risk Manage.*, 13, e12658, <https://doi.org/10.1111/jfr3.12658>, 2020a.
- Zsoter, E., Harrigan, S., Barnard, C., Blick, M., Ferrario, I., Wetterhall F., and Prudhomme, C.: Reforecasts of river discharge and related data by the Global Flood Awareness System, v2.2, Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [data set], <https://doi.org/10.24381/cds.2d78664e>, 2020b.

A8: Additional figures

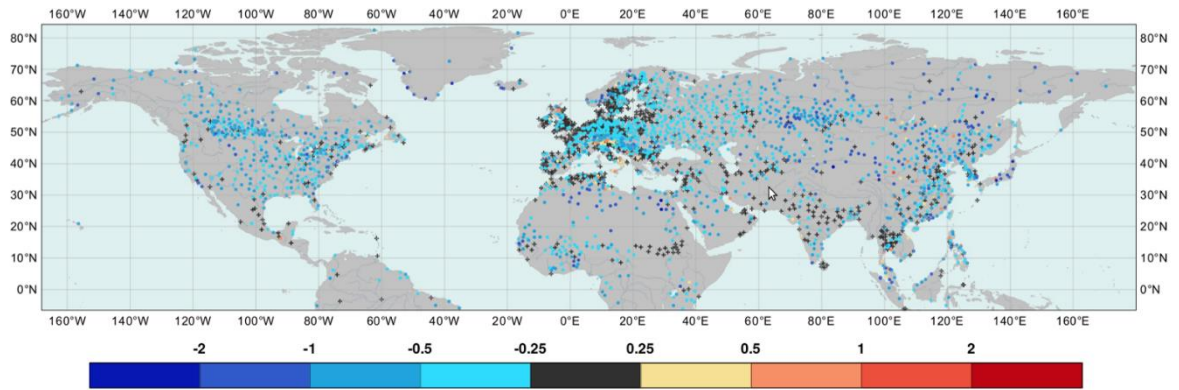


Figure 4-4b. Difference 2m temperature mean absolute errors between ONLINE and OFFLINE for January based on observations in 2000–07 (degrees). Points are shown where observations are available. Blue colours indicate lower errors in the ONLINE experiment. This figure is added to show error differences for 2m temperature, similar to Figure 4-4 for snow depth.

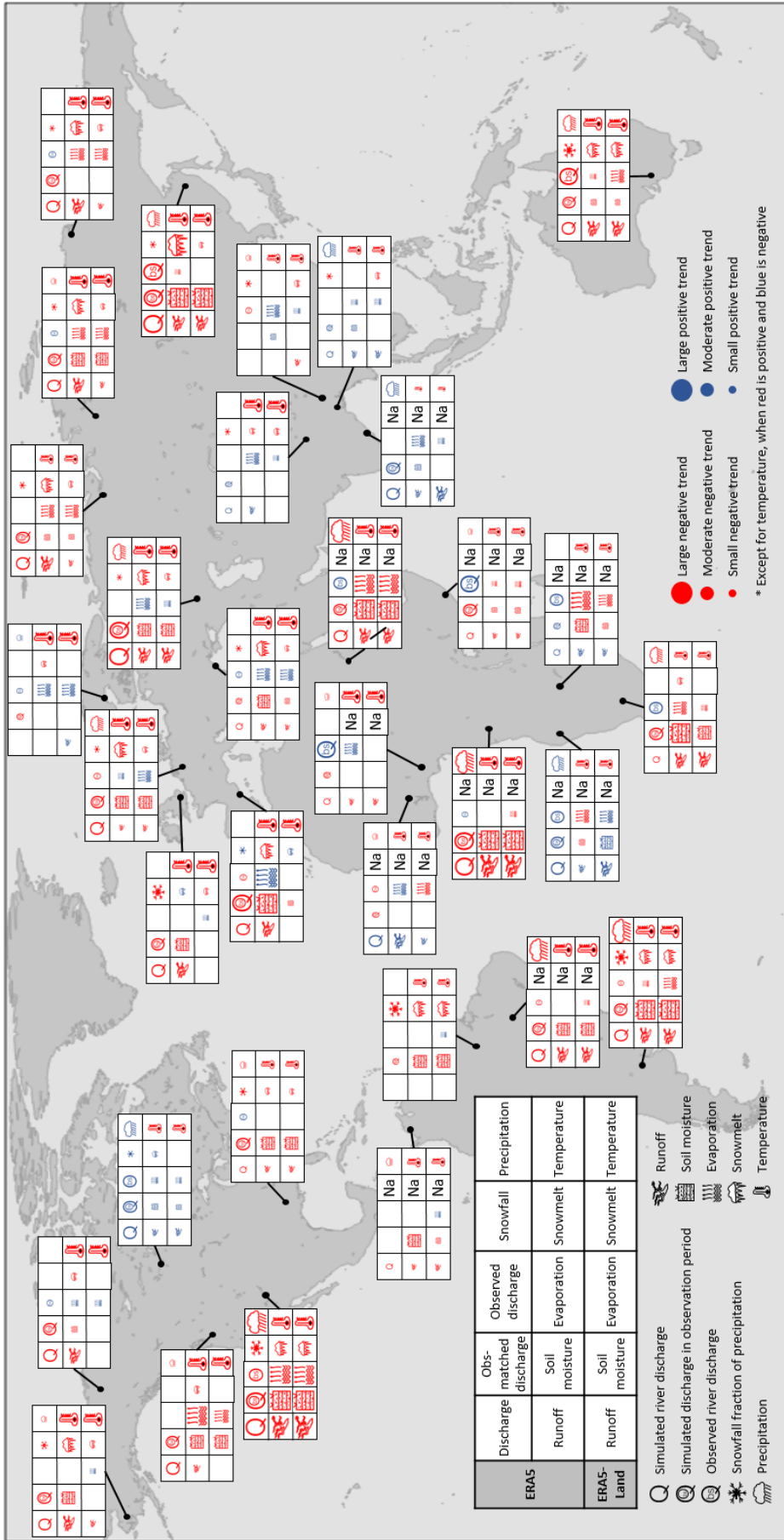


Figure 6-A. Geographical summary of the linear trend values for the analysed river discharge and other land-surface variables for the selected 33 global catchments.