

# *Using ensembles to analyse predictability links in the tropical cyclone flood forecast chain*

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Titley, H. A., Cloke, H. L. ORCID: <https://orcid.org/0000-0002-1472-868X>, Stephens, E. M., Pappenberger, F. and Zsoter, E. (2024) Using ensembles to analyse predictability links in the tropical cyclone flood forecast chain. *Journal of Hydrometeorology*, 25 (1). pp. 191-206. ISSN 1525-7541 doi: 10.1175/JHM-D-23-0022.1 Available at <https://centaur.reading.ac.uk/114239/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1175/JHM-D-23-0022.1>

Publisher: American Meteorological Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

# Using Ensembles to Analyze Predictability Links in the Tropical Cyclone Flood Forecast Chain

H. A. TITLEY<sup>a,b</sup>, H. L. CLOKE<sup>b</sup>, E. M. STEPHENS<sup>b,c</sup>, F. PAPPENBERGER<sup>d</sup>, AND E. ZSOTER<sup>d</sup>

<sup>a</sup> *Met Office, Exeter, United Kingdom*

<sup>b</sup> *University of Reading, Reading, Berkshire, United Kingdom*

<sup>c</sup> *Red Cross Red Crescent Climate Centre, The Hague, Netherlands*

<sup>d</sup> *ECMWF, Reading, Berkshire, United Kingdom*

(Manuscript received 8 February 2023, in final form 15 June 2023, accepted 10 November 2023)

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

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


**KEYWORDS:** Hurricanes/typhoons; Flood events; Tropical cyclones; Ensembles; Probability forecasts/models/distribution; Numerical weather prediction/forecasting

## 1. Introduction

Despite the major role of rain-induced flooding in deaths and damages related to tropical cyclones (TCs) (Rappaport 2014; Czajkowski and Kennedy 2010; Czajkowski et al. 2013) and predictions that TC rainfall rates and associated flooding impacts will increase due to climate change (Knutson et al. 2010, 2013, 2015, 2020; Wang et al. 2015; Wright et al. 2015), the main determining factor in evacuation decision-making remains the intensity of the TC (Stein et al. 2010; Whitehead et al. 2000; Senkbeil et al. 2019). The potential for heavy rainfall is present regardless of the storm's category (Titley et al. 2021) so it is vital to increase public awareness of the dangers of river flooding in TCs. The importance of increasing the focus on the probabilistic forecasting of downstream hazards

has been recognized by key international research programs (Golding 2022; Titley et al. 2019). Improving the understanding of the factors that influence the predictability of flood-related hazards during TCs can help optimize forecast guidance and therefore improve preparedness in advance of an approaching storm. However, investigating this predictability is hampered by a limited sample size of forecast TCs affecting a single region, and the difficulties of aggregating across regions due to geographical differences (e.g., river length and orientation, topography). Ensemble forecasts have the potential to help overcome this issue by providing a larger sample of potential outcomes for what are rare events in any given location.

This work therefore aims to develop a novel methodology to use ensemble forecasts to explore how each part of the TC flood forecast chain (track, intensity, precipitation, and hydrology) influences the overall predictability of river flooding in TC cases. The ensemble forecasts are examined across several forecast runs to create hundreds of sample forecasts for a particular high-impact case. Taking inspiration from previous ensemble-based sensitivity analysis (Hakim and Torn 2008;

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Helen Titley, helen.titley@metoffice.gov.uk

DOI: 10.1175/JHM-D-23-0022.1

© 2024 American Meteorological Society. This published article is licensed under the terms of a Creative Commons Attribution 4.0 International (CC BY 4.0) License



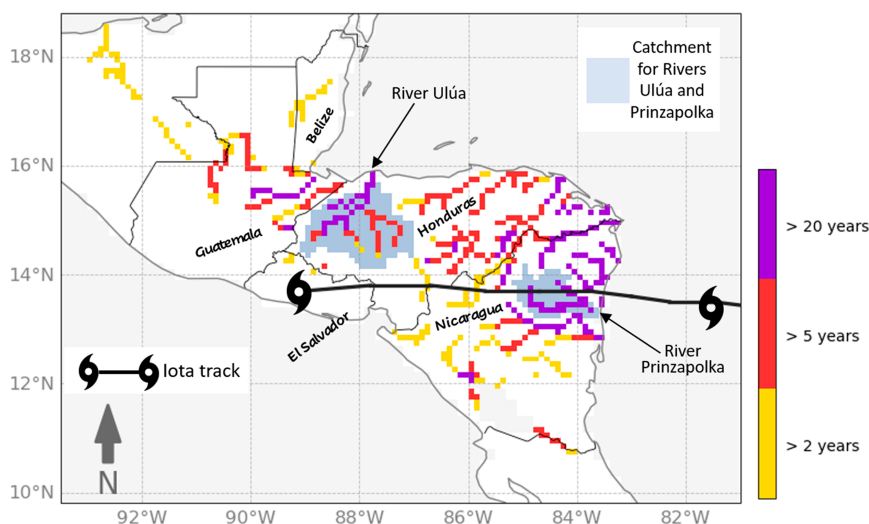


FIG. 1. The return period estimated to have been exceeded from 16 to 20 Nov 2020 in the GloFAS-ERA5 reanalysis. Two of the most severely affected rivers are labeled (the Ulúa and the Prinzapolka, for which the forecasts are examined in section 4), with their catchment areas shown in light blue.

Lynch and Schumacher 2014; Zhang and Meng 2018; Shen et al. 2020), in this study we develop methods to use the ensemble forecasts to investigate the interconnectivity of the various forecast components of the TC flood forecast chain and give valuable insights as to the controls on TC flood predictability. The following questions will be addressed: (i) How do the errors and the spread in the TC track and intensity forecasts relate to the errors and spread in the precipitation forecasts and the river discharge forecasts? (ii) Do the members with the best/worst track forecasts necessarily lead to good/poor precipitation and river discharge forecasts?

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

## 2. Case background

Hurricane Iota was a strong category-4 hurricane on the Saffir–Simpson scale that caused severe and widespread river flooding in Nicaragua, Honduras, and Guatemala, resulting in at least 67 deaths and total damages estimated at \$1.4 billion (Stewart 2021). Iota originated as a tropical wave that moved into the eastern Caribbean on 10 November 2020 before strengthening and being named Tropical Storm Iota on 13 November. Iota became hurricane strength on 15 November, and rapidly intensified to a category-4 hurricane, with peak maximum wind speeds of 135 kt ( $1 \text{ kt} \approx 0.51 \text{ m s}^{-1}$ ), before making landfall with an intensity of 125 kt at 0340 UTC 17 November, south of Puerto Cabezas in northeast Nicaragua. Several sites across Nicaragua and Honduras reported event rainfall totals over 200 mm, with a maximum total of 510 mm in Guatemala (Stewart 2021). The extreme rainfall led to estimated floods exceeding the 1-in-20-yr thresholds across major rivers in these countries (Fig. 1), as well as widespread flash flooding and mudslides. Flooding was exacerbated by pre-existing flood conditions caused by Hurricane Eta, which had made landfall in a similar location two weeks earlier. In total 5.2 million people were affected, with widespread damage to properties and infrastructure (IFRC 2020; UN OCHA 2020a,b).

## 3. Data and methods

The GloFAS forecast system is driven by ECMWF ensemble (ENS) forecasts. The ECMWF ENS consists of an ensemble of 51 forecasts with a current spatial resolution of 18 km: one unperturbed control member and 50 perturbed members whose initial states and model physics have been perturbed to explore the currently understood range of uncertainty in the observations and the model. The initial condition uncertainty

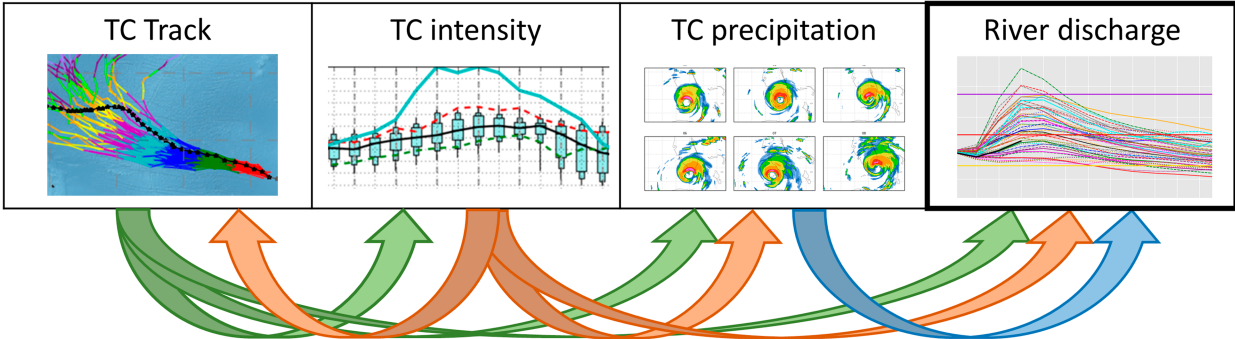


FIG. 2. Conceptual diagram illustrating four of the main components of the TC river flood forecast process, along with the potential links in predictability between them. The boxes contain the four aspects of the forecast that are verified in each ensemble member in each forecast run in the lead up to the flood event, while the arrows represent the potential of each component to influence the skill of the others (green arrows = from TC track; orange arrows = from TC intensity; blue arrows = from TC precipitation). It is these links and dependencies that are evaluated in this paper (e.g., each arrow has a matching scatterplot in Fig. 9).

is sampled using perturbations from an ensemble of data assimilations (EDA) and perturbations constructed from the leading singular vectors, while the model uncertainties are represented with the stochastically perturbed parameterization tendencies scheme (SPPT) that simulates the effect on forecast uncertainty of random model errors due to the parameterized physical processes (ECMWF 2020). The ensemble members provide a range of possible future weather states, with the methods designed to ensure that each ENS member is truly independent of all the others (Owens and Hewson 2018).

From the setup of the GloFAS forecast system it is possible to assess the performance of each component of the flood forecast chain (track, intensity, precipitation, and discharge) within each ensemble member from each successive forecast in the leadup to the flood event. Comparing the performance of the ensemble forecasts between components and from run to run

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

a. Tropical cyclone track (position and intensity)

1) FORECAST DATA

The forecasted TC track position and intensity information from each of the 51 ECMWF ENS ensemble members were extracted from the TIGGE CXML archive (NCEP et al. 2008), for each 0000 UTC forecast run from 8 to 17 November 2020.

TABLE 1. A summary of the forecast data, verification data, and metrics computed for each component of the TC flood forecast chain.				
	TC track	TC intensity	TC precipitation	River discharge
Forecast data	ECMWF ENS tracks from the TIGGE CXML archive (NCEP et al. 2008)	ECMWF ENS tracks from the TIGGE CXML archive (NCEP et al. 2008)	ECMWF ENS precipitation forecast data ( $0.2^{\circ} \times 0.2^{\circ}$ resolution) (ECMWF 2022)	Global Flood Awareness System (GloFAS) (v2.1, $0.1^{\circ} \times 0.1^{\circ}$ resolution) (Zsoter et al. 2019)
Verification data	International Best Track Archive for Climate Stewardship (IBTrACS) (Knapp et al. 2010, 2018)	IBTrACS (Knapp et al. 2010, 2018)	GPM IMERG (final version, $0.1^{\circ} \times 0.1^{\circ}$ resolution) precipitation data (Huffman et al. 2019)	GloFAS-ERA5 operational global river discharge reanalysis ( $0.1^{\circ} \times 0.1^{\circ}$ resolution) (Harrigan et al. 2019, 2020)
Verification metric	Track error at landfall Track error in the first 24 h over land	Error in mean sea level pressure minima at landfall	Mean absolute error of storm total precipitation	Modified Kling–Gupta efficiency (KGE') (Kling et al. 2012) computed for key river points and across all severely impacted river points

## 2) VERIFICATION DATA AND METHOD

The verification data for TC track position and intensity for Hurricane Iota were obtained from the International Best Track Archive for Climate Stewardship (IBTrACS) (Knapp et al. 2010, 2018). Forecast tropical cyclone tracks are deemed to be matched with Hurricane Iota if any point on the track is within 500 km of a matching observed position from IBTrACS. All points from that track are then included in the analysis provided that their forecast position is within 1000 km of the matching observed position. The TC track error in each of the ensemble members with a matching tropical cyclone was then calculated as follows: (i) the landfall track error, defined as the distance between the observed and forecast position, in km, at the time of the last pre-landfall observed track point (0000 UTC 17 November 2020); (ii) the average track error across land; in the Iota case this is defined as the average of the distances (in km) from each of the 6-hourly forecast and observed track positions in the first 24 h that the storm was over land, to ensure that storm tracks are available for most members. The intensity error calculated is the error in mean sea level pressure (mslp) minima at the last prelandfall point, prior to any weakening on landfall.

### b. Tropical cyclone precipitation

#### 1) FORECAST DATA

ECMWF ENS precipitation forecast data for each 0000 UTC forecast run from 8 to 17 November 2020 were extracted from the ECMWF archive (ECMWF 2022). The storm total precipitation was calculated as being the 96-h accumulation from 0000 UTC 16 November to 0000 UTC 20 November within 500 km of the Hurricane Iota track. This represents the area of precipitation defined as being associated with the cyclone, as used in multiple studies that assign a set radius when defining TC-related rainfall (Prat and Nelson 2013, 2016; Jiang et al. 2011; Luitel et al. 2018; Tittley et al. 2021). A land/sea mask was applied to focus on precipitation over land, as relevant for river flooding.

#### 2) VERIFICATION DATA AND METHOD

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

### c. River discharge

#### 1) FORECAST DATA

The setup of the operational GloFAS forecast at the time of Hurricane Iota (v2.1) (Alfieri et al. 2013; Harrigan et al. 2023)

utilized ECMWF ENS meteorological forecasts coupled with the land surface module HTESSEL (Balsamo et al. 2011) and the Lisflood hydrological routing model (Van Der Knijff et al. 2010) to produce 51 possible forecast evolutions of the streamflow through the global river network. GloFAS version 2.1 forecasts (Zsoter et al. 2019) were downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) for each 0000 UTC forecast run from 8 to 17 November 2020.

## 2) VERIFICATION DATA AND METHOD

Ongoing difficulties in the availability of reliable spatially distributed river discharge observations from river gauges (Lavers et al. 2019) mean that reanalyses provide the only practical source of river discharge verification data for the purposes of this study. Therefore, the verification dataset utilized in this study is the GloFAS-ERA5 operational global river discharge reanalysis that accompanies GloFAS v2.1 (Harrigan et al. 2019, 2020). An objective analysis of the GloFAS forecasts is carried out by calculating the modified Kling-Gupta efficiency (KGE') (Gupta et al. 2009; Kling et al. 2012) to compare the forecast and verification discharge data. The KGE' is an expression of the distance away from ideal model performance as described by three components (correlation, variability bias and mean bias).  $KGE' = 1$  indicates perfect agreement between simulations and observations. The KGE' score for a mean flow benchmark is  $KGE' \approx -0.41$  (Knoben et al. 2019). The KGE' is calculated for the 1–15-day forecasts from all GloFAS ensemble members for key river points and for all severely impacted river points across the region (defined as those points with an upstream area greater than 1000 km<sup>2</sup> that experienced greater than a 5-yr return period during the event). The return period thresholds used in the study were as adopted within the GloFAS web interface, calculated by fitting a Gumbel extreme value distribution to the GloFAS-ERA5 river discharge reanalysis over the 1979–2018 period (Alfieri et al. 2019; Harrigan et al. 2023; Zsoter et al. 2020).

### d. Proposed methodology to evaluate the predictability linkages along the forecast chain

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

- (i) Box-and-whisker plots are used to represent how the track, intensity, and precipitation errors and the discharge KGE' scores (across all ensemble members) compare in terms of their mean and spread for each forecast run in the leadup to the event.
- (ii) A “divergence” score (Richardson et al. 2020) is also calculated to allow a comparison of the run-to-run differences between each component of the forecast. In this score the difference between two ensembles is

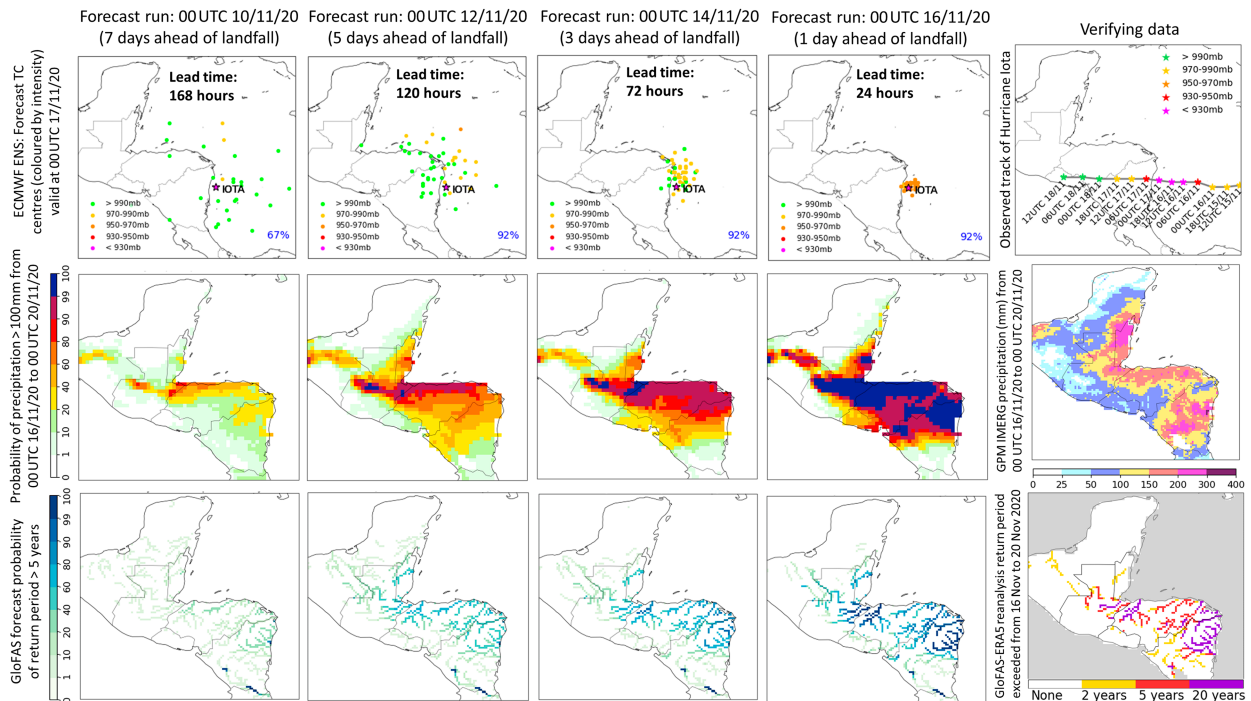


FIG. 3. Hurricane Iota cyclone (top) position, (middle) rainfall, and (bottom) flood forecasts from the forecast runs of ECMWF ENS and GloFAS in the lead up to the event. The first four columns show the forecast runs from 7, 5, 3, and 1 days prior to landfall, while the right column shows the corresponding verification data for track, rainfall, and flood return period exceedance. In the top row, the star shows the observed center position of TC Iota at 0000 UTC 17 Nov 2020 (just prior to landfall), and the dots show the forecast positions valid at this time from each ensemble member, colored by the mean sea level pressure minima. The percentages indicate the percentage of members that had a matched cyclone center forecast at this lead time. The second row shows the precipitation probability forecasts from EC ENS for 4-day event precipitation totals  $> 100$  mm. The bottom row shows the GloFAS forecast probability for river discharge exceeding the 5-yr return period on each river point.

calculated using the divergence function associated with the continuous ranked probability score (CRPS) (Gneiting and Raftery 2007). In our case the two ensembles we are comparing are the ensemble member scores from each component on two successive runs, with the divergence score indicating how much the forecast performance is changing from run to run.

- (iii) Comparing the divergence scores from each part of the forecast (the track, intensity, precipitation, and discharge components) allows us to assess how closely matched the run-to-run jumps in forecast performance are between each component.
- (iv) The correlation between the individual ensemble member scores for each component from each forecast lead time is also calculated using Spearman rank correlation, and displayed alongside scatterplots to show how closely each aspect of forecast performance is tied to each other.
- (v) “Good” and “bad” members are identified from their performance in each part of the forecast chain and their performance in other components examined, in order to delve deeper into the controls on predictability and how these vary from place to place and from run to run.

## 4. Results

### a. Evolution of forecasts for Iota

A summary of the TC position and intensity, precipitation, and discharge forecasts for Hurricane Iota is shown in Fig. 3. At seven days ahead of landfall around two-thirds of members are forecasting a TC to develop, but with a very large spread in forecast position. By five days ahead of landfall over 90% of members are forecasting a TC, and by three days ahead the forecast storm location is much more confident. The intensity of Hurricane Iota is underpredicted at all lead times, as the forecast models struggled to resolve the rapid intensification process. The observed mean sea level pressure (mslp) minimum at the prelandfall point was 918 mb (1 mb = 1 hPa), but even at a one-day lead time the minimum mslp forecasted by any of the ensemble members was 951 mb (Fig. 3, top row). Despite this, the heavy rainfall along the north coast of Honduras is predicted with high probabilities ( $>60\%$ ) around 5 days ahead of landfall (Fig. 3, middle row), but heavy rainfall close to the landfall location on the east coast of Nicaragua was not predicted with high probabilities until later forecast runs closer to the time of landfall. This suggests that the heavy rainfall close to the landfall location was more reliant on having a more accurate forecast for the landfall location and intensity, which was only possible

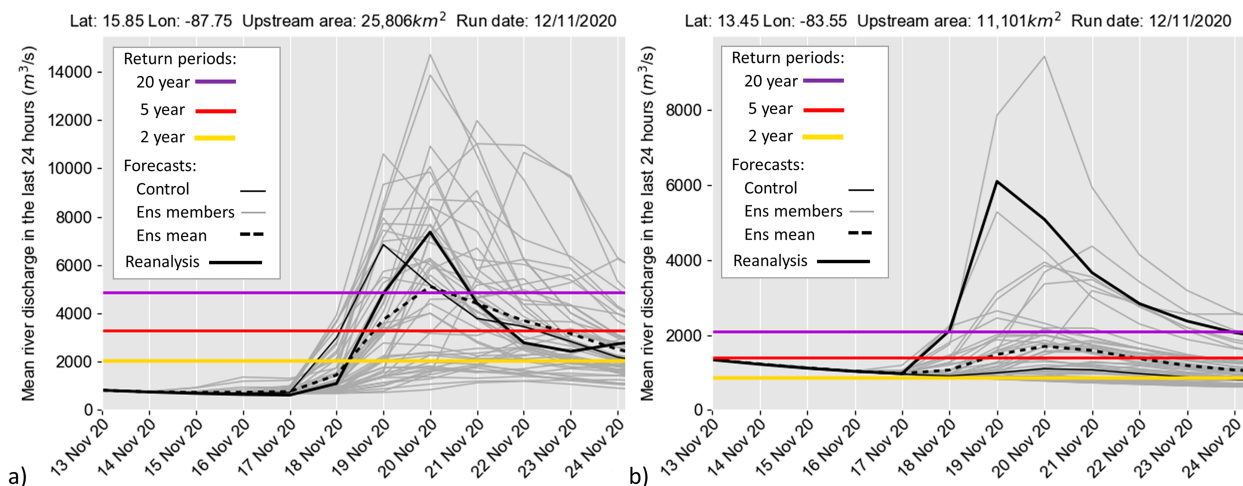


FIG. 4. GloFAS forecast hydrographs from 0000 UTC run on 12 Nov 2020 for the rivers (a) Ulúa and (b) Prinzapolka (ensemble members: gray; control member: thin black; ensemble mean discharge forecast: black dashed). The yellow, red, and purple horizontal lines show the 2-, 5-, and 20-yr return period discharge levels for that river. The reanalysis discharge is overlain in a thick black line.

at shorter lead times. The northward bias of some earlier runs (Fig. 3, top row) still produced forecasts of significant rainfall for northern Honduras.

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

Two of the rivers that experienced the most severe flooding from Hurricane Iota were the River Ulúa in northern Honduras and the River Prinzapolka in eastern Nicaragua. To examine the river discharge forecasts for these rivers, the GloFAS

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

#### b. Flow of uncertainty and predictability through the forecast chain

The previous section presented the evolution of forecasts for Hurricane Iota, identifying potential linkages between the different forecast components. To investigate the flow of uncertainty and predictability more thoroughly, a detailed analysis comparing the ensemble performance at each part of the forecast chain is proposed and illustrated. Forecast performance across the ensemble members is presented side-by-side for each component of the forecast: track error (Fig. 5a), intensity error (Fig. 5b), precipitation error (Fig. 5c), and KGE' score for the discharge forecasts (Fig. 5d) across all severely impacted river points as well as the two rivers that were examined in Fig. 4 (the Ulúa and the Prinzapolka). Flooding on the River Ulúa was well predicted further in advance than the flooding on the River Prinzapolka, which was closer to the landfall location, confirming the findings in section 4a. Using the Towner et al. (2019) thresholds to evaluate the KGE' scores as “good” ( $KGE' \geq 0.75$ ), “intermediate”

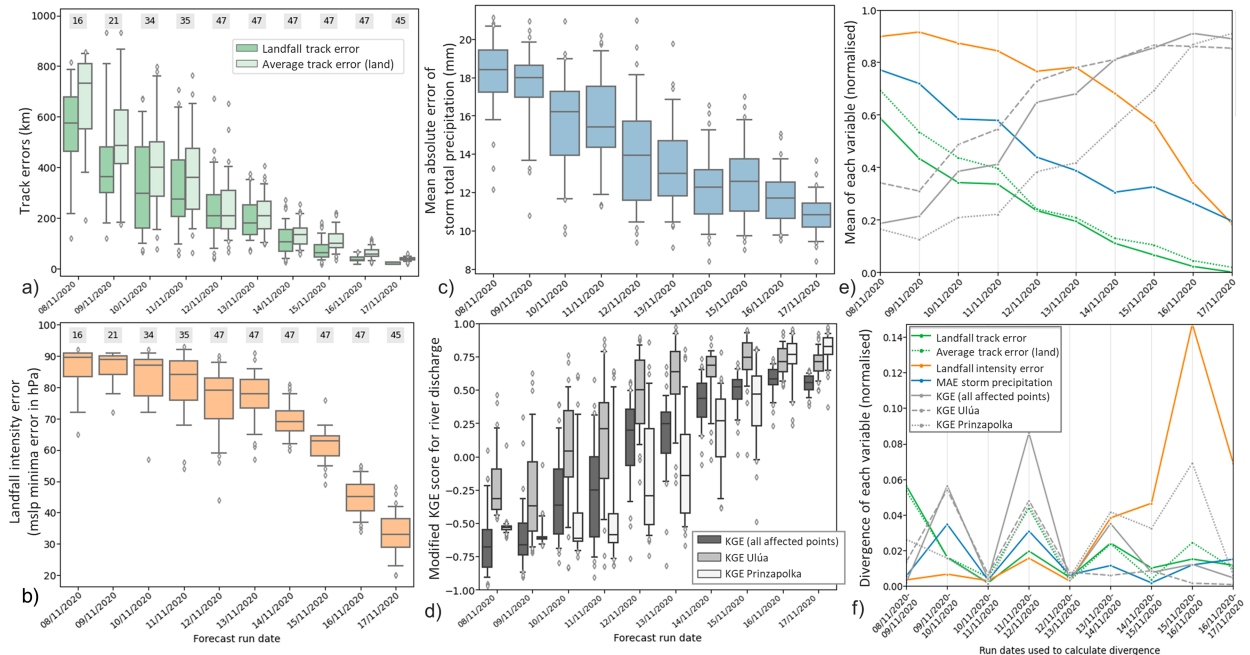


FIG. 5. Boxplots comparing forecast performance across the ensemble members for each 0000 UTC forecast run in the lead up to the Hurricane Iota event: (a) forecast track error (landfall error and average track error over land); (b) forecast landfall intensity (mslp minima) error; (c) forecast precipitation error (mean absolute error of the storm total precipitation (within 500 km of TC track)); (d) discharge forecast performance as measured by the average KGE' score at all severely impacted river points (where 5-yr return period is exceeded in the GloFAS reanalysis), and at the River Ulúa/River Prinzapolka outlet river points; (e) a comparison of the means of these metrics, normalized by the y-axis range in each subplot and color-coded from the plots and key as displayed in (f); and (f) divergence score of these metrics, normalized by the y-axis range and color coded as shown in the key. In (a) and (b), errors can only be computed for those ensemble members with a tracked TC (numbers overlain in gray).

( $0.75 > \text{KGE}' \geq 0.5$ ), “poor” ( $0.5 > \text{KGE}' > 0$ ) and “very poor” ( $\text{KGE}' \leq 0$ ), the majority of ensemble members have intermediate performance 5 days ahead of landfall for the River Ulúa, but only 2 days ahead of landfall for the River Prinzapolka. An examination of IMERG satellite data (Fig. 6) reveals that the Ulúa catchment rainfall was mainly associated with onshore flow of the outer rainbands with orographic enhancement, which fell over a longer period and is less sensitive to the exact landfall location. For the River Prinzapolka,

the center of the storm passed directly over the catchment leading to very heavy rainfall in a shorter period during and after landfall. Successful flood forecasts therefore relied on the accurate prediction of landfall location and of the heavy rainfall close to the TC center, which was less predictable at longer lead times. The average KGE' score across all impacted points (all river points where the 5-yr return period was exceeded in the reanalysis) falls between these two extremes.

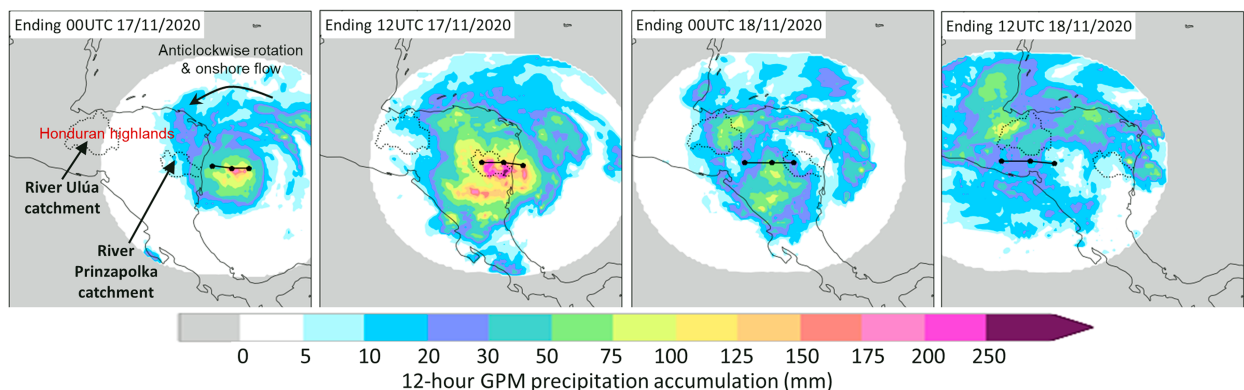


FIG. 6. Precipitation accumulation across Central America from Hurricane Iota in four 12-h periods from 1200 UTC 16 to 18 Nov 2020. The river catchments for the River Ulúa and River Prinzapolka are shown in dashed lines, with labels on the left plot.

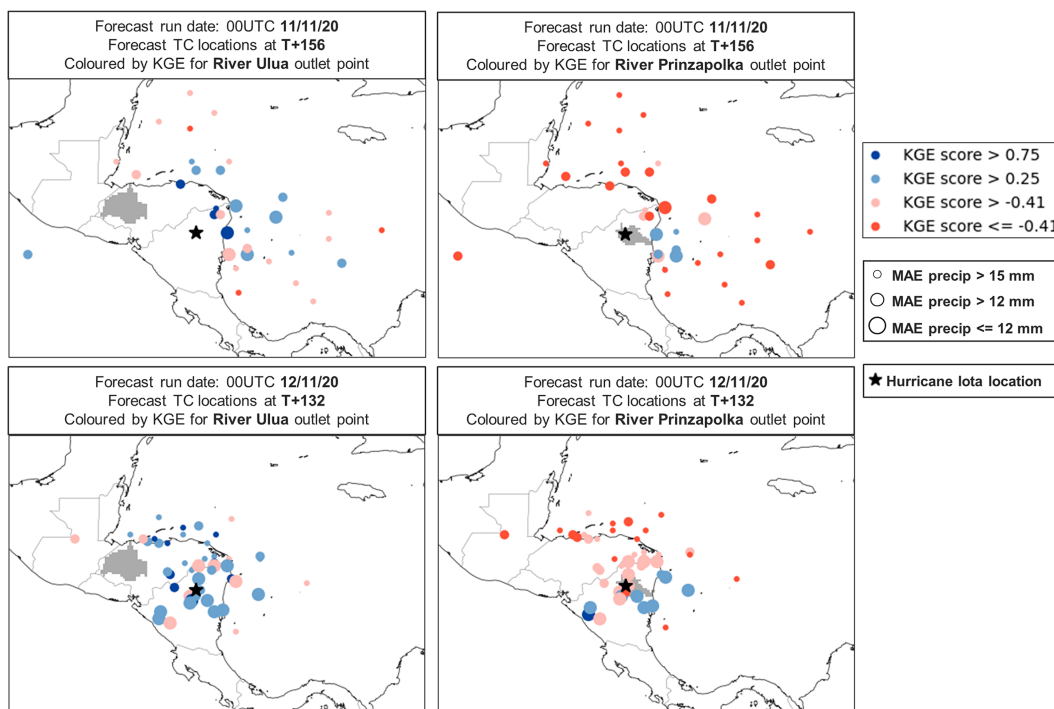


FIG. 7. Forecast TC locations in the ECMWF ENS forecasts from (top) 0000 UTC 11 Nov 2020 and (bottom) 0000 UTC 12 Nov 2020. The centers are colored according to the modified KGE' score for the (left) Ulúa and (right) Prinzapolka outlet river points as defined in section 4a. The centers are sized by the mean absolute error in the storm total precipitation field (larger circles = better forecasts). The verifying position of Hurricane Iota at this time is marked with a black star, and the catchment areas for the two river outlet points are shown in gray for reference.

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

closer to the landfall point, the rainfall it experiences is predominately from the area close to the center of the storm (e.g., the eyewall), and the intensity of the precipitation in this part of the storm is greater for intense storms.

### c. Member-specific analysis

The consecutive forecast runs on 11 and 12 November were found in Fig. 5f to have high divergence scores (i.e., significant run-to-run changes in ensemble performance). The increase in forecast confidence from 11 to 12 November in Iota's position forecast the day after landfall is notable with a much more concentrated pattern of ensemble member TC positions (Fig. 7). The additional member-specific analysis here illustrated in Fig. 7 is to color the forecast positions by their KGE' scores on two rivers' points that have been focused on: the Ulúa (left) and the Prinzapolka (right) outlet points. This allows us to provide an impact-relevant hydrological assessment of which ensemble members provided the most accurate flood forecast for each river (those colored in blue). Both before and after the increase in forecast confidence, the members with a good forecast for the River Prinzapolka all have the TC forecast center relatively close to the observed location, although there are also several ensemble members with forecast centers close to the observed location that do not have good KGE scores. A broader range of ensemble members based on TC position perform well for the Ulúa catchment.

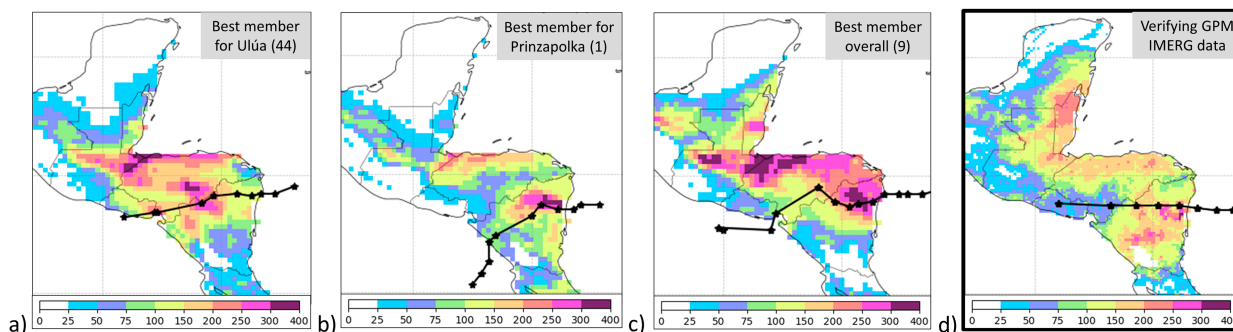


FIG. 8. Hurricane Iota forecast track and forecast event precipitation forecast totals (mm) from the 0000 UTC 12 Nov 2020 forecast run from the EC ENS ensemble members with the highest KGE' scores for (a) the River Ulúa outlet point, the (b) River Prinzapolka outlet point, and (c) the average of all severely impacted river points. (d) The verifying GPM IMERG precipitation (mm) along with the observed track of Hurricane Iota.

The TC centers in Fig. 7 are sized according to the MAE in storm total precipitation to allow an inspection of the relationship between three aspects of the forecast: the TC track (position), precipitation, and river discharge skill. The ensemble members with better overall precipitation forecasts (lower MAE), shown as larger dots, are mainly those that are forecasting the TC center to be close to its observed position. Several of the ensemble members with TC positions too far to the north are shaded blue in the left plots as they score well for river discharge on the Ulúa, but they have a small size indicating they do not score well overall for precipitation. Their better score for the Ulúa discharge is partly due to their more northerly tracks bringing forecasts of heavy rain from the central part of the storm over the Ulúa catchment. The center of the storm tracked farther south in reality and so the overall precipitation errors in these members are higher as they fail to forecast the precipitation farther south.

To take this member-specific analysis a step further, the discharge scores from each ensemble member from a particular run can be ranked to determine the “best” performing members. This allows us to examine the other aspects of forecast performance for those members, to better understand why they led to good flood forecasts. This is illustrated for the 0000 UTC 12 November run in Fig. 8. Interestingly in this case, the landfall position of all three of the identified best members is similar, and is close to the observed landfall location. However, their forecast tracks after landfall are different, with the best performing member for the River Ulúa being closest to the observed track (Fig. 8a), while the best member for the River Prinzapolka turns southwest shortly after landfall, and therefore underestimates precipitation totals farther north (Fig. 8b). Despite their similar landfall locations, the forecast precipitation patterns are very different, with large variation in the precipitation totals even close to the landfall location in all three. When comparing to the verifying IMERG data, the member that had the highest KGE' score when averaged over affected river points across the region (Fig. 8c) had the best match in terms of the overall extent of the rain event, but was forecasting significantly greater totals in several areas compared to IMERG. However, rain totals of

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

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

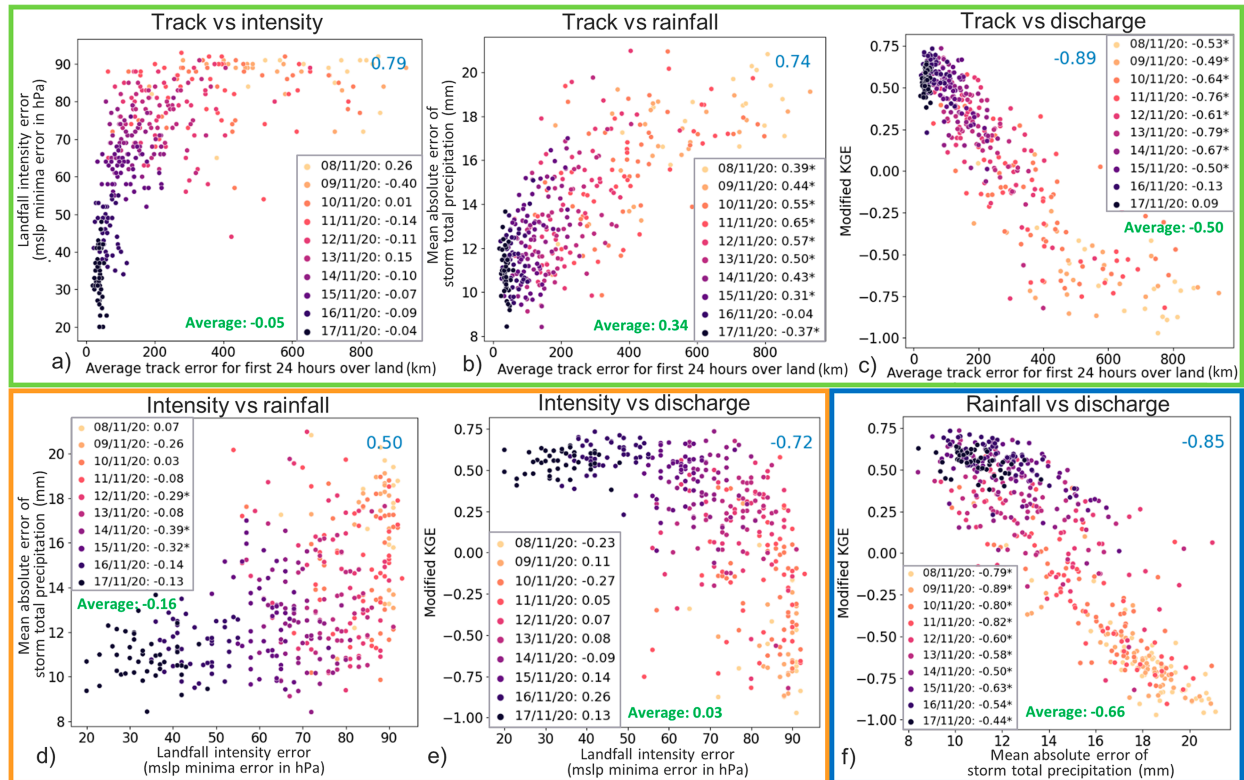


FIG. 9. Scatterplots of each error type/score where each dot is one member: (a) average track error over land vs landfall intensity error; (b) average track error over land vs MAE of storm total precipitation; (c) average track error over land vs average KGE' score at all severely impacted river points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

Looking at the correlations at each forecast run time (in tables in Fig. 9, with average of these in green) shows the predictability links between each component for particular forecast runs. The correlation between track and intensity error is not significant at any lead time (Fig. 9a). There are, however, significant correlations between track and precipitation, and track and river discharge at all forecast times other than the latest two, by which point the track errors were very small (Figs. 9b,c). Interestingly, the correlation between the track errors and river discharge skill is stronger than track errors against the precipitation errors, which may be because discharge skill depends on the total catchment rainfall, where the smaller grid point to grid point differences that influence the precipitation errors are balanced out to some degree. The slightly lower correlation with precipitation may also be due to the independence of the precipitation verification data, as opposed to the model-based reanalysis data that is used to verify the discharge forecast, and may also be impacted by potential underestimates of the rainfall totals by GPM IMERG. The intensity error at landfall is not significantly correlated with river discharge skill at any forecast lead time (Fig. 9e). Although there is a positive correlation overall between the

intensity error and rainfall error, there is actually a counterintuitive negative correlation at three of the forecast runs, when you take an average of the correlations at each forecast run (Fig. 9d). So, in a particular forecast cycle, the members with better forecasts for TC intensity do not tend to be those with better precipitation forecasts or river discharge forecasts. There are however significant correlations as expected between the precipitation error and the river discharge score (Fig. 9f). It is therefore crucial to verify and understand the skill of ensemble TC rainfall forecasts and to focus future research on understanding and improving rainfall prediction in addition to track and intensity in order to advance flood forecasting in TC cases.

Scatterplots of the two track metrics (landfall error and track error over land) against the KGE' river discharge score for the River Prinzapolka reveal a strong "L-shaped" relationship, with a very quick drop off in forecast skill as track error increases, no forecast skill for higher track errors, and little difference between the plots for the two track error metrics (with just a slightly stronger correlation with landfall track error compared to the average track error over land both overall and in the run-to-run correlations) (Figs. 10b,d,e).

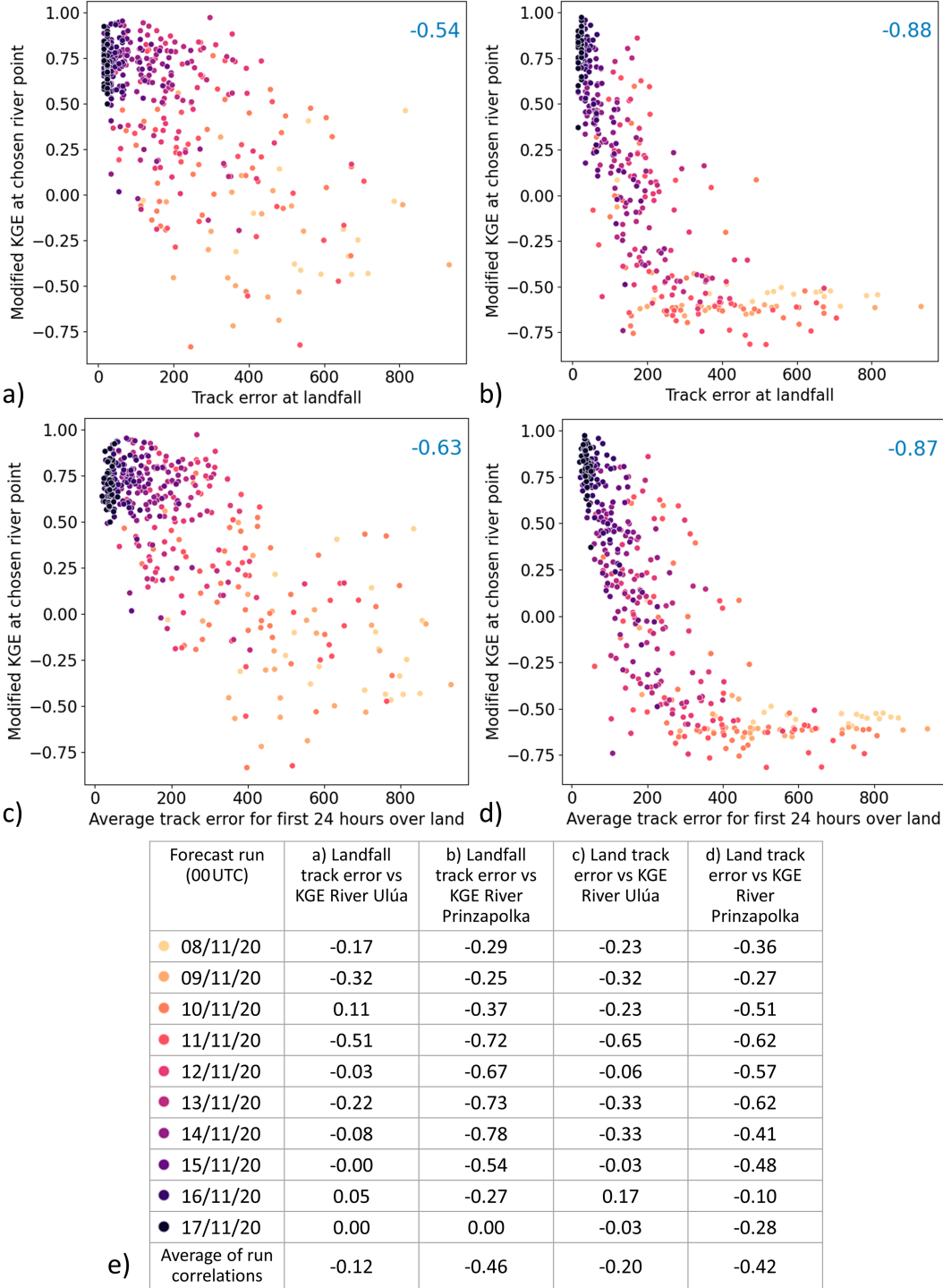


FIG. 10. Scatterplots of the various errors/scores where each dot is one member: (a) Track error at landfall vs KGE' at the River Ulúa outlet point; (b) track error at landfall vs KGE' at the River Prinzapolka outlet point; (c) average track error over land vs KGE' at the River Ulúa outlet point; and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

This shows how crucial an accurate landfall and track forecast was to a skillful flood forecast for this river. In the equivalent plots for the River Ulúa (Figs. 10a,c,e), this L-shaped relationship is not observed and there is a more gradual drop in KGE' as track error increases, and so an accurate landfall forecast was less crucial for this river. The correlation between track error and river discharge skill is lower compared to the Prinzapolka, and there is a stronger relationship with average land track error rather than landfall error (overall and in the run-to-run correlations), showing the greater influence of the detail of the forecast track after landfall on how much rain fell in the Ulúa catchment in northern Honduras.

## 5. Discussion

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

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

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

The average correlation between rainfall and river discharge skill is stronger than just track error against river discharge at most forecast runs and the analysis of a few strong performing members from one of the forecast runs reveals that even similar track forecasts can have very different precipitation patterns. This, combined with the finding that the spread in precipitation forecast errors across the ensemble remains relatively large even after the track spread reduces, highlights the importance of understanding and optimizing precipitation forecasts in the ensemble models that are used to drive ensemble flood forecasts. Traditionally ensemble TC forecast verification at numerical weather prediction (NWP)

centers has focused on track and intensity (Magnusson et al. 2019, 2021; Titley et al. 2020; Zhou et al. 2022), and the relatively crude precipitation verification done here does not provide all the answers as to which aspects of the precipitation forecasts would benefit from improvement. More detailed verification of ensemble precipitation forecasts in TC cases is recommended, particularly using spatial or object-oriented methods to highlight which specific aspects of rainfall forecasts would benefit from improvement in order to feed into improvements in flood forecasts.

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

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

There are limitations to the methods as described here that could be improved in subsequent case studies. The use of reanalyses as the river discharge verification dataset has the advantage of being available at every global river point, and further work is ongoing to address their known limitations (Harrigan et al. 2020; Zsoter et al. 2020), but if reliable discharge observations were more widespread and openly available this would benefit studies of this kind. The rainfall verification metric used here is a very simple point-based metric and other metrics that remove the point-by-point sensitivity such as the fractions skill score (Roberts and Lean 2008) could be used. It would also be useful to calculate the forecast catchment-total rainfall in addition to the overall storm-total rainfall when analyzing the forecasts for specific

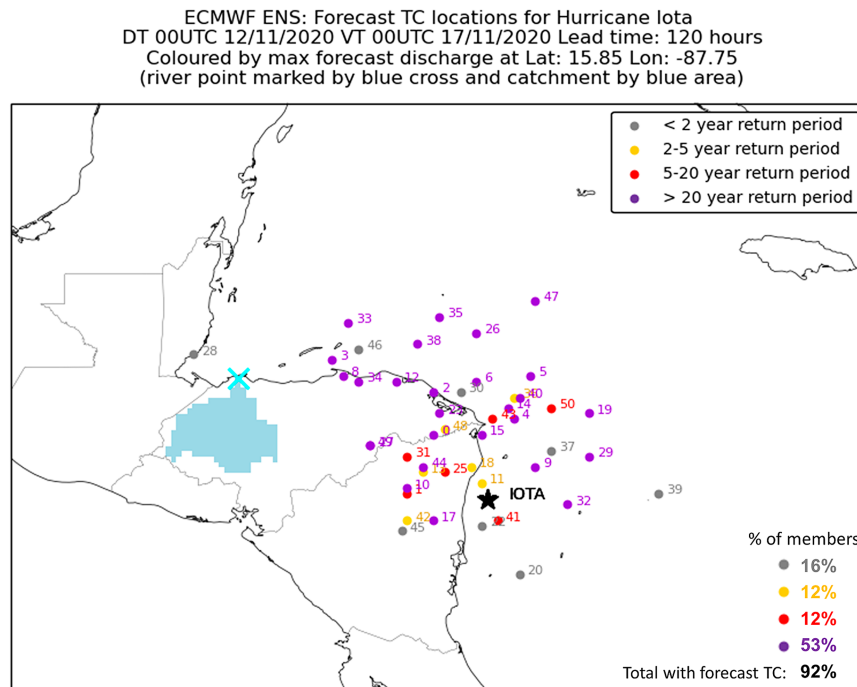


FIG. 11. Prototype plot that could be produced in the run up to a TC event to highlight the forecast TC locations of those ensemble members that are forecasting a significant flood event at a particular river point. Here the outlet point on the River Ulúa has been used (cyan cross). The day 5 ECMWF ENS forecast positions for Hurricane Iota from 0000 UTC 12 Nov 2020 forecast run are displayed (valid at 0000 UTC 17 Nov 2020 close to the time of landfall). The forecast positions are numbered by the ensemble member number (to allow the individual ensemble members to be followed through the forecast chain) and colored specifically by the return period that is forecast to be exceeded on the River Ulúa. The thresholds and colors are the same as those used on the GloFAS web interface <https://www.globalfloods.eu/>.

ivers. A further limitation of this study is the reliance on GPM IMERG data to verify the TC rainfall. GPM is an incredibly useful resource due to its global coverage and wide accessibility, and the Final Run data used here incorporates additional microwave passes and a monthly gauge correction. However, some of the ground-based observations in Iota indicate potential errors in capturing the true rain rates in some regions, which may impact the robustness of the analysis of the errors of the precipitation forecasts. More detailed investigation of biases in observations of TC precipitation data is recommended, comparing IMERG data with other sources of precipitation “truth” data such as radar and gauge-based datasets, in order to develop better understanding of observation uncertainty and improve the confidence in TC rainfall verification results.

This study has focused on a particular case, Hurricane Iota, and the resultant flooding that occurred in Honduras and Nicaragua. In the event of a cyclone with a similar forecast track impacting the region, the findings of this study could help in giving greater confidence to longer-range flood forecasts for those catchments that are forecasted to be impacted by the wider storm circulation and orographically enhanced rainfall, while understanding that the locations and severity of flooding closer to the potential landfall site will be more difficult to predict until

closer to the landfall time. However, it is important to understand that even in the same region, no two cyclones are the same in terms of their track, speed, size, and wider environmental conditions, and so care should be taken in applying the understanding gained in this study to future storms. The wider applicability of these results in other geographical contexts is also uncertain because flood severity also depends on the local landscape and catchment characteristics. The purpose of this study was to provide a suggested approach and set of methods to carry out an analysis of the flow of predictability through the flood forecast chain in TC cases. The wider application of these methods across a range of cases would help to inform 1) the robustness and wider applicability of the findings of this case and 2) how the flow of predictability through the flood forecast chain varies with different geographical and forecast contexts.

## 6. Conclusions

The work adds a new hydrological perspective to the traditional methods of verifying global ensemble TC forecast performance. It proposes the use of ensemble-based methods to analyze the predictability links between the track, intensity, precipitation, and discharge forecast components of the ensemble flood forecast chain

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

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

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

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

**Acknowledgments.** Author Helen Titley acknowledges the support of the Weather and Climate Science for Service Partnership (WCSSP) India and Southeast Asia projects, supported by the Department for Science, Innovation & Technology (DSIT). Liz Stephens and Hannah Cloke acknowledge funding from the U.K.'s Natural Environment Research Council (NERC)

and Foreign Commonwealth and Development Office (FCDO) (formerly DFID) Science for Humanitarian Emergencies and Resilience (SHEAR) research program Forecasts for Anticipatory Humanitarian Action (FATHUM) project Grant NE/P000525/1. Author Hannah Cloke also acknowledges funding from the U.K.'s Natural Environment Research Council (NERC) The Evolution of Global Flood Risk (EVOFLOOD) project Grant NE/S015590/1.

**Data availability statement.** The data in this study were accessed from the following available sources: (i) ECMWF ENS tracks from the TIGGE CXML archive at the NCAR Research Data Archive (NCEP et al. 2008; <https://doi.org/10.5065/D6GH9GSZ>); (ii) ECMWF ENS precipitation forecast data from ECMWF Meteorological Archival and Retrieval System (MARS) (ECMWF 2022); (iii) GloFAS river discharge forecasts from the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) (Zsoter et al. 2019; <https://doi.org/10.24381/cds.ff1aef77>); (iv) IBTrACS data from the National Centers for Environmental Information (NCEI) (Knapp et al. 2018; <https://doi.org/10.25921/82ty-9e16>); (v) GPM IMERG (final version) precipitation data from NASA (Huffman et al. 2019); and (vi) GloFAS-ERA5 operational global river discharge reanalysis from the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) (Harrigan et al. 2019; <https://doi.org/10.24381/cds.a4fdd6b9>).

## REFERENCES

- Alfieri, L., P. Burek, E. Dutra, B. Krzeminski, D. Muraro, J. Thielen, and F. Pappenberger, 2013: GloFAS—Global ensemble streamflow forecasting and flood early warning. *Hydrol. Earth Syst. Sci.*, **17**, 1161–1175, <https://doi.org/10.5194/hess-17-1161-2013>.
- , E. Zsoter, S. Harrigan, F. Aga Hirpa, C. Lavaysse, C. Prudhomme, and P. Salamon, 2019: Range-dependent thresholds for global flood early warning. *J. Hydrol.*, **4**, 100034, <https://doi.org/10.1016/j.hydroa.2019.100034>.
- Balsamo, G., 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. Proc.*, **25**, 1046–1054, <https://doi.org/10.1002/hyp.7808>.
- Czajkowski, J., and E. Kennedy, 2010: Fatal tradeoff? Toward a better understanding of the costs of not evacuating from a hurricane in landfall counties. *Popul. Environ.*, **31**, 121–149, <https://doi.org/10.1007/s11111-009-0097-x>.
- , G. Villarini, E. Michel-Kerjan, and J. A. Smith, 2013: Determining tropical cyclone inland flooding loss on a large scale through a new flood peak ratio-based methodology. *Environ. Res. Lett.*, **8**, 044056, <https://doi.org/10.1088/1748-9326/8/4/044056>.
- ECMWF, 2020: IFS Documentation CY47R1—Part V: Ensemble prediction system. ECMWF Doc. 5, 23 pp., [https://www.ecmwf.int/sites/default/files/elibrary/2020/81190-ifs-documentation-cy47r1-part-v-ensemble-prediction-system\\_1.pdf](https://www.ecmwf.int/sites/default/files/elibrary/2020/81190-ifs-documentation-cy47r1-part-v-ensemble-prediction-system_1.pdf).
- , 2022: Access to archive datasets. Accessed 24 May 2022, <https://www.ecmwf.int/en/forecasts/access-forecasts/access-archive-datasets>.

- Emerton, R., and Coauthors, 2020: Emergency flood bulletins for Cyclones Idai and Kenneth: A critical evaluation of the use of global flood forecasts for international humanitarian preparedness and response. *Int. J. Disaster Risk Reduct.*, **50**, 101811, <https://doi.org/10.1016/j.ijdr.2020.101811>.
- Gneiting, T., and A. E. Raftery, 2007: Strictly proper scoring rules, prediction, and estimation. *J. Amer. Stat. Assoc.*, **102**, 359–378, <https://doi.org/10.1198/016214506000001437>.
- Golding, B., Ed., 2022: *Towards the “Perfect” Weather Warning: Bridging Disciplinary Gaps through Partnership and Communication*. Springer, 280 pp., <https://doi.org/10.1007/978-3-030-98989-7>.
- 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>.
- Hakim, G. J., and R. D. Torn, 2008: Ensemble synoptic analysis. *Synoptic–Dynamic Meteorology and Weather Analysis and Forecasting: A Tribute to Fred Sanders*, Meteor. Monogr., No. 55, Amer. Meteor. Soc., 147–161, [https://doi.org/10.1007/978-0-933876-68-2\\_7](https://doi.org/10.1007/978-0-933876-68-2_7).
- Harrigan, S., E. Zsoter, C. Barnard, F. Wetterhall, P. Salamon, and C. Prudhomme, 2019: River discharge and related historical data from the global flood awareness system, v2.1. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), accessed 3 February 2022, <https://doi.org/10.24381/cds.a4fdd6b9>.
- , and Coauthors, 2020: 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>.
- , 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.*, **27**, 1–19, <https://doi.org/10.5194/hess-27-1-2023>.
- Huffman, G. J., and Coauthors, 2019: NASA Global Precipitation Measurement (GPM) Integrated Multi-satellite Retrievals for GPM (IMERG). Algorithm Theoretical Basis Doc. (ATBD), version 06, 38 pp., [https://gpm.nasa.gov/sites/default/files/document\\_files/IMERG\\_ATBD\\_V06.pdf](https://gpm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V06.pdf).
- IFRC, 2020: Revised emergency appeal: Central America—Hurricane Eta and Iota. Red Cross and Red Crescent International Rep., 36 pp., <https://reliefweb.int/report/honduras/central-america-hurricane-eta-iota-revised-emergency-appeal-n-mdr43007-revision-no-1>.
- Jiang, H., C. Liu, and E. J. Zipser, 2011: A TRMM-based tropical cyclone cloud and precipitation feature database. *J. Appl. Meteor. Climatol.*, **50**, 1255–1274, <https://doi.org/10.1175/2011JAMC2662.1>.
- 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>.
- Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone data. *Bull. Amer. Meteor. Soc.*, **91**, 363–376, <https://doi.org/10.1175/2009BAMS2755.1>.
- , H. J. Diamond, J. P. Kossin, M. C. Kruk, and C. J. Schreck, 2018: International Best Track Archive for Climate Stewardship (IBTrACS) project, version 4. NOAA National Centers for Environmental Information, accessed 9 August 2022, <https://doi.org/10.25921/82ty-9e16>.
- Knoben, W., J. Freer, and R. 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>.
- Knutson, T. R., and Coauthors, 2010: Tropical cyclones and climate change. *Nat. Geosci.*, **3**, 157–163, <https://doi.org/10.1038/ngeo779>.
- , and Coauthors, 2013: Dynamical downscaling projections of twenty-first-century Atlantic hurricane activity: CMIP3 and CMIP5 model-based scenarios. *J. Climate*, **26**, 6591–6617, <https://doi.org/10.1175/JCLI-D-12-00539.1>.
- , J. J. Sirutis, M. Zhao, R. E. Tuleya, M. Bender, G. A. Vecchi, G. Villarini, and D. Chavas, 2015: Global projections of intense tropical cyclone activity for the late twenty-first century from dynamical downscaling of CMIP5/RCP4.5 scenarios. *J. Climate*, **28**, 7203–7224, <https://doi.org/10.1175/JCLI-D-15-0129.1>.
- , and Coauthors, 2020: Tropical cyclones and climate change assessment: Part II: Projected response to anthropogenic warming. *Bull. Amer. Meteor. Soc.*, **101**, E303–E322, <https://doi.org/10.1175/BAMS-D-18-0194.1>.
- Lavers, D. A., S. Harrigan, E. Andersson, D. S. Richardson, C. Prudhomme, and F. Pappenberger, 2019: A vision for improving global flood forecasting. *Environ. Res. Lett.*, **14**, 121002, <https://doi.org/10.1088/1748-9326/ab52b2>.
- Luitel, B., G. Villarini, and G. A. Vecchi, 2018: Verification of the skill of numerical weather prediction models in forecasting rainfall from U.S. landfalling tropical cyclones. *J. Hydrol.*, **556**, 1026–1037, <https://doi.org/10.1016/j.jhydrol.2016.09.019>.
- Lynch, S. L., and R. S. Schumacher, 2014: Ensemble-based analysis of the May 2010 extreme rainfall in Tennessee and Kentucky. *Mon. Wea. Rev.*, **142**, 222–239, <https://doi.org/10.1175/MWR-D-13-00020.1>.
- Magnusson, L., and Coauthors, 2019: ECMWF activities for improved hurricane forecasts. *Bull. Amer. Meteor. Soc.*, **100**, 445–458, <https://doi.org/10.1175/BAMS-D-18-0044.1>.
- , and Coauthors, 2021: Tropical cyclone activities at ECMWF. ECMWF Tech. Memo. 888, 180 pp., <https://www.ecmwf.int/sites/default/files/elibrary/2021/20228-tropical-cyclone-activities-ecmwf.pdf>.
- NCEP, and Coauthors, 2008: THORPEX Interactive Grand Global Ensemble (TIGGE) model tropical cyclone track data. National Center for Atmospheric Research, Computational and Information Systems Laboratory, accessed 26 May 2022, <https://doi.org/10.5065/D6GH9GSZ>.
- Owens, R. G., and T. D. Hewson, 2018: ECMWF forecast user guide. ECMWF Doc. 16559, <https://doi.org/10.21957/m1cs7h>.
- Prat, O. P., and B. R. Nelson, 2013: Mapping the world’s tropical cyclone rainfall contribution over land using the TRMM Multi-Satellite Precipitation Analysis. *Water Resour. Res.*, **49**, 7236–7254, <https://doi.org/10.1002/wrcr.20527>.
- , and —, 2016: On the link between tropical cyclones and daily rainfall extremes derived from global satellite observations. *J. Climate*, **29**, 6127–6135, <https://doi.org/10.1175/JCLI-D-16-0289.1>.
- Rappaport, E. N., 2014: Fatalities in the United States from Atlantic tropical cyclones: New data and interpretation. *Bull. Amer. Meteor. Soc.*, **95**, 341–346, <https://doi.org/10.1175/BAMS-D-12-00074.1>.
- Richardson, D. S., H. L. Cloke, and F. Pappenberger, 2020: Evaluation of the consistency of ECMWF ensemble forecasts. *Geophys. Res. Lett.*, **47**, e2020GL087934, <https://doi.org/10.1029/2020GL087934>.

- Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, <https://doi.org/10.1175/2007MWR2123.1>.
- Senkbeil, J., J. Collins, and J. Reed, 2019: Evacuee perception of geophysical hazards for Hurricane Irma. *Wea. Climate Soc.*, **11**, 217–227, <https://doi.org/10.1175/WCAS-D-18-0019.1>.
- Shen, Y., Y. Du, and G. Chen, 2020: Ensemble sensitivity analysis of heavy rainfall associated with three MCSs coexisting over southern China. *J. Geophys. Res. Atmos.*, **125**, e2019JD031266, <https://doi.org/10.1029/2019JD031266>.
- Speight, L., and Coauthors, 2023: Recommendations to improve the interpretation of global flood forecasts to support international humanitarian operations for tropical cyclones. *J. Flood Risk Manage.*, e12952, <https://doi.org/10.1111/jfr3.12952>, in press.
- Stein, R. M., L. Dueñas-Osorio, and D. Subramanian, 2010: Who evacuates when hurricanes approach? The role of risk, information, and location. *Soc. Sci. Quart.*, **91**, 816–834, <https://doi.org/10.1111/j.1540-6237.2010.00721.x>.
- Stewart, S. R., 2021: Tropical cyclone report: Hurricane Iota (AL312020). NHC Tech. Rep., 44 pp., [https://www.nhc.noaa.gov/data/tcr/AL312020\\_Iota.pdf](https://www.nhc.noaa.gov/data/tcr/AL312020_Iota.pdf).
- Titley, H. A., M. Yamaguchi, and L. Magnusson, 2019: Current and potential use of ensemble forecasts in operational TC forecasting: Results from a global forecaster survey. *Trop. Cyclone Res. Rev.*, **8**, 166–180, <https://doi.org/10.1016/j.tcr.2019.10.005>.
- , R. L. Bowyer, and H. L. Cloke, 2020: A global evaluation of multi-model ensemble tropical cyclone track probability forecasts. *Quart. J. Roy. Meteor. Soc.*, **146**, 531–545, <https://doi.org/10.1002/qj.3712>.
- , 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. *J. Hydrometeor.*, **22**, 1801–1817, <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>.
- UN OCHA, 2020a: Latin America and The Caribbean: 2020 hurricane seasons—Situation Report no. 4. 13 pp., [https://reliefweb.int/sites/reliefweb.int/files/resources/20201120\\_CA%20Eta%20SitRep%204%20ENG.pdf](https://reliefweb.int/sites/reliefweb.int/files/resources/20201120_CA%20Eta%20SitRep%204%20ENG.pdf).
- , 2020b: Central America and Mexico: 2020 Hurricane season—Situation Report no. 5. 16 pp., <https://reliefweb.int/sites/reliefweb.int/files/resources/20201126%20CA%20Eta%20SitRep%205%2028ENG%29.pdf>.
- Van Der Knijff, J. M., J. Younis, and A. P. J. De Roo, 2010: LISFLOOD: A GIS-based distributed model for river basin scale water balance and flood simulation. *Int. J. Geogr. Info. Sci.*, **24**, 189–212, <https://doi.org/10.1080/13658810802549154>.
- Wang, C.-C., B.-X. Lin, C.-T. Chen, and S.-H. Lo, 2015: Quantifying the effects of long-term climate change on tropical cyclone rainfall using a cloud-resolving model: Examples of two landfall typhoons in Taiwan. *J. Climate*, **28**, 66–85, <https://doi.org/10.1175/JCLI-D-14-00044.1>.
- Whitehead, J. C., R. Edwards, M. Van Willigen, J. R. Maiolo, K. Wilson, and K. T. Smith, 2000: Heading for higher ground: Factors affecting real and hypothetical hurricane evacuation behavior. *Global Environ. Change*, **2B**, 133–142, [https://doi.org/10.1016/S1464-2867\(01\)00013-4](https://doi.org/10.1016/S1464-2867(01)00013-4).
- Wright, D., T. R. Knutson, and J. A. Smith, 2015: Regional climate model projections of rainfall from U.S. landfalling tropical cyclones. *Climate Dyn.*, **45**, 3365–3379, <https://doi.org/10.1007/s00382-015-2544-y>.
- Zhang, M., and Z. Meng, 2018: Impact of synoptic-scale factors on rainfall forecast in different stages of a persistent heavy rainfall event in South China. *J. Geophys. Res. Atmos.*, **123**, 3574–3593, <https://doi.org/10.1002/2017JD028155>.
- Zhou, X., and Coauthors, 2022: The development of the NCEP Global Ensemble Forecast System version 12. *Wea. Forecasting*, **37**, 1069–1084, <https://doi.org/10.1175/WAF-D-21-0112.1>.
- Zsoter, E., S. Harrigan, C. Barnard, F. Wetterhall, P. Salamon, and C. Prudhomme, 2019: River discharge and related forecasted data from the global flood awareness system, v2.1. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), accessed 3 February 2022, <https://doi.org/10.24381/cds.ff1aef77>.
- , H. L. Cloke, C. Prudhomme, S. Harrigan, P. de Rosnay, J. Munoz-Sabater, and E. Stephens, 2020: Trends in the GloFAS-ERA5 river discharge reanalysis. ECMWF Tech. Memo. 871, 73 pp., <https://www.ecmwf.int/en/eLibrary/81194-trends-glofas-era5-river-discharge-reanalysis>.