

The Effect of Reference Climatology on Global Flood Forecasting

FEYERA A. HIRPA, PETER SALAMON, LORENZO ALFIERI, AND JUTTA THIELEN-DEL POZO

Institute of Environment and Sustainability, European Commission Joint Research Centre, Ispra, Italy

ERVIN ZSOTER

European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

FLORIAN PAPPENBERGER

European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom, and College of Hydrology and Water Resources, Hohai University, Nanjing, China, and School of Geographical Sciences, University of Bristol, Bristol, United Kingdom

(Manuscript received 27 March 2015, in final form 7 January 2016)

ABSTRACT


The Global Flood Awareness System (GloFAS) is a preoperational suite performing daily streamflow simulations to detect severe floods in large river basins. GloFAS defines the severity of a flood event with respect to thresholds estimated based on model-simulated streamflow climatology. Hence, determining accurate and consistent critical thresholds is important for its skillful flood forecasting. In this work, streamflow climatologies derived from two global meteorological inputs were compared, and their impacts on global flood forecasting were assessed. The first climatology is based on precipitation-corrected reanalysis data (ERA-Interim), which is currently used in the operational GloFAS forecast, while the second is derived from reforecasts that are routinely produced using the latest weather model. The results of the comparison indicate that 1) flood thresholds derived from the two datasets have substantial dissimilarities with varying characteristics across different regions of the globe; 2) the differences in the thresholds have a spatially variable impact on the severity classification of a flood; and 3) ERA-Interim produced lower flood threshold exceedance probabilities (and flood detection rates) than the reforecast for several large rivers at short forecast lead times, where the uncertainty in the meteorological forecast is smaller. Overall, it was found that the use of reforecasts, instead of ERA-Interim, marginally improved the flood detection skill of GloFAS forecasts.

1. Introduction

Flood, which knows no political, economic, or social boundaries, is one of the world's deadliest and costliest natural disasters. Since the beginning of the twentieth century, almost 5000 major hydrological disasters are known to have occurred worldwide, affecting a total of over 3.5 billion people, claiming more than 7 million lives, and causing estimated economic damages of USD \$650

billion (CRED 2014). Flood-related disasters alone are responsible for 51% of the total people affected, 22% of the total deaths, and 24% of the total economic damages caused by all major natural disasters combined.

Wide-ranging global efforts (UNISDR 2015; Ward et al. 2015) have been aimed at reducing flood damages by increasing flood-risk awareness. The Global Flood Partnership (GFP; De Groeve et al. 2015) has recently been launched with the vision of bringing together the scientific community, end users, and decision-makers for the common goal of improving flood-risk management and reducing the impact of floods. The partnership promotes the integration of available flood forecasting, detection, and mapping tools and facilitates the provision of timely flood-related information to relevant end users and disaster-response organizations. These scientific tools play a key role in flood damage reduction.

 Denotes Open Access content.

Corresponding author address: Feyera A. Hirpa, Institute of Environment and Sustainability, European Commission Joint Research Centre, Via Enrico Fermi 2749 TP 122, I-21027 Ispra, Italy.
E-mail: feyera-aga.hirpa@jrc.ec.europa.eu

DOI: 10.1175/JHM-D-15-0044.1

Flood detection and recording systems such as the Dartmouth Flood Observatory (DFO; [Brakenridge 2014](#)) and the Global Disaster Alert and Coordination System (GDACS 2014; [De Groeve and Riva 2009](#)) collect, archive, and share real-time and historical flood information. These comprehensive flood-record databases with common data standards for recording warnings, events, and impacts will ensure that flood-risk research and applications have sufficient data available to study flood disasters. Near-real-time flood monitoring systems using a combination of satellite-based precipitation estimates and hydrologic models (e.g., [Wu et al. 2014](#)) provide real-time flood detection and short-term flood prediction mechanisms. Flood inundation mapping tools (e.g., [Schumann et al. 2014](#)) are useful for delineating vulnerable areas that are under flood risk. Flood forecasting and early warning systems (e.g., [Alfieri et al. 2013](#); [Pappenberger and Brown 2013](#); [Pappenberger et al. 2013](#); [Dale et al. 2014](#)) allow for prediction of a flood disaster beforehand and preparation days or weeks before it strikes ([Pagano et al. 2014](#)).

Flood early warning systems play a major role in saving lives and reducing economic damages ([Jha et al. 2012](#)). Recent studies show that early warning systems bring substantial financial benefits ([Sampson et al. 2014](#); [Pappenberger et al. 2014](#)). There are a number of probabilistic flood forecasting and early warning systems at national and continental scales (see [Cloke and Pappenberger 2009](#); [Pappenberger et al. 2015](#)). Examples of continental-scale hydrologic ensemble prediction systems are the European Flood Awareness System (EFAS; [Thielen et al. 2009](#); [Bartholmes et al. 2009](#)) and the African Flood Forecasting System (AFFS; [Thiemig et al. 2014](#)).

At the global scale, however, there are only a few hydrological and flood-risk assessment modeling (e.g., [Kim et al. 2009](#); [Decharme et al. 2012](#); [Pappenberger et al. 2012](#); [Yamazaki et al. 2013](#); [Winsemius et al. 2013](#)) and operational flood forecasting systems. The Global Flood Awareness System (GloFAS; [Alfieri et al. 2013](#)) has been producing ensemble streamflow forecasts for major global rivers since its launch in July 2011. On a daily basis, GloFAS also produces forecasts for probability of streamflow exceeding critical thresholds corresponding to medium, high, and severe flood levels. The flood levels are expressed in terms of return levels based on streamflow climatology derived from long-term model simulation. Simulated streamflow is used for the derivation of the return levels primarily because of the limited availability of discharge observation in most parts of the world.

Several factors could have an effect on the forecast skill of GloFAS, including errors in the meteorological forecasts, sparse observational network for meteorological and hydrological variables, uncertainty in the

hydrologic model and initial conditions, human-made influences in the river basins, information on the local flood vulnerability, and the uncertainty in reference climatology ([Zsoter et al. 2014](#)) used for threshold calculation. The last factor has been evaluated in the context of flood-level forecasting and is the subject of the current work, while investigating the possible impacts of the remaining sources of uncertainty ([Wood and Lettenmaier 2008](#)) is beyond the scope of this study.

It is important that the characteristics of the reference climatology are fully understood since it dictates the process of determining the occurrence and severity of a flood event. If, for example, the reference threshold is set too low compared to what the operational forecast produces, there will be a high frequency of false alarms produced. Conversely, if the threshold is set too high, the flood detection rate will be too low and thus the forecast will miss large floods. In this work, we evaluate the impact of the meteorological dataset underlying the reference discharge climatology on GloFAS forecast skill using two inputs produced at the European Centre for Medium-Range Weather Forecasts (ECMWF) using different versions of the Integrated Forecast System (IFS). The first one is ERA-Interim ([Dee et al. 2011](#)), which is an atmospheric reanalysis dataset with precipitation bias corrected using Global Precipitation Climatology Project (GPCP) data ([Balsamo et al. 2015](#)). The GPCP-corrected ERA-Interim represents a long-term global atmospheric dataset that is the closest to reality as possible. The second is the ECMWF reforecast dataset, which is continuously produced using the latest version of the operational forecast model. The reforecast datasets are more consistent with operational weather forecasts, which could make them more suitable to reproduce statistical extreme values that are consistent with predicted discharge extremes.

The objective of this study is to assess the effect of the reference climatology on the skill of GloFAS flood occurrence probability. The remainder of the paper is organized as follows. The methods and data, including the hydrologic modeling, reference climatology, and the data used for evaluation of the flood forecast are presented in [section 2](#). The evaluation methods are described in [section 3](#). In [section 4](#), results and a discussion are presented, and conclusions are summarized in [section 5](#).

2. Methodology and data

A schematic of the methodology used in this study is shown in [Fig. 1](#). There are two main outputs from the GloFAS operational system on daily time steps: ensemble streamflow forecasts and flood exceedance probabilities. The exceedance probabilities are estimated based

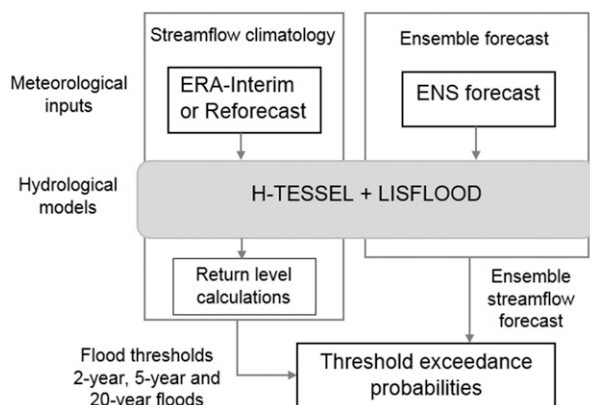


FIG. 1. Schematics of the study design.

on whether or not the forecasted streamflow exceeds critical flood thresholds that are derived from historical streamflow climatology. A common hydrological model setup was used to produce the ensemble streamflow forecast and the streamflow climatology. In this study, we compare two streamflow climatologies derived from ERA-Interim and reforecast meteorological forcings. The following sections describe the details of the global streamflow forecasting system, the procedure to estimate the streamflow climatology, and the data used (see also Table 1 for a summary).

a. Global hydrological forecasting

As described in Alfieri et al. (2013), GloFAS employs a cascade of a hydrological and routing model. The first is a land surface scheme of the ECMWF's IFS used for calculating the surface and subsurface runoff, while the second is used for routing the simulated runoff through the river network.

1) H-TESEL FOR HYDROLOGICAL MODELING

The precipitation (and snowmelt) reaching the land surface is partitioned between surface evaporation, vegetation root extraction, surface and subsurface runoff, and free drainage to the groundwater. In GloFAS,

this grid-based partitioning process is modeled using the land surface scheme of the ECMWF's IFS, referred to as Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land (H-TESEL; Balsamo et al. 2009; Dutra et al. 2010). H-TESEL performs the energy and water balance calculations at each pixel of gridded land and water surfaces. Each land surface grid is represented with six tiles depending on the type and the degree of land cover (bare land, low or high vegetation, shaded or exposed snow, or intercepted water), and each water surface grid is denoted with two tiles for open or frozen water. The physical characteristics of each grid box are parameterized using global datasets. The land surface vegetation characteristics, such as vegetation type and area fraction, are derived from the climate database based on Global Land Cover Characterization (GLCC) data (Loveland et al. 2000). Spatially variable soil texture classes are used for determining the soil hydraulic properties and the subsequently calculated surface runoff varies based on the soil texture. The global soil texture classes are obtained from the Food and Agricultural Organization of the United Nations (FAO 2003). Subsurface water flows are determined by Darcy's law, and the vertical water fluxes are modeled using a representation of four soil layers covered with one layer of snow (Dutra et al. 2010). The surface and subsurface runoff at variable spatial resolutions (i.e., 32–79 km depending on the IFS model) are the final outputs from the H-TESEL used for streamflow modeling.

2) LISFLOOD FOR FLOW ROUTING

A simplified version of LISFLOOD (Van Der Knijff et al. 2010; Burek et al. 2013) is used for routing the surface and subsurface runoff produced by H-TESEL [see section 2a(1)]. LISFLOOD is a two-soil-layer distributed rainfall–runoff model used for flood forecasting and water resources modeling. Besides its standalone capability to model the complete surface and subsurface hydrological processes as used in EFAS (e.g., Thielen

TABLE 1. Summary of the hydrologic models and meteorological inputs used for the calculation of the flood thresholds and operational forecasts. A common hydrological model setup was used for all simulations.

	Threshold calculation		
	ERA-Interim	Reforecast	Operational forecast
Meteorological forcing	ERA-Interim (deterministic) Resolution: ~79 km, daily	Reforecast (deterministic) Resolution: ~32 km, daily	ENS with 51 members Resolution: ~32 km for 1–10-day lead, daily
Hydrologic model used	H-TESEL coupled with LISFLOOD Estimated return levels to get flood thresholds at 0.1° grids	H-TESEL coupled with LISFLOOD Estimated return levels to get flood thresholds at 0.1° grids	H-TESEL coupled with LISFLOOD Output: Streamflow at 0.1° grids and daily time steps

et al. 2009; Bartholmes et al. 2009) and AFFS (Thiemig et al. 2014), LISFLOOD can also provide an offline channel flow routing module that can be coupled with other land surface models. In the GloFAS setup, a one-dimensional LISFLOOD channel routing module is used for estimating the groundwater storages and for routing the surface and subsurface runoffs along river channels. The routing scheme uses a four-point implicit finite difference solution of the kinematic wave equation (Chow et al. 1988). It assumes rectangular-shaped channel cross sections and calculates channel flow velocity at each of the $0.1^\circ \times 0.1^\circ$ grid points of the global setup using Manning's equation. The variables needed to apply Manning's equation (such as river network and river width) are obtained from global databases. In this work, the river network map, the global flow direction map, the upstream area, and the flow length at $0.1^\circ \times 0.1^\circ$ resolution were obtained from the global river network database (Wu et al. 2012). The Manning's roughness coefficients were approximated to values ranging from a minimum of 0.025 for large lowland rivers to a maximum of 0.07 for small mountainous streams (McCuen 2004). The channel gradient between adjacent grids on the river network was estimated using the flow length and the change in elevation between the respective grids. The river widths were obtained from the Global Width Database for Large Rivers (GWD-LR; Yamazaki et al. 2014). The GWD-LR was calculated from a water body mask obtained mainly from Shuttle Radar Topography Mission (SRTM) water body data and flow direction datasets derived from Hydrological Data and Maps Based on Shuttle Elevation Derivatives at Multiple Scales (HydroSHEDS; Yamazaki et al. 2014; NASA/NGA 2003; Lehner et al. 2008). The bankfull water depth was empirically estimated from long-term average discharge and other parameters (described above) by applying Manning's equation. For simplicity, for all simulations in this study, the large lakes and reservoir modules are not included in the hydrologic model. Moreover, other parameters of the routing model such as time constants for water in the upper and lower zones and maximum rate of percolation from upper to lower zone were set to constant values, though no specific calibration was performed to fine-tune output discharge values. While this may have an effect on the overall (absolute) forecast skill, it is not expected to affect the relative skill scores since the same model setup is used for the streamflow forecast and the climatology simulations.

3) METEOROLOGICAL FORCING FOR OPERATIONAL FLOOD FORECASTS

The meteorological variables used as input to the hydrologic models were obtained from the Ensemble Prediction System (ENS) of the ECMWF. The ENS,

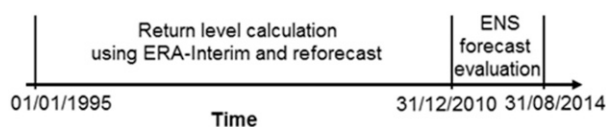


FIG. 2. Time window considered for the calculation of flood thresholds and for the forecast evaluation.

produced with the most recent IFS update (ECMWF 2015) at the time, is a 51-member ensemble of global forecasts generated twice per day (at 0000 and 1200 UTC) at a horizontal resolution of ~ 32 km for 1–10-day lead times and at a coarser resolution (~ 65 km) for 11–15-day lead times. In this work, we used the first 10-day lead forecasts initialized once per day at 0000 UTC to produce a daily ensemble streamflow forecast for a total period of 44 months (from January 2011 to August 2014; see Fig. 2).

4) INITIAL CONDITIONS FOR THE ROUTING MODEL

The initial states of the LISFLOOD routing model for the first day (i.e., 1 January 2011) were produced using a long-term run (starting on 1 January 1980) using ERA-Interim meteorological forcing. Then for the subsequent days, model initial conditions were extracted from the simulations with the ENS control run as an input. This means that all members of the 51-ensemble model simulation started from common initial conditions on every first day of the 10-day forecast window, and thus, the differences that arise in the 1–10-day lead forecast among the streamflow ensemble members are due to the spread in the ensemble meteorological forecasts. This approach is analogous to the operational GloFAS (www.globalfloods.eu), while the only difference is that in the operational mode a higher-resolution deterministic forecast (HRES) is used for the generation of the initial conditions of the routing model, instead of the ENS control run used in this study.

b. Streamflow climatology

To determine the occurrence and severity of a flood event, the common practice is to express the forecasted hydrographs in terms of whether they exceed certain critical thresholds (e.g., Reed et al. 2007; Thielen et al. 2009). In GloFAS, thresholds corresponding to medium, high, and severe floods are expressed as 2-, 5-, and 20-yr return levels, respectively. The impact of the different-magnitude floods is not currently accounted for in the GloFAS as, for example, a 5-yr flood can cause a catastrophic effect in vulnerable areas with no flood protection levels, while the same magnitude may not have considerable impact in other areas where flood vulnerability or exposure is low.

The return levels for each $0.1^\circ \times 0.1^\circ$ model grid are estimated from the streamflow climatology based on model simulations using long-term meteorological forcing. The return levels were computed by fitting Gumbel extreme value distribution (Gumbel 1941) to the annual maxima of the daily flows over the period 1995–2010 (see Fig. 2). The two-parameter Gumbel distribution is shown to produce a comparable (in some cases better) estimate of extreme discharge compared to a three-parameter Generalized Extreme Value (GEV) distribution (Dankers and Feyen 2008). The cumulative distribution function (CDF) of the Gumbel distribution is expressed as

$$\text{CDF} = e^{-(e^{-Z})}, \quad (1)$$

where $Z = (Q - \mu)/\beta$ is a dimensionless constant, Q is discharge ($\text{m}^3 \text{s}^{-1}$), μ is a location parameter equal to a discharge value corresponding to the peak of the probability distribution function (PDF), and β is a scale parameter that determines the shape of the PDF.

The procedure described above was used to estimate flood warning thresholds for the two sets of streamflow climatologies derived from ERA-Interim and reforecast. One may note that the GPCP-corrected ERA-Interim represents an attempt to produce a long-term global atmospheric dataset that is as close to reality as possible. The atmospheric reforecast represents a long-term dataset that is more consistent with the operational forecasts, meaning that it is more suitable to reproduce statistical extremes consistent with discharge extremes produced by the operational forecast.

1) THRESHOLDS FROM ERA-INTERIM

The first sets of flood thresholds were estimated using ERA-Interim near-surface meteorological fields as input to the hydrologic models. The initial conditions for the routing model were extracted from a long-term run (more than 15 years starting in 1980) using the ERA-Interim forcing. The ERA-Interim is a global atmospheric reanalysis dataset produced based on the 2006 version of the ECMWF IFS (Berrisford et al. 2009; Dee et al. 2011). It was produced at a reduced Gaussian grid of approximately 79-km uniform horizontal resolution. The ERA-Interim precipitation was bias corrected to match the monthly averages from GPCP, which is a globally merged precipitation product from satellite and gauge observations (Huffman et al. 2009; Balsamo et al. 2015).

2) THRESHOLDS FROM REFORECAST

The second set of thresholds was calculated using the unperturbed (control) run of the reforecast dataset,

which is generated based on a retrospective run of the most recent version of the IFS model at the run time (see ECMWF 2015). At the ECMWF, the IFS model is run once per week to generate ensemble meteorological reforecasts for the same weekday of the five surrounding weeks in the past 20 years. The purpose of the model rerun is to make the datasets consistent with the updated operational weather forecasts. However, it should be noted that if there is a frequent model upgrade, there could be a possibility of having nonhomogeneous datasets produced by different model versions. Since the reforecast is run only once per week, it was necessary to reconstruct the forecasts from multiple lead times to obtain a continuous time series needed to run the LISFLOOD model and then estimate flood threshold maps. The effect of the mixed forecast lead times (1–7-day lead) on the skill of simulated discharge was found to be not significant when evaluated against in situ measurements for several global discharge measurement stations. Alfieri et al. (2014) showed that such an approach based on a reforecast dataset leads to estimates of flood thresholds consistent with operational forecasts.

Compared to the ERA-Interim, the reforecast has a higher spatial resolution (~ 32 km) and is produced with the latest version of the IFS model (ECMWF 2015), which makes it more similar to the operational forecast. An additional difference is that there is no bias correction applied to the reforecast precipitation, while the ERA-Interim precipitation was bias corrected using GPCP. The initial conditions for the routing model were obtained from a long-term run using ERA-Interim meteorological inputs.

c. Data

1) DISCHARGE MEASUREMENTS

Daily discharge measurements for several stations from different continents were used for the assessment of the extreme value distribution used for estimating the flood return levels and the skill of the exceedance probability forecast. A total of 375 stations with continuous daily records during the 1995–2010 period were obtained from the Global Runoff Data Centre (GRDC). The global distribution of the stations is, however, not homogenous, with no continuous daily discharge data for South America and Asia, only two for Africa, four for Australia, 10 representing Europe, and the remaining coming from North America.

2) GLOBAL FLOOD RECORDS

Flood events from the DFO (Brakenridge 2014) were also used for evaluating the skill of the GloFAS forecast. The DFO database provides a list of major global flood

events with start and end dates, estimates of the affected areas, approximate centroids of the affected areas, and related information extracted from various news reports. A total of 102 global events were selected from the list based on three criteria: 1) the start and end dates of the flood event lie within the study period (between January 2011 and August 2014), 2) the flood event is also recorded in the Emergency Events Database (EM-DAT; [CRED 2014](#)), and 3) flash flood events are excluded from the list. After the selection was made, for each event a flood point (a single model pixel) on a river channel was manually determined based on the centroid of the affected area and other geographical information extracted from news reports included in the DFO database.

The ensemble streamflow forecast from the GloFAS is extracted at all model grids (0.1°) containing the 102 flood points worldwide for a period between the start and end of the flooding. Then, using the flood thresholds computed from the ERA-Interim and reforecasts, exceedance probabilities are calculated at each location for all days during the duration of the flood, using the two sets of three maps with constant levels associated with the 2-, 5-, and 20-yr flood. The maximum exceedance probability during the duration of the flood is used to assess the flood detection skill.

3. Evaluation

The difference between the ERA-Interim and the reforecast streamflow climatologies and their impacts on the operational flood forecasting were evaluated using various methods. First the Gumbel extreme value distributions estimated for each of the ERA-Interim and reforecast were compared with those of in situ discharge. Then, 2-, 5-, and 20-yr flood levels computed at each model grid ($0.1^\circ \times 0.1^\circ$) globally using ERA-Interim and reforecast were directly compared. The percentage difference between the thresholds for each flood level was calculated as

$$\text{difference in threshold} = \left(\frac{\text{Thr}_{\text{reforecast}} - \text{Thr}_{\text{ERA-Interim}}}{\text{Thr}_{\text{ERA-Interim}}} \right) (100\%), \quad (2)$$

where $\text{Thr}_{\text{ERA-Interim}}$ and $\text{Thr}_{\text{reforecast}}$ are thresholds derived from ERA-Interim and reforecast, respectively.

Additionally, the impacts of the reference climatology on flood exceedance probability of the GloFAS forecast for lead times ranging from 1 to 10 days were assessed. This was done using the 51-member ensemble forecast generated for lead times up to 10 days over the period from 1 January 2011 to 31 August 2014 (a total of 1339 days). The total number of days for which the

median of the ensemble streamflow forecast exceeded the thresholds from ERA-Interim and reforecasts were compared. The percentage difference was computed as

$$\text{difference in flood days} = \left(\frac{N_{\text{reforecast}} - N_{\text{ERA-Interim}}}{N_{\text{ERA-Interim}}} \right) (100\%), \quad (3)$$

where $N_{\text{ERA-Interim}}$ and $N_{\text{reforecast}}$ are the number of flood days when ERA-Interim and reforecasts were used for deriving the thresholds, respectively.

Additionally, an evaluation based on observed discharge over three years (2011–13) was done using the Brier score (BS) for each station:

$$\text{BS} = \frac{1}{T} \sum_{t=1}^T (P_F - P_O)^2, \quad (4)$$

where P_F is the probability of the forecast exceeding a reference threshold (e.g., ERA-Interim or reforecast), P_O is 0 or 1 depending on whether or not the observed discharge exceeds the flood threshold (observed), and T is the total number of days. The change in forecast skill when the reforecast thresholds are used instead of ERA-Interim is expressed using the Brier skill score (BSS):

$$\text{BSS} = 1 - \frac{\text{BS}_{\text{reforecast}}}{\text{BS}_{\text{ERA-Interim}}}. \quad (5)$$

Furthermore, flood data obtained from global flood records were used for assessing the effect on the probability of the streamflow forecast exceeding the warning thresholds. The flood exceedance probabilities were calculated as the proportion of the ensemble streamflow forecast exceeding the flood thresholds. Differences in the exceedance probabilities based on ERA-Interim and reforecast thresholds were estimated using Eq. (6):

$$\begin{aligned} \text{difference in exceedance probability} &= P(Q_{\text{Flood}} > \text{Thr}_{\text{ERA-Interim}}) \\ &\quad - P(Q_{\text{Flood}} > \text{Thr}_{\text{reforecast}}), \end{aligned} \quad (6)$$

where $P(Q_{\text{Flood}} > \text{Thr}_{\text{ERA-Interim}})$ is the probability of discharge exceeding the ERA-Interim threshold and $P(Q_{\text{Flood}} > \text{Thr}_{\text{reforecast}})$ is the probability of discharge exceeding the reforecast threshold.

4. Results

a. Comparison of flood thresholds

The Gumbel extreme value distribution was fit for each of the ERA-Interim, reforecast, and in situ

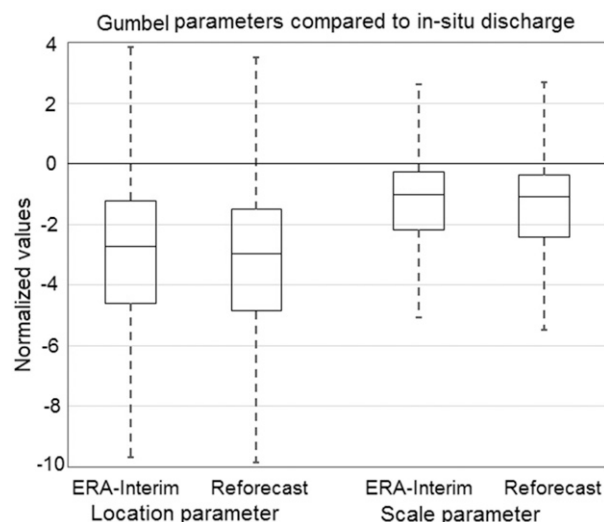


FIG. 3. Difference between parameters of the Gumbel distribution derived from simulated streamflow climatology and in situ discharge for 375 global stations. Values shown are normalized using the mean observed discharge of each station.

discharge climatologies. Figure 3 shows differences between the Gumbel distribution parameters estimated for the simulated and observed discharge for the 375 stations. Note that the time period and length of data used for calculation of the return levels are identical for the ERA-Interim, reforecast, and observed discharge (as shown in Fig. 2). Parameters of the extreme value distribution based on both ERA-Interim and the reforecasts were on average lower than those based on in situ discharge. The difference in location parameter (normalized with the mean observed discharge of each location) has a median of -2.72 (10th percentile of -7.45 and 90th percentile of 1.42) for ERA-Interim and a slightly lower value of -2.95 (10th percentile of -7.60 and 90th percentile of -0.30) for reforecast climatology. Similarly, the medians of the scale parameter are found to be -1.01 (10th percentile of -5.27 and 90th percentile of 0.94) and -1.07 (10th percentile of -5.51 and 90th percentile of 0.12) for ERA-Interim and reforecast, respectively. The null hypothesis test on the two parameters sets showed significant similarity between ERA-Interim and the reforecasts ($p = 0.109$ for the location parameter and $p = 0.247$ for the scale parameter). The negative values of the parameters indicate that the flood return levels estimated using ERA-Interim and the reforecast have lower magnitudes compared to the return levels estimated from observed discharge. Such behavior can be partly explained by the smoothing effect on discharges produced by the use of meteorological input and hydrological models with space–time resolution comparatively coarser than the real-world hydrometeorological

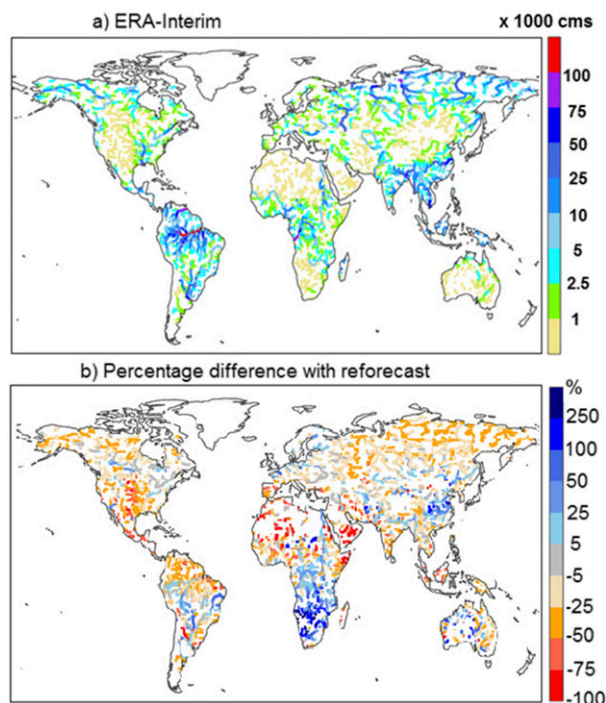


FIG. 4. (a) ERA-Interim flood thresholds ($\times 10^3 \text{ m}^3 \text{ s}^{-1}$) and (b) percentage difference with reforecasts [Eq. (2)] shown for 5-yr return flood. Return levels were calculated based on 16-yr (1995–2010) model run for both ERA-Interim and reforecasts.

processes, particularly in extreme event situations. Note that there were no discharge data for some regions (e.g., South America, Asia, and most parts of Africa), and hence, the results presented based on the discharge datasets are not a complete representation of the globe.

The direct comparison between return levels estimated based on the two sets of climatologies from ERA-Interim and the reforecast showed substantial differences with varying characteristics across different regions of the globe. The 5-yr return flood from ERA-Interim and percentage difference with reforecast are shown in Fig. 4. The 2- and 20-yr amounts are not shown in the figure as they have similar patterns and lead to the same conclusions. Return levels based on reforecasts have a higher magnitude compared to those based on ERA-Interim for most African basins, including the Nile, Zambezi, Okavango, Congo, Orange, and Limpopo Rivers; for some rivers in Asia, such as the Ganges, Brahmaputra, Yangtze, and downstream reaches of the Yellow and Indus Rivers; and for some in South America, including the Paraná, São Francisco, and Madeira Rivers. ERA-Interim produces higher thresholds for pan-Arctic rivers, the Mississippi and Niger Rivers, and most of the Amazon tributaries. While the difference in most European rivers was within $\pm 25\%$ of

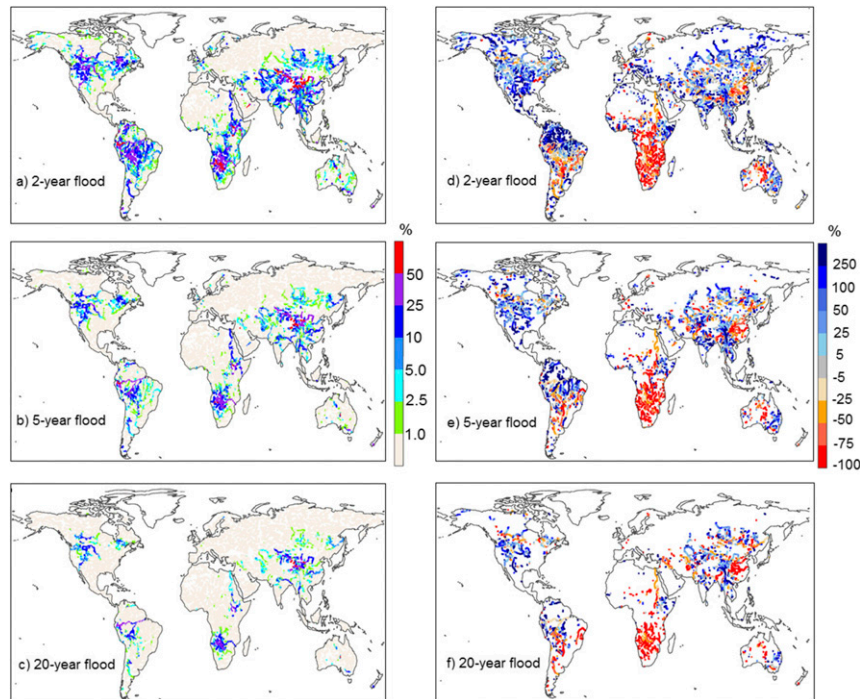


FIG. 5. For return levels of (a),(d) 2; (b),(e) 5; and (c),(f) 20 years, the (left) percentage of days derived from ERA-Interim for which the 1-day lead ensemble median exceeds the flood threshold and (right) difference (%) between the reforecast-based exceeded days and ERA-Interim [Eq. (3)]. Negative values indicate that there are more flood warnings based on ERA-Interim thresholds compared to those of reforecast and vice versa.

the ERA-Interim, the relative difference is larger in central and southern Africa with some values going above 100% (e.g., Okavango basin).

Note that the daily annual maximum flows are the basis for calculation of the return levels, and the differences in the climatologies reflect the disagreements between the annual peak flows of the two datasets. Two factors may have contributed to the disagreements. First, the IFS model has undergone several improvement cycles since its 2006 release, which was used for creating ERA-Interim datasets. The model changes include, among others, improved representation of clouds and precipitation (e.g., Forbes and Tompkins 2011), change in vertical and horizontal resolutions (ECMWF 2015), and the use of an ensemble of data assimilation for initial perturbation of the ensemble forecast (Forbes et al. 2015). Studies (Haiden et al. 2014; Forbes et al. 2015) found that the successive IFS model enhancements contributed to a better forecast skill of heavy precipitation events relative to the ERA-Interim (before precipitation correction).

Second, the bias correction applied to the ERA-Interim precipitation using the GPCP monthly climatology could contribute to the disagreements with the reforecast climatology. This is particularly important in the tropics, where ERA-Interim precipitation was found to be larger

(Betts et al. 2009; Balsamo et al. 2015). The tropics and Southern Hemisphere were identified as regions where the bias correction produces a substantial improvement in the ERA-Interim precipitation (Balsamo et al. 2015). This may have reduced the magnitudes of peak flows in areas where ERA-Interim climatology is significantly lower than reforecast (e.g., southern Africa). In some other regions (e.g., the United States; see Balsamo et al. 2010) there was a smaller bias between the ERA-Interim (before correction) and the GPCP precipitation, which may result, at global scale, in a spatially heterogeneous precipitation bias correction applied to the ERA-Interim.

b. Effect on operational flood forecasts

Figures 5a–c show the percentage of days (out of 1339) for which the 1-day lead-time ensemble median exceeded the ERA-Interim flood threshold for different flood magnitudes. Note that the number of flood days presented in the figure should not be confused with the number of flood events, since a flood event can last for a duration of several days to weeks.

The percentage of days when the median of operational forecasts exceeds the 2-yr ERA-Interim flood level (Fig. 5a) is the highest for the Okavango and Zambezi Rivers in Africa, the Yellow River in Asia, and

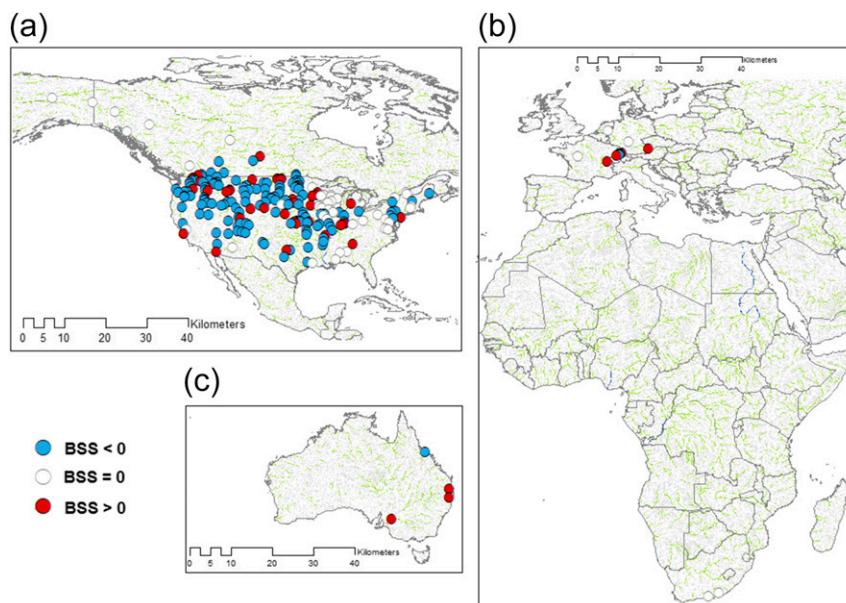


FIG. 6. BSS of the 1-day lead ensemble flood forecasts resulting from reforecast thresholds relative to ERA-Interim thresholds: (a) North America, (b) Europe and Africa, and (c) Australia. Results refer to the 2-yr flood threshold.

the main channels of the Amazon and Nile, for some river channels going above 500 days. The upper Mississippi River, the Colorado, the Amazon tributaries, and most Asian rivers also exhibited high number of flood days. A similar pattern was observed for high-magnitude floods (5-yr flood; Fig. 5b). A surprisingly high percentage was also forecasted to exceed the 20-yr flood in some regions (Fig. 5c). The reason for the high percentage needs to be investigated in future research.

Results were then compared with those from reforecast-based thresholds. As shown in Figs. 5d–f, the total number of flood days are significantly lower (red colors) for most African rivers when reforecast-based thresholds are used as a reference compared to ERA-Interim. This is also the case for the main reaches of the Ganges and Brahmaputra, the Yangtze, and lower parts of the Yellow River. This indicates that in those regions there is a higher tendency of detecting floods, and possibly producing false alarms, with the use of ERA-Interim reference climatology. Conversely, for the majority of basins in North America and Russia (blue colors; Figs. 5d–f), results show that using reforecast climatology likely produces higher flood detection rates and possibly more false alarms. These gaps in percentage of flood days between the datasets are direct reflections of the relative differences between the thresholds (see Fig. 4 and section 4a).

Figure 6 presents the overall BSS [see Eq. (5)] of the 2-yr flood discharge of reforecasts with reference to

ERA-Interim. The skill score, calculated using in situ discharges, denotes the added forecast skill when the reforecast thresholds are used instead of those from ERA-Interim. The results show that the forecast skill depreciates ($BSS < 0$) for the majority of the stations in North America (59%) when the reforecast is used instead of ERA-Interim, while the skill improves ($BSS > 0$) for 21% of the stations. The decline in the BSS for the majority of locations in the region can be attributed to the fact that the reforecast thresholds are generally lower than those of ERA-Interim in North America (see Fig. 4), which leads to a tendency of higher false alarm rates produced when the reforecast thresholds are used (explained next). In the regions outside of North America, however, the threshold exceedance skill improves for 7 out of the 16 (44%) stations.

When only observed severe flood events (observed discharge exceeding the 20-yr flood thresholds estimated using in situ data) are considered, the Brier score for 1-day lead forecast of 2-yr return flood improves for 37% (26 out of 70) of the global stations when reforecast thresholds are used instead of those from ERA-Interim. There is no change observed for the remaining 63%. The fact that there was no skill depreciation (no negative BSS) when only the observed flood events were considered indicates that the number of false alarms is the cause for the relatively lower skill score for reforecast when calculated over the whole time series (including no floods). A similar pattern was observed for longer lead

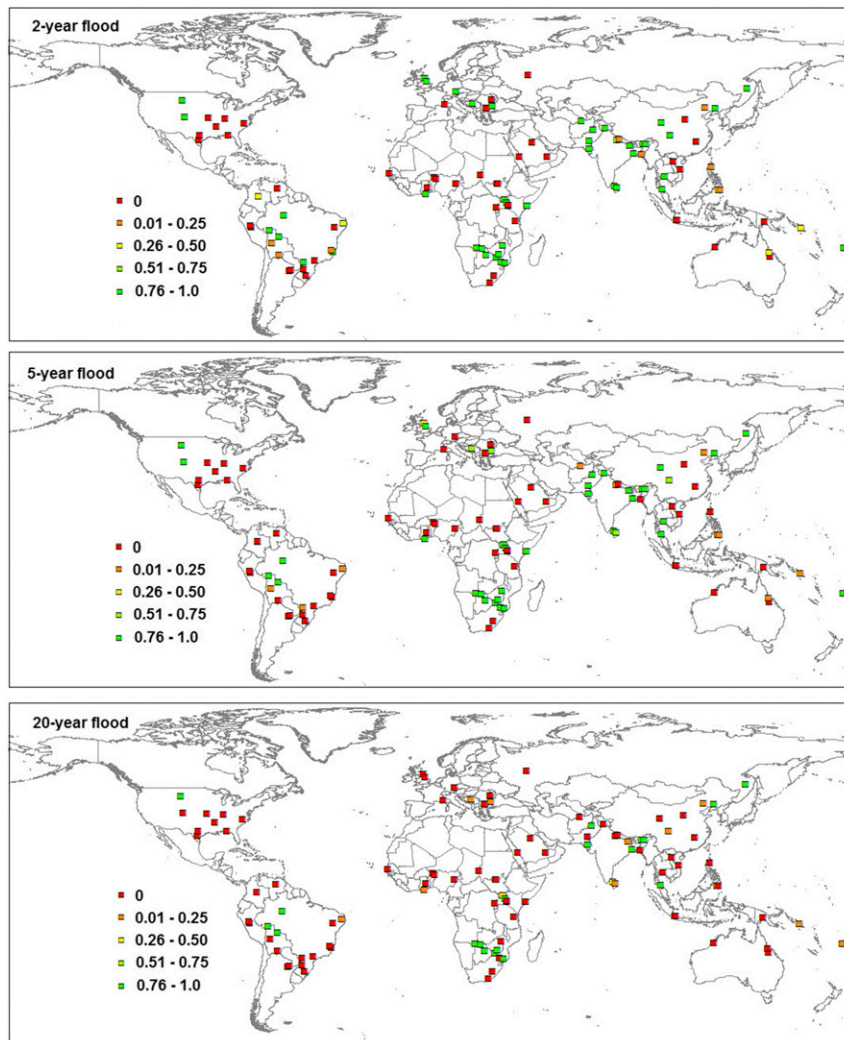


FIG. 7. Probability of forecasted flood threshold exceedances: (top) 2-, (middle) 5-, and (bottom) 20-yr return levels calculated from ERA-Interim for 1-day lead time.

times and higher-magnitude floods. It should be noted, however, that the lack of adequate discharge data leaves the vast majority of rivers out of this analysis. The BSS was calculated for locations that have complete discharge data during 1995–2010 (for the estimation of the return levels) and at least 9 months of discharge data for the forecast evaluation during 2011–14. This limits the number of stations mostly to North America (93%) and only a handful coming from the other regions, with none from South America and Asia.

c. Comparison with flood-record data

1) ERA-INTERIM

Figure 7 shows a global map of forecast exceedance probability for the 102 global flood events for 2, 5, and 20

years from ERA-Interim. The 1-day lead-time forecasts show that 44% of the events were forecasted to exceed the 2-yr flood level with a probability larger than 0.5 [$P(\text{flood} \geq 0.5)$], 51% with a probability more than 0.25 [$P(\text{flood} \geq 0.25)$], and 67% of the recorded floods were forecasted by any ensemble member [$P(\text{flood} > 0)$] when ERA-Interim thresholds were used. Here, we considered only the recorded flood events for evaluation of the streamflow forecast, and hence, we did not quantify the contribution of possible false alarms.

The variation of the 2-yr flood exceedance probability with forecast lead time is shown in Fig. 8. The results for different ensemble percentiles indicate that the flood detection rate decreases with lead time. For example, the correct flood forecast rate of the ensemble median drops from 44% of the flood events at 1-day lead to 26%

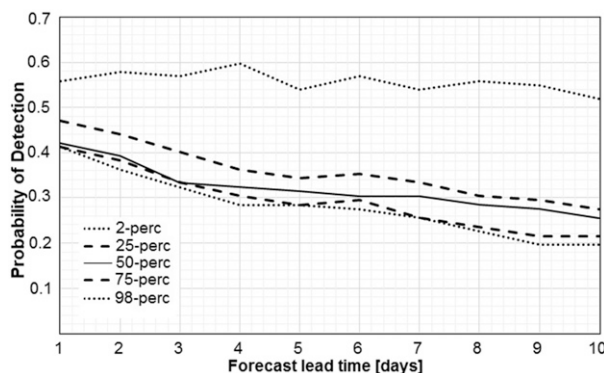


FIG. 8. Probability of detection (ERA-Interim based) with forecast lead time. The y axis shows the proportion of flood events (out of the 102) for which the forecasted flood exceeds the 2-yr threshold.

at 10-day lead forecast. Similarly, the detection rate decreases for other ensemble percentiles shown (98th, 75th, 25th, and 2nd percentiles) with longer forecast lead times. The trend of relatively high detection rate at short lead forecast can be attributed to the lower uncertainty in meteorological forecast compared to long lead times.

There are several possible factors causing the operational ENS forecast to miss some of the floods in general. The first reason could be the uncertainty in the hydro-meteorological modeling. Under the test setup used in this work, the hydrological model parameters are not calibrated and the lake and reservoir modules are not included, both of which may affect the flood detection probability. Besides, the meteorological forecast also has its own uncertainty that could affect the detection. Further, no data assimilation system is in place to update initial discharge conditions at the start of the forecast run. An accurate representation of the initial conditions was shown to be a key component in forecasting high flows in large river basins due to persistence (Alfieri et al. 2013; Yossef et al. 2013), where a considerable component of the runoff contributing to the flood hydrograph is already on the land surface at the start of the forecast run and is simply conveyed downstream through the river network with a travel time proportional to the basin time of concentration.

Second, the spatial location of the flood event might be inaccurately matched with the flood point on the river network in the forecasting model (Revilla-Romero et al. 2015b). The spatial extent of the area affected by the floods (which is not necessarily flooded area) was estimated by the DFO based on information acquired from news sources (Brakenridge 2014). In this work, the centroid of the large affected area was approximated as a flooding point on a nearest major flood network. An alternative approach to selecting a single flood point on

the river channel for the assessment of the ENS flood detection rate would be to consider the entire upstream area from the flood point and then take any exceedance forecasted at any location within the boundary of the catchment (e.g., Wu et al. 2014). This approach considers a much larger spatial domain in contrast to a single model pixel used in this study, but whether it produces an improved detection rate is not clear.

The third reason could be inaccurate estimation of the reference threshold, which is the subject of this work. There is a tendency of missing floods in regions where ERA-Interim thresholds are higher than the ENS forecast, even if high flow magnitudes are forecasted. In such cases, without changing the forecast, it is possible to improve the exceedance probability forecast skill with only using consistent reference climatology with the forecast system. This was tested using reforecast thresholds and assessing the changes in flood detection rates as presented next.

2) ERA-INTERIM VERSUS REFORECASTS

Forecast exceedance probabilities using reforecast thresholds are compared with those of ERA-Interim-based thresholds to evaluate the differences between the two approaches. Figure 9 shows a global map of the difference in exceedance probabilities between ERA-Interim and reforecast thresholds [see Eq. (6)] for 1-day forecast lead time. In 23% of the flood events the probability of peak flow exceeding the 2-yr flood level increased when reforecast thresholds were used instead of ERA-Interim, while in 5% of the cases the probability decreased and there was no change for 72% of the events. Results for the shorter lead times showed that there is an overall tendency of improvement in detection rate (in terms of probability of exceedance) when thresholds derived from reforecast were used instead of ERA-Interim. Similar results were found for 5- and 20-yr floods that the exceedance probability increased for the 1-day lead forecast when the reforecast was used.

The distribution of probabilities for different flood magnitudes and forecast lead times (Fig. 10) indicates that the difference in exceedance probabilities between ERA-Interim and reforecast thresholds narrows for longer forecast lead times (mostly within ± 0.3). It was found that over 70% of the flood events have 5-yr exceedance probability difference between ERA-Interim and reforecast thresholds within ± 0.10 for 5-day lead forecast (Fig. 10b), and it goes close to 80% for 10-day lead (Fig. 10c).

Based on the overall results, a case can be made that, in the current setup of GloFAS, a better flood detection could be achieved when using reforecasts instead of the ERA-Interim to derive threshold levels. This is because of two major reasons: 1) it would reduce the false alarm

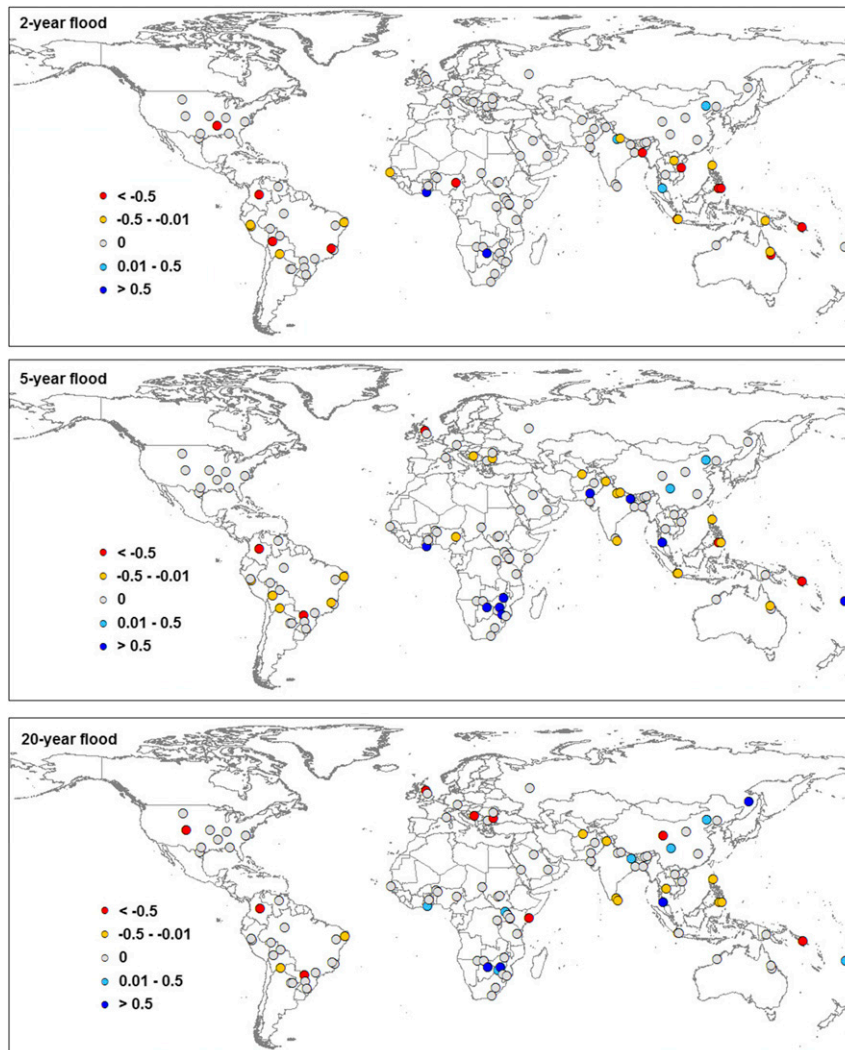


FIG. 9. Difference in threshold exceedance between ERA-Interim and reforecast for 1-day lead ENS (51 member) forecasts [Eq. (6)]. Positive values indicate that the use of ERA-Interim as a reference threshold produces higher exceedance probability compared to reforecast and vice versa. Shown are (top) 2-, (middle) 5-, and (bottom) 20-yr floods.

rate due to low ERA-Interim thresholds in several large rivers (e.g., most rivers in Africa and others shown in Fig. 4) and 2) it would improve the skill of the exceedance probability forecast for short lead-time forecasts for which the meteorological uncertainty is relatively lower. It should also be noted that the reforecasts are more consistent with the operational forecast in terms of model version and resolution. While our study showed that streamflow climatology has an important impact on the flood forecasting skill, the use and effects of other global reanalysis datasets (Kalnay et al. 1996; Rienecker et al. 2011; Kobayashi et al. 2015) for a similar purpose or in other global flood forecasting systems needs to be investigated.

5. Conclusions

In this work, we assessed the effect of the streamflow climatology on GloFAS flood severity forecast by comparing two long-term datasets: ERA-Interim and reforecasts. The evaluation was performed through direct intercomparison of the flood thresholds derived from the two datasets, using the in situ discharge and flood data from a global flood archive. The results can be summarized as follows.

- 1) The return levels based on ERA-Interim have lower magnitude compared to those based on the reforecast for most African basins, for some large rivers in Asia, and in South America, but they have higher

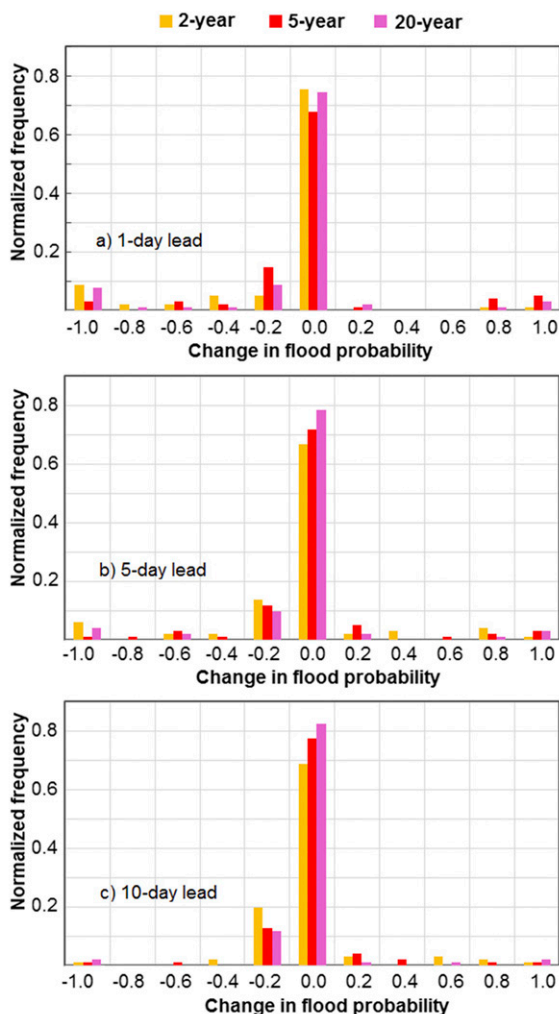


FIG. 10. Distribution of the differences in probability of flood forecast exceeding the 2-, 5-, and 20-yr thresholds between ERA-Interim and the reforecasts (see Fig. 9): (a) 1-, (b) 5-, and (c) 10-day lead times.

magnitude for pan-Arctic rivers, which could be due to the successive IFS model improvements and the precipitation bias correction applied to the ERA-Interim datasets. The relative differences were found to be minor in most European rivers while largest disagreements were observed in central and southern Africa.

- 2) The thresholds have a direct effect on the exceedance probabilities of the operational forecast. Comparison for flood events obtained from global flood-record databases showed that differences between the ERA-Interim and reforecast thresholds have an impact on the flood detection rates for several rivers (ranging from 20% to 40% of the events depending on lead time). Overall, ERA-Interim produced lower

flood threshold exceedance probabilities than the reforecast for short forecast lead times (up to 5-day lead), while the difference narrows for longer lead times (beyond 5-day lead).

- 3) Comparison with in situ discharge showed a regionally variable result. ERA-Interim thresholds produced better streamflow forecast skill for the majority of stations in North America compared to reforecast thresholds, while in other regions it is found that reforecast produced better forecast skills. Flood detection skill considerably improved with the use of reforecast thresholds when flood events were considered.
- 4) The overall results confirmed the importance of defining accurate reference thresholds for exceedance probability forecast and showed that using reforecast-based streamflow climatology produces higher flood detection rates in GloFAS especially at shorter lead times.

While streamflow climatology is shown to be important, additional steps are needed for improving the overall skill of the GloFAS forecast. Works foreseen are calibrating the LISFLOOD parameters and inclusion of the lake and reservoir module into the model. The model calibration using in situ discharge and satellite-derived surface water information [following Revilla-Romero et al. (2015a)] and the impact of lakes and reservoirs on the forecast skill are currently being investigated in separate works. Satellite-based discharge measurements (Brakenridge et al. 2012; Alsdorf et al. 2007; Durand et al. 2010) could be used to fill the gap in discharge observations (e.g., Hirpa et al. 2013) and may provide an opportunity for the calibration and evaluation of the forecast and will be considered in the future.

Acknowledgments. We thank the Dartmouth Food Observatory for providing the list of global floods, as well as the Global Runoff Data Centre, Koblenz, Germany, for providing the discharge data. The GloFAS is developed and maintained jointly by the European Commission Joint Research Centre (JRC) and the European Centre for Medium-Range Weather Forecasts (ECMWF). The paper has greatly benefited from useful comments provided by the editor and two insightful reviewers.

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, doi:10.5194/hess-17-1161-2013.

- , F. Pappenberger, and F. Wetterhall, 2014: The extreme runoff index for flood early warning in Europe. *Nat. Hazards Earth Syst. Sci.*, **14**, 1505–1515, doi:[10.5194/nhess-14-1505-2014](https://doi.org/10.5194/nhess-14-1505-2014).
- Alsdorf, D. E., E. Rodríguez, and D. P. Lettenmaier, 2007: Measuring surface water from space. *Rev. Geophys.*, **45**, RG2002, doi:[10.1029/2006RG000197](https://doi.org/10.1029/2006RG000197).
- 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, doi:[10.1175/2008JHM1068.1](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. [Available online at <http://www.ecmwf.int/sites/default/files/elibrary/2010/7926-evaluation-era-interim-and-era-interim-gpcp-rescaled-precipitation-over-usa.pdf>.]
- , and Coauthors, 2015: ERA-Interim/Land: A global land water resources dataset. *Hydrol. Earth Syst. Sci.*, **19**, 389–407, doi:[10.5194/hess-19-389-2015](https://doi.org/10.5194/hess-19-389-2015).
- Bartholmes, J. C., J. Thielen, M. H. Ramos, and S. Gentilini, 2009: The European Flood Alert System EFAS—Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts. *Hydrol. Earth Syst. Sci.*, **13**, 141–153, doi:[10.5194/hess-13-141-2009](https://doi.org/10.5194/hess-13-141-2009).
- Berrisford, P., D. Dee, K. Fielding, M. Fuentes, P. Kallberg, S. Kobayashi, and S. Uppala, 2009: The ERA-Interim archive. ERA Rep. Series 1, ECMWF, 16 pp.
- Betts, A. K., M. Köhler, and Y.-C. Zhang, 2009: Comparison of river basin hydrometeorology in ERA-Interim and ERA-40 reanalyses with observations. *J. Geophys. Res.*, **114**, D02101, doi:[10.1029/2008JD010761](https://doi.org/10.1029/2008JD010761).
- Brakenridge, G. R., 2014: Global active archive of large flood events. Dartmouth Flood Observatory, University of Colorado, accessed 2 October 2014. [Available online at <http://floodobservatory.colorado.edu/Archives/index.html>.]
- , S. Cohen, A. J. Kettner, T. De Groeve, S. V. Nghiem, J. P. M. Syvitski, and B. M. Fekete, 2012: Calibration of satellite measurements of river discharge using a global hydrology model. *J. Hydrol.*, **475**, 123–136, doi:[10.1016/j.jhydrol.2012.09.035](https://doi.org/10.1016/j.jhydrol.2012.09.035).
- Burek, P., J. Van Der Knijff, and A. De Roo, 2013: LISFLOOD, distributed water balance and flood simulation model: Revised user manual. JRC Tech. Rep., 138 pp.
- Chow, V. T., D. R. Maidment, and L. M. Mays, 1988: *Applied Hydrology*. McGraw-Hill, 572 pp.
- Cloke, H. L., and F. Pappenberger, 2009: Ensemble flood forecasting: A review. *J. Hydrol.*, **375**, 613–626, doi:[10.1016/j.jhydrol.2009.06.005](https://doi.org/10.1016/j.jhydrol.2009.06.005).
- CRED, 2014: Centre for Research on the Epidemiology of Disasters Emergency database. Accessed 15 October 2014. [Available online at <http://www.emdat.be/>.]
- Dale, M., J. Wicks, J. Mylne, F. Pappenberger, S. Laeger, and S. Taylor, 2014: Probabilistic flood forecasting and decision-making: An innovative risk-based approach. *Nat. Hazards*, **70**, 159–172, doi:[10.1007/s11069-012-0483-z](https://doi.org/10.1007/s11069-012-0483-z).
- Dankers, R., and L. Feyen, 2008: Climate change impact on flood hazard in Europe: An assessment based on high resolution climate simulations. *J. Geophys. Res.*, **113**, D19105, doi:[10.1029/2007JD009719](https://doi.org/10.1029/2007JD009719).
- Decharme, B., R. Alkama, F. Papa, S. Faroux, H. Douville, and C. Prigent, 2012: Global off-line evaluation of the ISBA-TRIP flood model. *Climate Dyn.*, **38**, 1389–1412, doi:[10.1007/s00382-011-1054-9](https://doi.org/10.1007/s00382-011-1054-9).
- 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, doi:[10.1002/qj.828](https://doi.org/10.1002/qj.828).
- De Groeve, T., and P. Riva, 2009: Global real-time detection of major floods using passive microwave remote sensing. *Proc. 33rd Int. Symp. on Remote Sensing of Environment*, Stresa, Italy, ISRSE, 587–590.
- , and Coauthors, 2015: Joining forces in a global flood partnership. *Bull. Amer. Meteor. Soc.*, **96**, ES97–ES100, doi:[10.1175/BAMS-D-14-00147.1](https://doi.org/10.1175/BAMS-D-14-00147.1).
- Durand, M., E. Rodríguez, D. E. Alsdorf, and E. Trigg, 2010: Estimating river depth from remote sensing swath interferometry measurements of river height, slope, and width. *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, **3**, 20–31, doi:[10.1109/JSTARS.2009.2033453](https://doi.org/10.1109/JSTARS.2009.2033453).
- Dutra, E., G. Balsamo, P. Viterbo, P. M. A. Miranda, A. Beljaars, C. Schar, and K. Elder, 2010: An improved snow scheme for the ECMWF land surface model: Description and offline validation. *J. Hydrometeorol.*, **11**, 899–916, doi:[10.1175/2010JHM1249.1](https://doi.org/10.1175/2010JHM1249.1).
- ECMWF, 2015: Changes in ECMWF models. Accessed 9 September 2015. [Available online at <http://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model>.]
- FAO, 2003: Digital soil map of the world and derived soil properties. FAO Land and Water Digital Media Series 1, CD-ROM.
- Forbes, R., and A. Tompkins, 2011: An improved representation of cloud and precipitation. *ECMWF Newsletter*, No. 129, ECMWF, Reading, United Kingdom, 13–18.
- , T. Haiden, and L. Magnusson, 2015: Improvements in IFS forecasts of heavy precipitation, *ECMWF Newsletter*, No. 144, ECMWF, Reading, United Kingdom, 21–26.
- GDACS, 2014: Global Disaster Alert and Coordination System. Accessed 9 October 2014. [Available online at <http://www.gdacs.org/alerts/default.aspx?profile=archive>.]
- Gumbel, E. J., 1941: The return period of flood flows. *Ann. Math. Stat.*, **12** (2), 163–190.
- Haiden, T., and Coauthors, 2014: ECMWF forecast performance during the June 2013 flood in central Europe. ECMWF Tech. Memo. 723, 34 pp. [Available online at <http://www.ecmwf.int/en/elibrary/9731-ecmwf-forecast-performance-during-june-2013-flood-central-europe>.]
- Hirpa, F. A., T. Hopson, T. De Groeve, G. R. Brakenridge, M. Gebremichael, and P. J. Restrepo, 2013: Upstream satellite remote sensing for river discharge prediction: Application to major rivers in South Asia. *Remote Sens. Environ.*, **131**, 140–151, doi:[10.1016/j.rse.2012.11.013](https://doi.org/10.1016/j.rse.2012.11.013).
- 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**, L17808, doi:[10.1029/2009GL040000](https://doi.org/10.1029/2009GL040000).
- Jha, A. K., R. Bloch, and J. Lamond, 2012: *Cities and Flooding: A Guide to Integrated Flood Risk Management for the 21st Century*. World Bank, 632 pp., doi:[10.1596/978-0-8213-8866-2](https://doi.org/10.1596/978-0-8213-8866-2).
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–470, doi:[10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- Kim, H., P. J.-F. Yeh, T. Oki, and S. Kanae, 2009: Role of rivers in the seasonal variations of terrestrial water storage over global basins. *Geophys. Res. Lett.*, **36**, L17402, doi:[10.1029/2009GL039006](https://doi.org/10.1029/2009GL039006).
- Kobayashi, S., and Coauthors, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, doi:[10.2151/jmsj.2015-001](https://doi.org/10.2151/jmsj.2015-001).
- Lehner, B., K. Verdin, and A. Jarvis, 2008: New global hydrography derived from spaceborne elevation data. *Eos, Trans. Amer. Geophys. Union*, **89**, 93–94, doi:[10.1029/2008EO100001](https://doi.org/10.1029/2008EO100001).
- Loveland, T. R., B. C. Reed, J. F. Brown, D. O. Ohlen, Z. Zhu, L. Youing, and J. W. Merchant, 2000: Development of a global

- land cover characteristics database and IGBP DISCover from the 1 km AVHRR data. *Int. J. Remote Sens.*, **21**, 1303–1330, doi:[10.1080/014311600210191](https://doi.org/10.1080/014311600210191).
- McCuen, R. H., 2004: *Hydrologic Design and Analysis*. Prentice Hall, 888 pp.
- NASA/NGA, 2003: SRTM water body data product specific guidance, version 2.0. Accessed 15 August 2014. [Available online at http://dds.cr.usgs.gov/srtm/version2_1/SWBD/SWBD_Documentation/SWBD_Product_Specific_Guidance.pdf.]
- Pagano, T. C., and Coauthors, 2014: Challenges of operational river forecasting. *J. Hydrometeor.*, **15**, 1692–1707, doi:[10.1175/JHM-D-13-0188.1](https://doi.org/10.1175/JHM-D-13-0188.1).
- Pappenberger, F., and J. D. Brown, 2013: HP today: On the pursuit of (im)perfection in flood forecasting. *Hydrol. Processes*, **27**, 162–163, doi:[10.1002/hyp.9465](https://doi.org/10.1002/hyp.9465).
- , 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, doi:[10.5194/hess-16-4143-2012](https://doi.org/10.5194/hess-16-4143-2012).
- , F. Wetterhall, E. Dutra, F. Di Giuseppe, K. Bogner, L. Alfieri, and H. L. Cloke, 2013: Seamless forecasting of extreme events on a global scale. *IAHS Publ.*, **359**, 3–10.
- , and Coauthors, 2014: HEPEx—Operational HEPs systems around the globe. Accessed 3 October 2014. [Available online at <http://hepex.irstea.fr/operational-heps-systems-around-the-globe/>.]
- , H. L. Cloke, F. Wetterhall, D. J. Parker, D. Richardson, and J. Thielen, 2015: The monetary benefit of early flood warnings in Europe. *Environ. Sci. Policy*, **51**, 278–291, doi:[10.1016/j.envsci.2015.04.016](https://doi.org/10.1016/j.envsci.2015.04.016).
- Reed, S., J. Schaake, and Z. Zhang, 2007: A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged locations. *J. Hydrol.*, **337**, 402–420, doi:[10.1016/j.jhydrol.2007.02.015](https://doi.org/10.1016/j.jhydrol.2007.02.015).
- Revilla-Romero, B., H. E. Beck, P. Burek, P. Salamon, A. de Roo, and J. Thielen, 2015a: Filling the gaps: Calibrating a rainfall-runoff model using satellite-derived surface water extent. *Remote Sens. Environ.*, **171**, 118–131, doi:[10.1016/j.rse.2015.10.022](https://doi.org/10.1016/j.rse.2015.10.022).
- , F. A. Hirpa, J. Thielen, P. Salamon, G. R. Brakenridge, F. Pappenberger, and T. De Groeve, 2015b: On the use of global flood forecasts and satellite-derived inundation maps for flood monitoring in data-sparse regions. *Remote Sens.*, **7**, 15 702–15 728, doi:[10.3390/rs71115702](https://doi.org/10.3390/rs71115702).
- Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, doi:[10.1175/JCLI-D-11-00015.1](https://doi.org/10.1175/JCLI-D-11-00015.1).
- Sampson, C. C., T. J. Fawcett, F. O'Loughlin, F. Pappenberger, P. B. Bates, J. E. Freer, and H. L. Cloke, 2014: The impact of uncertain precipitation data on insurance loss estimates using a flood catastrophe model. *Hydrol. Earth Syst. Sci.*, **18**, 2305–2324, doi:[10.5194/hess-18-2305-2014](https://doi.org/10.5194/hess-18-2305-2014).
- Schumann, G. J. P., K. M. Andreadis, and P. D. Bates, 2014: Downscaling coarse grid hydrodynamic model simulations over large domains. *J. Hydrol.*, **508**, 289–298, doi:[10.1016/j.jhydrol.2013.08.051](https://doi.org/10.1016/j.jhydrol.2013.08.051).
- 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, doi:[10.5194/hess-13-125-2009](https://doi.org/10.5194/hess-13-125-2009).
- Thiemig, V., B. Bisselink, F. Pappenberger, and J. Thielen, 2014: A pan-African flood forecasting system. *Hydrol. Earth Syst. Sci.*, **11**, 5559–5597, doi:[10.5194/hessd-11-5559-2014](https://doi.org/10.5194/hessd-11-5559-2014).
- UNISDR, 2015: Sendai Framework for Disaster Risk Reduction 2015–2030. United Nations, 37 pp. [Available online at http://www.unisdr.org/files/43291_sendaiframeworkfordrren.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. Inf. Sci.*, **24**, 189–212, doi:[10.1080/13658810802549154](https://doi.org/10.1080/13658810802549154).
- Ward, P. J., and Coauthors, 2015: Usefulness and limitations of global flood risk models. *Nat. Climate Change*, **5**, 712–715, doi:[10.1038/nclimate2742](https://doi.org/10.1038/nclimate2742).
- Winsemius, H. C., L. P. H. Van Beek, B. Jongman, P. J. Ward, and A. Bouwman, 2013: A framework for global river flood risk assessments. *Hydrol. Earth Syst. Sci.*, **17**, 1871–1892, doi:[10.5194/hess-17-1871-2013](https://doi.org/10.5194/hess-17-1871-2013).
- Wood, A. W., and D. P. Lettenmaier, 2008: An ensemble approach for attribution of hydrologic prediction uncertainty. *Geophys. Res. Lett.*, **35**, L14401, doi:[10.1029/2008GL034648](https://doi.org/10.1029/2008GL034648).
- Wu, H., J. S. Kimball, H. Li, M. Huang, L. R. Leung, and R. F. Adler, 2012: A new global river network database for macroscale hydrologic modeling. *Water Resour. Res.*, **48**, W09701, doi:[10.1029/2012WR012313](https://doi.org/10.1029/2012WR012313).
- , R. F. Adler, Y. Thian, 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, doi:[10.1002/2013WR014710](https://doi.org/10.1002/2013WR014710).
- Yamazaki, D., G. A. M. de Almeida, and P. D. Bates, 2013: Improving computational efficiency in global river models by implementing the local inertial flow equation and a vector-based river network map. *Water Resour. Res.*, **49**, 7221–7235, doi:[10.1002/wrcr.20552](https://doi.org/10.1002/wrcr.20552).
- , F. O'Loughlin, M. A. Trigg, Z. F. Miller, T. M. Pavelsky, and P. D. Bates, 2014: Development of the global width database for large rivers. *Water Resour. Res.*, **50**, 3467–3480, doi:[10.1002/2013WR014664](https://doi.org/10.1002/2013WR014664).
- Yossef, N. C., H. Winsemius, A. Weerts, R. van Beek, and M. F. P. Bierkens, 2013: Skill of a global seasonal streamflow forecasting system, relative roles of initial conditions and meteorological forcing. *Water Resour. Res.*, **49**, 4687–4699, doi:[10.1002/wrcr.20350](https://doi.org/10.1002/wrcr.20350).
- Zsoter, E., F. Pappenberger, and D. Richardson, 2014: Sensitivity of model climate to sampling configurations and the impact on the Extreme Forecast Index. *Meteor. Appl.*, **22**, 236–247, doi:[10.1002/met.1447](https://doi.org/10.1002/met.1447).