Hydrol. Earth Syst. Sci. Discuss., 9, 12293–12332, 2012
www.hydrol-earth-syst-sci-discuss.net/9/12293/2012/
doi:10.5194/hessd-9-12293-2012


Hydrology and Earth System Sciences Discussions



# GloFAS – global ensemble streamflow forecasting and flood early warning

**L. Alfieri[1,2], P. Burek[2], E. Dutra[1], B. Krzeminski[1], D. Muraro[2], J. Thielen[2], and F. Pappenberger[1,3]**

[1]European Centre for Medium Range Weather Forecasts, Reading, UK
[2]European Commission – Joint Research Centre, Institute for Environment and Sustainability, Ispra, Italy
[3]College of Hydrology and Water Resources, Hohai University, Nanjing, China

Received: 15 October 2012 – Accepted: 25 October 2012 – Published: 2 November 2012

Correspondence to: L. Alfieri (lorenzo.alfieri@ecmwf.int)

Published by Copernicus Publications on behalf of the European Geosciences Union.





**Abstract**

Anticipation and preparedness for large-scale flood events have a key role in mitigating their impact and optimizing the strategic planning of water resources. Although several developed countries have well-established systems for river monitoring and flood early warning, figures of population affected every year by floods in developing countries are unsettling. This paper presents the Global Flood Awareness System, which has been set up to provide an overview on upcoming floods in large world river basins. The Global Flood Awareness System is based on distributed hydrological simulation of numerical ensemble weather predictions with global coverage. Streamflow forecasts are compared statistically to climatological simulations to detect probabilistic exceedance of warning thresholds. In this article, the system setup is described, together with an evaluation of its performance over a two-year test period and a qualitative analysis of a case study for the Pakistan flood, in summer 2010. It is shown that hazardous events in large river basins can be skilfully detected with a forecast horizon of up to 1 month. In addition, results suggest that an accurate simulation of initial model conditions and an improved parameterization of the hydrological model are key components to reproduce accurately the streamflow variability in the many different runoff regimes of the Earth.

## 1 Introduction

Weather-driven natural hazards, including storm surges, floods, flash floods, and subsequent mass movements, are the most prominent natural disasters in worldwide statistics (CRED, 2011). 57 % of the reported number of victims in 2011 is associated with so-called "hydrological disasters". These caused a total economic damage of more than 70 billion US dollars, meaning a 230 % average increase compared to the previous decade (Guha-Sapir et al., 2012). According to the United Nations International Strategy for Disaster Reduction (UN/ISDR, 2002) and statistics from insurance companies, the socio-economic impact of floods is increasing. With steadily rising world population, the need for optimizing the use of water resources for drinking water as well as energy production demands more and more technologically driven solutions for controlling water quantity and quality in river systems. In addition, floods can no longer be treated as isolated events, as they are heavily linked with issues such as food insecurity, disease outbreaks and environmental degradation (IFRC, 2011).

With increasing vulnerability and the likelihood of changes in frequency and intensity of future weather extremes (Trenberth et al., 2003), anticipation of severe events is becoming a key element to protect the society and favor timely reaction, thus effectively reducing socio-economic damage. While anticipation is essential at local level, it is equally important on national or trans-national level. The management of the response and aid for major upcoming disasters (e.g. through international organizations) requires a substantial planning and information at different levels. The earlier the planning phase starts, the better preparatory actions, coordination and gathering of information are achieved, thus limiting the consequences of potential humanitarian and economic disasters. While some countries have mechanisms in place to mitigate the effects of natural disasters, the European Union Solidarity Fund (European Commission, 2002) being the main example for Europe, developing countries often struggle through a much longer recovery process. Increasing preparedness can be achieved by flood hazard maps, which are available on national or regional level (Hagen and Lu, 2011; Prinos et al., 2008) as well as on global level (Pappenberger et al., 2012; Winsemius et al., 2012). These static maps can be used to define flood hazard zones, but do not incorporate changes in daily conditions, which require a real-time observing system.

The availability of remote sensing data, such as satellite imagery, has fostered the development of flood detection techniques at global scale (e.g., de Groeve, 2010; Proud et al., 2011; Westerhoff et al., 2012; Wu et al., 2012), that promptly produce overviews of affected areas and improve the management of rescue actions. To increase the preparedness towards floods and in general to water-related hazards, a number of research institutes and national hydro-meteorological services run





operationally flood forecasting systems, often focused on specific river basins or, most commonly, limited to national boundaries (Alfieri et al., 2012a). Several flood forecasting systems are based on observed river level, while future values are extrapolated through river routing models or by coupling observed rainfall fields into hydrological models. The extension of the forecast horizon beyond the response time of a river basin is enabled by the use of Numerical Weather Predictions (NWP) as input to hydrological-hydraulic models (e.g., He et al., 2010; Hopson and Webster, 2010; Paiva et al., 2012; Thiemig et al., 2010). Recent review articles by Cloke and Pappenberger (2009) and by Alfieri et al. (2012a) showed the strong potential of using ensemble NWP to further extend the forecasting horizon in early warning systems.

Weather forecasting models are set-up at global scale in different meteorological centers, producing deterministic and ensemble products. Nevertheless, only few attempts have been made so far, to move towards operational systems with coupled hydro-meteorological models providing streamflow predictions at the global scale (Candogan Yossef et al., 2011; see Sperna Weiland et al., 2010; Voisin et al., 2011; Wang et al., 2011) and, to the authors' knowledge, none of these runs operationally with ensemble predictions. Indeed real-time hydrological modeling requires a large amount of information, including not only static maps describing the surface and sub-surface basin features, but also a long-term balance of water fluxes to give an estimate of the initial conditions, from which the forecast is run. At the continental scale, the European Flood Awareness System (EFAS) has demonstrated that ensemble flood forecasting and early warning based on critical flood thresholds can be produced also with limited amount of data, by applying probabilistic methods and model consistent climatologies (Bartholmes et al., 2009; Pappenberger et al., 2010b; de Roo et al., 2003; Thielen et al., 2009a).

The aim of this study is to assess the feasibility of transferring methodologies and concepts from the EFAS system to the global scale and to evaluate the system performance at its initial stage, where no model parameter has been specifically calibrated. A Global Flood Awareness System (GloFAS) has been set up jointly between the Joint

Research Centre (JRC) of the European Commission and the European Centre for Medium-Range Weather Forecasts (ECMWF), and runs operationally on a daily basis since July 2011. GloFAS produces global flood forecasting products which are shown on a password-protected web interface. The system performance is currently being monitored and results are already accessed for research and testing purposes by partner organizations such as the Mekong River Commission (http://www.mrcmekong.org/) and the CEMADEN (http://www.cemaden.gov.br/), the newly established Brazilian center for monitoring of natural disasters.

## 2 Data and methods

The GloFAS system is composed by a well-integrated hydro-meteorological forecasting chain and by a monitoring system which analyzes daily results and shows forecast flood events on a dedicated web platform. An overview of the system structure is shown in Fig. 1.

### 2.1 Meteorological data

To set up a forecasting system that runs on daily basis with global coverage, initial conditions and input forcing data must be provided seamlessly to every point within the domain. To this end, two products are used. The first consists of operational ensemble forecasts of near surface meteorological parameters. The second is a long-term dataset consistent with daily forecasts, used to derive a reference climatology. These products are described in the next sub-sections. They are both computed by the Integrated Forecast System (IFS) of the ECMWF, whose main components (see Fig. 1) are a Data Assimilation System (DAS) and a Global Circulation Model (GCM).





#### 2.1.1 Daily forecasts

The Variable Resolution Ensemble Prediction System (VarEPS) is the operational ensemble forecasting product of the ECMWF IFS. VarEPS consist of 51-member ensemble global forecasts with 50 perturbed members and one unperturbed control run. The weather forecast component has horizontal grid resolution of about 32 km for 10 days, increasing to 65 km from day 11 to 15 (Miller et al., 2010). The forecast is produced twice per day, at 00:00 UTC and 12:00 UTC. In the GloFAS system, VarEPS weather forecasts are not handled explicitly. Forecast values of the predicted meteorological parameters of the 00:00 UTC forecast are processed by the land surface module (HTESSEL, see Sect. 2.2.1) of the IFS, which in turn creates the VarEPS runoff fields for the ensemble streamflow prediction.

#### 2.1.2 Reference climatology

The second meteorological product used is ERA-Interim (Dee et al., 2011), the latest global atmospheric reanalysis produced by the ECMWF. The ERA-Interim archive contains 6-hourly gridded estimates of three-dimensional (3-D) meteorological variables, 3-hourly estimates of a large number of surface parameters and other two-dimensional (2-D) fields. It has horizontal resolution of about 80 km, it covers the period from January 1989 onwards, and continues to be extended forward in near-real time. ERA-Interim makes use of a forecast model, so that information can be extrapolated from locally observed weather parameters to unobserved parameters in a physically meaningful way. ERA-Interim precipitation dataset has been bias corrected using the Global Precipitation Climatology project (GPCP) version 2.1 (Huffman et al., 2009). Bias corrected precipitation is used as input, together with other ERA-Interim meteorological fields, for an offline long-term HTESSL simulation (ERA-Interim/Land). The land surface reanalysis used for this study covers the period from January 1990 to December 2010, providing land surface conditions (e.g. soil moisture) and water fluxes (e.g. evapotranspiration, runoff).

## 2.2 Hydrological modeling

River discharge is simulated by the Lisflood hydrodynamic model (van der Knijff et al., 2010) for the flow routing in the river network and the groundwater mass balance. The model is set up on global coverage with horizontal grid resolution of 0.1° (about 10 km in mid-latitude regions) and daily time step for input/output data. Details of Lisflood and the HTESSEL are given in the following. Two types of simulations are performed to estimate discharge in the river network, which use the input runoff forcing described in the previous section and appropriate initial model state.

- Forecasting simulations are run every day using the latest VarEPS runoff prediction and result in 51 possible evolutions of the streamflow for the selected forecast horizon (i.e., 45 days in the current setting).
- A deterministic climatological simulation is run in offline mode using ERA-Interim/Land input data for a 21 yr period starting in 1990. Seamless streamflow climatology is derived and maps of annual maxima are extracted and fitted with a Gumbel extreme value distribution to estimate corresponding discharge warning thresholds for selected return periods.

### 2.2.1 HTESSEL

HTESSEL (Balsamo et al., 2009, 2011a) is the land surface component of the ECMWF IFS. It is a revised land surface Hydrology, derived from the former Tiled ECMWF Scheme for Surface Exchanges over Land (TESSEL). HTESSEL computes the land surface response to atmospheric forcing, and estimates the surface water and energy fluxes and the temporal evolution of soil temperature, moisture content and snowpack conditions. At the interface to the atmosphere each grid box is divided into fractions (tiles), with up to six fractions over land (bare ground, low and high vegetation, intercepted water, shaded and exposed snow). Vegetation types and cover fractions are



derived from an external climate database, based on the Global Land Cover Characteristic (Loveland et al., 2000).

The grid box surface fluxes are calculated separately for each tile, leading to a separate solution of the surface energy balance equation and the skin temperature. The latter represents the interface between the soil and the atmosphere. Below the surface, the vertical transfer of water and energy is performed using four vertical layers to represent soil temperature and moisture. Soil heat transfer follows a Fourier law of diffusion, modified to take into account soil water freezing/melting. Water movement in the soil is determined by Darcy's Law, and surface runoff accounts for the subgrid variability of orography. In the case of a partially (or fully) frozen soil, water transport is limited, leading to a redirection of most of the rainfall and snowmelt to surface runoff when the uppermost soil layer is frozen. The snow scheme (Dutra et al., 2010) represents an additional layer on top of the soil, with an independent prognostic thermal and mass content. The model has been successfully tested in a river routing settings (Balsamo et al., 2011b; Pappenberger et al., 2010a). HTESSEL is part of the IFS at ECMWF with operational applications ranging from the short-range to monthly and seasonal weather forecasts.

For this work, operational ensemble forecasts of surface and sub-surface runoff are extracted from the daily output of the ECMWF forecasts. These are produced by the HTESSEL module of the IFS using VarEPS weather forecasts as input. Further, an offline simulations of HTESSEL forced by ERA-Interim near surface fields was performed to derive a 21-yr climatology starting in 1990, including surface and sub-surface runoff (ERA-Interim/Land). Balsamo et al. (2012) presented a detailed description of the simulation set-up of ERA-Interim/Land and a general overview of the model performance.

### 2.2.2 Lisflood global

Lisflood is a GIS-based spatially-distributed hydrological model, which includes a one-dimensional hydrodynamic channel routing model (van der Knijff et al., 2010). The Lisflood model is currently running within the European Flood Awareness System (EFAS) on an operational basis (Pappenberger et al., 2010b; Thielen et al., 2009a) covering the whole Europe on a 5 km grid.

In the context of global flood modeling, the transformation from precipitation to surface and sub-surface runoff is done by the HTESSEL module of the IFS. Lisflood global is stripped down to the groundwater and routing procedures and uses surface runoff and sub-surface runoff from HTESSEL as input fluxes on a resolution of $0.1^\circ$. Surface runoff is routed to the outlet of each cell using a four-point implicit finite-difference solution of the kinematic wave equations (Chow et al., 1988). The Global Land Cover 2000 dataset (Bartholomé and Belward, 2005) is used to derive surface roughness coefficients.

Subsurface storage and transport are modeled using two parallel linear reservoirs. The upper zone represents a quick runoff component, which includes fast groundwater and subsurface flow through macropores in the soil. The lower zone represents the slow groundwater component that generates the base flow. As for the sub-surface runoff, all water that flows out of the upper and lower groundwater zones is routed to the outlet of each grid cell within one time step. Runoff produced for every grid cell from surface, upper and lower groundwater zones is routed through the river network using a kinematic wave approach. The river network is taken from the Hydrosheds project (Lehner et al., 2008).

In arid and semiarid regions one can observe a loss of water among the channel reaches. In order to include this effect into the model we use the simplified approach by Rao and Maurer (1996) to simulate transmission losses in a stream. This method uses a power function with two parameters to describe the relationship between inflow and outflow in cells. In a first attempt the yearly average potential evapotranspiration rate is used to fit the transmission loss function. The resulting loss function give emphasis to transmission losses in Africa, the Arabic Peninsula, India, Australia and the southern part of North America whereas discharge in Europe and northern part of Asia remains unaffected. With this approach the model is able to mimic the river-aquifer and





river-floodplain interaction (e.g. the big Sudan swamps in the Nile River) as well as the influence of evaporation from braided rivers.

### 2.3 Operational monitoring

Ensemble streamflow predictions (ESP) are run operationally on global scale by feeding VarEPS surface and sub-surface runoff into the Lisflood hydrological model. Although the precipitation input spans 15 days, hydrological simulations are computed for a 45-day time horizon, to account for the delayed routing of flood waves in large river basins, with time of concentration of the order of one month. Initial condition maps to start up the model are first taken from the last available day of ERA-Interim dataset. Initial conditions for subsequent simulations are then extracted from the results of the model run with the VarEPS control run, after the first day of simulation. As this procedure is based on forecast meteorological variables as input, rather than observed, results may possibly drift in time from the reality. Therefore, periodical updating of initial condition maps based on ERA-Interim dataset is foreseen for future system developments.

Resulting ESP maps for each daily time step and ensemble member are compared with reference threshold maps derived from the streamflow climatology, corresponding to return periods of 2, 5 and 20 yr. Summary threshold exceedance maps are calculated accordingly, which show the maximum probability of exceeding the 5 and 20-yr return period within the forecast horizon. In addition, reporting points are chosen at fixed and dynamic locations in the river network where upcoming flood hazard is detected, according to the following two-step procedure.

*Fixed points* are first selected from ~~a database of~~ about 4000 gauged river stations included in the Global Runoff Data Centre (GRDC, http://grdc.bafg.de/) database, where the maximum forecast value of the ESP mean, over the simulation horizon, is above the 2-yr return period threshold.

*Dynamic points* are then generated to provide similar information in river reaches where no fixed point is available. The following experience-based rules are adopted for obtaining a good overview of the potentially affected areas, yet avoiding the confusion of displaying too many points:

– The ESP mean is above the medium warning threshold on at least 5 contiguous pixels of the river network (~ 50 km long river reach), in at least one of the two most recent daily simulations.

– The upstream area of the selected point must be larger than 4000 km$^2$.

– Points are generated starting from the most downstream pixel complying with the selection criteria, proceeding upstream every 300 km to each other, unless a fixed point is encountered within a shorter distance.

The two sets of points are merged and classified into medium, high and severe alert level. Medium alert level (yellow color coding) is assigned to points with ESP mean between 2 and 5-yr return period. High alert level (red color coding) is assigned to points with ESP mean between 5 and 20-yr return period. Severe alert level (purple color coding) is assigned to points with ESP mean above 20-yr return period. At each point, ESP time series are plotted versus the forecast horizon, together with persistence diagrams (Bartholmes et al., 2009) showing the probability of exceeding the three warning thresholds for each day of simulation and the evolution over the latest consecutive forecasts.

## 3 Methods

### 3.1 Evaluation of the hydrological modeling

The first part of the work is focused on evaluating the skill of the Lisflood hydrological model forced by ERA-Interim/Land runoff in reproducing the hydrological processes for river basins in different regions and climates of the Earth. The 21-yr simulated discharge climatology has been compared with daily observations at a number of stations



included in the GRDC database. Stations for the comparison were chosen according to the three following criteria:

– Observed discharge time series at each station must include at least 5 yr of valid data within the simulation period (1990–2010).

– At each river station, the upstream area of the modeled river network must not differ by more than 10 % from the actual one, to prevent from matching incoherent data pairs. This typically occurs in small river basins, where the modeled river network is sometimes different from the real one – because of scaling issues – and as a result the station does not lie in the correct grid cell.

– A visual check has been performed on the observed time series to remove those stations with evident discharge regulation (e.g., through artificial reservoirs) or with clear errors in the data.

Overall, 620 stations from five continents (all but Antarctica) were selected for the comparison, with upstream area ranging between 450 and 4 680 000 km$^2$ and period of record between 5 and 21 yr. The distribution of stations (see Fig. 2) reflects the quantity and quality of daily discharge measurements, with most data coming from North America, Brazil, Europe, Japan and Australia. The aim of this analysis is to assess how the adopted model is capable to reproduce observed river discharge. The expected outcome is to assess the model performance and identify areas with the most significant mismatch between observation and simulations, which indicates where the modeling can be improved through different parameterization of the hydrological processes. For each station, observed and simulated discharge time series are plotted and compared through scatter plots, to give a first visual check of the collected data. An example is shown in Fig. 3 and Fig. 4 where simulated and observed discharges at Obidos – linigrafo (Brazil), in the Amazon River, are compared over 17 yr between 1990 and 2007. The performance of simulation is assessed for each station through different skill scores, namely the Nash-Sutcliffe efficiency (NS), Pearson correlation coefficient





Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

(PCC), Root mean square error (RMSE), Mean absolute error (MAE) and Coefficient of variation (CV) of the residuals towards the observed mean.

It is worth noting that the proposed system is designed for early warning purpose, rather than for quantitative streamflow forecasting. In other words, the main goal of the system is to assign each forecast value a correct probability of occurrence taken from its cumulative distribution function and thus identify extreme values in the upper tail of the distribution, which can possibly correspond to flooding conditions. Ideally, the percentile rank of each simulated value, compared to its climatology, should match that of observations (related to the observed time series), independently of any bias between observed and simulated time series. As a result, more emphasis is given to skill scores that are not affected by bias of estimation. Also, dimensionless indicators are preferred, as these enable straightforward comparison of results from different river stations having a wide range of quantitative runoff and hydrological regimes. Among such skill scores, the Coefficient of variation (CV) at each point is calculated as the ratio of the standard deviation ($\sigma(.)$) of the estimation residuals to the mean ($\overline{Q_{\text{obs}}}$) of observations,

$$\text{CV} = \frac{\sigma(Q_{\text{sim}} - Q_{\text{obs}})}{\overline{Q_{\text{obs}}}}. \quad (1)$$

The CV enables the comparison of the estimation variability at different locations through normalization by the average flow conditions. Furthermore, the Pearson correlation coefficient (PCC) of simulated versus observed discharges is calculated according to the equation:

$$\text{PCC} = \frac{\sum_{\forall i}(Q_{\text{obs}_i} - \overline{Q_{\text{obs}}})(Q_{\text{sim}_i} - \overline{Q_{\text{sim}}})}{\sqrt{\sum_{\forall i}(Q_{\text{obs}_i} - \overline{Q_{\text{obs}}})^2}\sqrt{\sum_{\forall i}(Q_{\text{sim}_i} - \overline{Q_{\text{sim}}})^2}}, \quad (2)$$

which considers all the $i$-th available daily data pairs. PCC is particularly fit to the desired verification strategy as it assesses the linear correlation between simulated



and observed discharges, without being penalized by multiplicative or additive bias. On the other hand the PCC is known for being sensitive to even few outlying data pairs, thus stressing significant shifts between the timing of simulated and observed flow peaks (Wilks, 2006).

The model performance in reproducing observed discharge has also been tested through threshold exceedance analysis, focused on discriminating events above a fixed threshold. This approach is more suitable for evaluating the performance of early warning systems, as it is independent of the quality of estimation for value ranges far from the threshold (e.g., the range of low flows when the threshold corresponds to high flows). Most scores for dichotomous evaluation are based on contingency tables, which include four variables calculated from the set of observations and of simulated values:

- Hit: event observed and simulated.
- Miss: event observed and not simulated.
- False alarm: event simulated and not observed.
- Correct negative: event not observed and not simulated.

The Peirce's skill score (PSS, Eq. 3) has been calculated for each station, taking the 90th percentile as threshold values (i.e., the 90th percentile from the sorted observations and from the sorted simulated values to discriminate each corresponding data series).

$$\text{PSS} = \frac{\text{hits}}{\text{hits} + \text{misses}} - \frac{\text{false alarms}}{\text{false alarms} + \text{correct negatives}} \quad (3)$$

Such percentile is a good tradeoff between being representative of high flow values and including a sufficient number of events to draw robust statistics. Data series for comparison include at least five years of data, which corresponds to more than 182 days above the 90 % threshold. The PSS accounts for all elements of the contingency table and is defined as the difference between Probability of detection (POD) and Probability of false detection (POFD), PSS = POD − POFD. Perfect forecasts have PSS = 1, while forecasts have no value when PSS ≤ 0.

### 3.2 Performance of the early warning system

The early warning system, as described in Sect. 2.3, has been set up and runs operationally since July 2011. To evaluate the forecast performance, the system was run in hindcast mode for the period 1 January 2009 to 31 December 2010. 730 sets (i.e., one per day) of 45-day ensemble streamflow predictions (ESP) were evaluated against discharge proxy simulations for the same period, taken from the simulated discharge climatology obtained using ERA-Interim/Land runoff as forcing. Differently from the analysis in the previous section this approach enables the performance evaluation at each grid point of the simulated river network. Furthermore, as the datasets of streamflow predictions and proxy simulations are generated by the same hydrological model, this type of analysis focuses on the skills of the ensemble weather predictions. Indeed, it allows one to draw indications on the maximum forecast horizon (or potential skill) for which the system yields valuable information. In general, we expect results to be mainly influenced by (i) the skills of 15-day weather predictions and by (ii) the upstream area of each selected river point, which is correlated to the lag time between rainfall events and the subsequent flow hydrographs. This can yield an extension of the forecast lead time beyond the time window for which weather forecasts are available and contribute to the assessment of the limits of predictability (Thielen et al., 2009b).

Initial conditions of the hydrological model were taken from the climatological run for the first day of simulation (i.e., 1 January 2009) and were then calculated for the following days, up to the 31 December 2010, by using the forecast fields of the first day of the VarEPS control run. Current ERA-Interim data availability would allow the model to update its initial conditions roughly on a monthly basis, to avoid significant drifts of the simulated initial conditions from the climatological run. To account for this improvement in the evaluation of the 2-yr forecasts, a bias correction technique was applied to adjust





the initial conditions of the starting day of each forecast with those of the climatological run. The correction was performed through a quantile matching over a 30-day window, which reproduces similar error structure as of a monthly update of initial conditions with ERA-Interim input data. The resulting discharge dataset is hereinafter referred to as corrected discharge climatology.

Ensemble streamflow predictions were evaluated by means of a twofold approach. The Continuous Rank Probability Skill Score (CRPSS, see e.g., Hersbach, 2000; Voisin et al., 2010) is used to evaluate the quantitative skills of prediction, while the Area under the Receiver Operating Characteristic (AROC, see Marzban, 2004; Wilks, 2006) is calculated to assess the performance in threshold exceedance analysis.

The CRPSS is defined as:

$$\mathrm{CRPSS} = \frac{\overline{\mathrm{CRPS}_{\mathrm{ref}}} - \overline{\mathrm{CRPS}_{\mathrm{forecast}}}}{\overline{\mathrm{CRPS}_{\mathrm{ref}}}} \tag{4}$$

where

$$\mathrm{CRPS} = \int_{-\infty}^{\infty} [F(y) - F_0(y)]^2 \mathrm{d}y \tag{5}$$

and

$$F_0(y) = \begin{cases} 0, & y < \text{observed value} \\ 1, & y \geq \text{observed value} \end{cases} \tag{6}$$

while $F(y)$ is the stepwise cumulative distribution function (cdf) of the ESP for each forecast day and lead time. The CRPS accounts for the integrated squared difference between the ESP and the step function of the proxy truth, here represented by the corrected climatological run for the 2 yr of forecast. The CRPSS is a dimensionless indicator of the skills of ensemble predictions, measured by $\overline{\mathrm{CRPS}_{\mathrm{forecast}}}$, compared to that of a reference forecast $\overline{\mathrm{CRPS}_{\mathrm{ref}}}$, which assumes all future values being equal to the latest observation (persistence criterion), meaning the value at timestep $i = 0$ which is used to initialize each forecast. CRPSS ranges between 1 (for perfect predictions) to $-\infty$, though ESP are only valuable when CRPSS > 0, i.e., when ensemble forecasts perform better than the reference persistent forecast.

Receiver Operating Characteristic (ROC) curves are widely used to measure the skill of dichotomous forecasts based on probabilistic information, as they plot the empirical relation between the Hit Rate (HR) and False Alarm Rate (FAR) for different probability thresholds (Alfieri et al., 2012b). The overall performance of ensemble forecasts in predicting threshold exceedances can be assessed though the area under the ROC curve, which summarizes the system skill for all the probability thresholds, which in the discrete case are as many as the ensemble size. AROC values range between 0 (i.e., forecasts are exactly the opposite of observations) and 1 (perfect match between predicted and observed threshold exceedances). AROC = 0.5 corresponds to random forecasts, while meteorological ensemble predictions are commonly considered as useful when AROC ≥ 0.7 (e.g., Buizza et al., 1999).

## 4 Results

### 4.1 Evaluation of the hydrological modeling

The coefficient of variation as defined in Eq. (1) is shown in the map in Fig. 2. In 60 % of points, the CV is found smaller than 1, denoting a variability of estimation lower than the observed mean discharge. Poorest performance is mainly found in arid and semi-arid regions, particularly in Australia, Mexico and in the Sahel. This can be due to incorrect modeling of some hydrological processes such as evapotranspiration, infiltration and lack of simulated water withdrawals for irrigation purpose. However, one should note that in arid regions, results calculated with the CV as defined above are penalized by rather low average discharges (which is in the denominator in Eq. 1), compared to





high flow conditions. Similar consideration can be drawn for small river basins, such as the yellow/orange circles in the USA and Europe shown in Fig. 2. Indeed, it is known that the ratio between peak flow and average flow rises with decreasing basin area, hence increasing the weight of estimation residuals in Eq. (1). In addition, clusters of points with CV > 1 are located in north-eastern Brazil and west-Africa, where the model performance is often substantially affected by dam regulation, and in the north-eastern Russia, where most discrepancies are related to the modeling of freezing cycles, snow accumulation and melting processes.

In Fig. 5 the PCC is plotted against the upstream area of each gauge. In addition the gauge latitude is shown with a color shading ranging from red at the Equator, to blue at high latitudes. Fig. 5 shows that the best correlations are on average achieved in large river basins (i.e., upstream area larger than 10 000 $\mathrm{km}^2$) in inter-tropical latitudes. Overall, 71 % of points have PCC larger than 0.5. The envelope curve of highest PCC values shows an increasing trend with the upstream area. In fact, the typical scales of weather events inducing floods in small river basins are below the spatial and temporal resolution of the hydrological model and of the meteorological input data used in simulation, as well as of the observations used for validation.

The Peirce's skill score (PSS) for the set of selected stations is shown in Fig. 6. 98.5 % of stations provide valuable simulated values (i.e., PSS > 0), while PSS > 0.25 and PSS > 0.5 is found in 79 % and 22 % of cases respectively. It is worth noting in Fig. 6 that positive skills are achieved at several stations in dry regions where the estimation error showed considerable variability in Fig. 2 (e.g., NE Brazil, Africa, Australia). In those regions medium to low flows are difficult to estimate accurately because of dam regulation and water abstraction for irrigation. On the other hand, floods and high flows, and particularly their percentile rank, are less influenced by reservoirs which often have limited storage for flood mitigation. Regarding negative PSS values, 8 out of 9 points in total are located in Canada and have relatively small upstream area, in all cases below 50 000 $\mathrm{km}^2$. Graphs comparing the observed and simulated time series (not shown) suggest that the mismatch in those points is due to incorrect modeling of the snow-related processes or to biased input temperatures in the model, which induces a substantial delay between observed and simulated flow peaks.

### 4.2 Performance of the early warning system

CRPSS maps for the 2 yr of ensemble streamflow prediction (i.e., 2009–2010) were calculated for each selected forecast lead time from 1 to 45 days. CRPSS maps with lead time of 5, 15 and 25 days are shown in Fig. 7. To improve the figure readability, only river pixels with upstream area larger than 50 000 $km^2$ are plotted. Valuable quantitative ESP are indicated with blue shadings in Fig. 7, that is where CRPSS > 0, while in red are indicated those rivers where a reference persistent forecast performs quantitatively better. As expected, the CRPSS deteriorates for increasing forecast horizons, particularly in smaller rivers. Poorest performance is shown in northern cold regions, mostly in Asian and North American rivers. In large river basins in inter-tropical and mid-latitude regions (e.g., Amazon, Mississippi, Congo, Nile, Paraná) the ESP perform better than the reference forecast, especially for longer lead times. In fact, in such rivers the runoff has very slow and delayed response, hence for short lead times (e.g., 5 days) the difference between the ESP and a persistent forecast is not substantial. On the other hand, smaller river basins often have their highest CRPSS for shorter lead times, while it decreases fast after 15-day lead time, when no meteorological forcing is used as input.

The threshold exceedance analysis is evaluated through the use of ROC curves and specifically the area under these curves. As discussed in Sect. 3.1, the threshold between events and non-events is set to the 90th percentile of the corrected discharge climatology. Although it usually does not correspond to flooding conditions, it is important to select a discharge value that was reached at every river pixel during the 2 yr of simulation, so that the skill score can be calculated for the whole domain. Results of this analysis are drawn in Fig. 8, which shows the maximum lead time over which forecasts are skillful (i.e., AROC > 0.7, as stated in Sect. 3.2). Spatial pattern of results in Fig. 8 is widely in agreement with those of Fig. 7. Longest lead times are found in large





river basins in South America, Africa, South Asia, with values exceeding 25 days in some areas. Smaller river basins mostly achieve maximum forecast lead times around 20 days, while in some cases they are limited within 10 days. Results from the ESP as calculated by the proposed model and shown in Fig. 7 and Fig. 8 should be filtered by excluding regions where no significant river network and runoff exists. These include deserted areas such as the Sahara, Arctic, Gobi, Arabian and Australian Desert, among the largest. Unexpectedly, in the lowest part of the Mississippi River, in North America, maximum values are within 10 days, despite having skillful CRPSS for lead times as long as 25 days (see Fig. 7). In other words, while quantitative streamflow predictions in the Mississippi are on average rather accurate even for long lead times, high flow events above the 90th percentiles are skillfully detected only for a shorter time horizon. The reasons for such behavior are mostly related to a delay in the discharge peak for the main event within the considered period, occurred in autumn 2009 (see Fig. 9).

In Fig. 9, ESP and corrected discharge climatology are compared for the 2 yr of available forecast (i.e., 2009–2010). These are shown at the outlet of six major river basins in different climatic regions, for forecast lead times of 15 and 25 days. Outlets location and name initials of each river are shown with red markers in Fig. 8. In all cases shown, the ensemble spread is relatively narrow, as in such large river basins the runoff is mostly driven by the initial conditions and, specifically, by water already in the river network at the start of the forecast, and that is conveyed downstream by the hydrological model. At all locations the ensemble spread is larger for the longest lead time shown, reflecting the increasing uncertainty range as the lead time increases. However, graphs with longer forecast lead time (not shown in the article) suggest that after reaching its maximum, the ensemble spread tends to reduce after the predicted rainfall has drained through the basin outlet. This is the consequence of using 15 days of rainfall but simulating a longer lead time, which means that the ESP spread is increasingly underestimated after the day 15 of simulation. In five out of six stations in Fig. 9, the runoff regime follows a clear seasonal trend, with peak flows always in the same range of months, depending on the rainfall regime and on the timing of snow and ice accumulation and melting. Differently, in the Mississippi River, the runoff regime is more variable and high flows occurred in different seasons. This partly explains the results shown in Fig. 7, where the ESP performs quantitatively better than a persistent forecast also for long lead times (i.e., 25 days). Graphs in Fig. 9 shows that the ESP spread is higher when the hydrographs have increasing trend, because of the uncertainty of forecast rainfall. On the other hand, as the reference simulation and the ESP are outputs of the same hydrological model, results matches very well in the recession part of the hydrographs, that is when little rainfall is forecast or during the period of snow accumulation. It is worth noting that the highest spread of the EPS occurs in the Yenisei River, where the snow and ice melting in the spring season play a prominent role in generating high flows. As a result the ensemble spread is amplified as the uncertainty of both rainfall and temperature affects the streamflow forecast.

### 4.3 Case study – 2010 Pakistan floods

The system demonstrated its potential by detecting a number of flooding events of the past 3 yr in major world rivers, with forecast lead time often larger than 10 days. A striking example is that of the severe floods that hit Pakistan in summer of 2010, triggered by exceptional monsoon rain beginning at the end of July. The flooding covered approximately one-fifth of the total land area of Pakistan, directly affecting about 20 millions and causing a death toll close to 2000 people (http://en.wikipedia.org/wiki/2010_Pakistan_floods). On that occasion, forecasts on 28 July 2010 showed probabilities up to 100 % of exceeding the severe alert level (i.e., 20-yr return period) in most of the Indus River Basin, with peak flow travelling downstream on the first half of August 2010. Figure 10 shows a 20-day ensemble streamflow prediction for a dynamic reporting point generated by GloFAS on the Indus River, few kilometers downstream the city of Sukkur, in the Sindh province of Pakistan. Forecasts show a sharp rise of discharge in the river, with expected peak of the ensemble mean on the 10 August, hence 13 days after the prediction was issued. The uncertainty range increases with





the lead time, though it almost completely exceeds the severe alert level from the 8 to the 12 August. The following forecasts confirmed these results indicating, for the same station, a maximum probability to exceed the severe alert level of 100 % from the 7 August onwards. On 9 August, the BBC reported that the measured discharge through the Sukkur Barrage was up to 1.4 million cubic feet per second (cusecs), i.e., about $39\,600\,\mathrm{m}^3\,\mathrm{s}^{-1}$, way higher than its design capacity of 900 000 cusecs. Also, Fig. 11 shows a comparison between satellite images taken on 10 July 2010 (top) and on 11 August 2010 (bottom) from MODIS Rapid Response. In the latter the extent of flooded areas (with dark shades) is clearly visible for a wide portion of the Indus River Basin. In the top panel of Fig. 11, the maximum probability of exceeding the severe warning threshold over the forecast range (i.e., 20-yr return period) is indicated with purple shades (ESP of 28 July 2010).

## 5 Discussions and conclusions

In this article we present a probabilistic flood early warning system running at global scale, aimed at forecasting the threshold exceedance of ensemble streamflow predictions on the basis of a model-consistent discharge climatology. The system has been set up and runs on a daily basis since July 2011. Results are shown on a password-protected website and are being monitored to assess qualitatively the system skills for flood events in large river basins. Quantitative performance has been assessed for 2 years of daily hindcasts starting on 1 January 2009, using a simulated climatology as reference run.

Findings of this analysis show that current ensemble weather predictions enable skillful detection of hazardous events with forecast horizon as long as 1 month, in large river basins. This anticipation depends on the skill of input weather forecasts and on the delay between the meteorological forcing and the hydrological response in the river basin. Interestingly, the uncertainty range of ensemble weather predictions has a reduced effect when propagated to discharge predictions in large river basins.

Indeed, flood events in major rivers are mostly caused by large scale weather systems that are skillfully predicted by state-of-the-art global forecasting models. In addition, when weather systems have smaller or similar size as that of the river basin, spatial shifts of predicted rainfall fields have limited effect on the resulting streamflow at the outlet.

With regard to the system performance in quantitative forecasting and early warning, the maximum added value is shown (i) in medium-size river basins, (ii) in those with relatively fast response and (iii) in basins with no definite trend in the seasonal runoff. At the lower boundary of the range of basin size, forecast performance deteriorates quickly with increasing lead time and with decreasing upstream area. Indeed, in these river basins, flood events are caused by small size weather systems which cannot be properly modeled by the current system, as the model space-time resolution is comparatively coarse for their typical hydro-meteorological dynamics. Consequently, on the basis of the analysis performed in this work the authors suggest a lower boundary of 10 000 $km^2$ as the minimum upstream area to consider for streamflow predictions provided by the model.

In contrast, in the largest world river basins (i.e., basin area larger than 1 million $km^2$) variations of river discharge occur at slow rates, hence the 1 to 10-day streamflow prediction does not differ substantially from a persistent forecast (i.e., the last observed discharge value). On the other hand, results for these basins show skillful predictions for lead times up to one month, whereas the highest added value compared to persistent forecast is provided for lead times of 10 ÷ 30 days (see Fig. 7). Besides the slow response, large river basins have long memory, so even small errors in model components such as snow accumulation and soil moisture can sum up over long time and induce a considerable bias in the water balance. An accurate estimation of the initial model state is therefore of crucial importance for the overall system performance. This can be achieved by regularly updating the water balance using the latest input data from ERA-Interim reanalysis, to improve the consistency between ensemble forecasts and the climatological warning thresholds.





This work shows the system setup and skills in its initial stage, that is, no calibration has been performed on the hydrological model behind. This is an important step for future improvements, particularly for a global system which therefore includes the full range of climates and hydrological regimes of the Earth. Results in Figs. 7–8 show the current system potential assuming that the simulated climatology corresponds to the actual river conditions, that is, for a perfectly calibrated hydrological model. The presented research work shows that there is substantial room for improving the current model parameterization, with particular focus on hydrological regimes in arid and cold regions. However, errors coming from the hydrological modeling and from the weather predictions do not sum up linearly in the assessment of the overall system performance. As stated in Sect. 3.1, the main goal of an early warning system is to match the percentile rank of each simulated and observed discharge, rather than minimizing quantitative values. In addition, the model capability would also benefit by improved weather forecasts and possibly by the use of input data with longer forecast horizon. In this regard, the use of monthly ECMWF VarEPS forecasts – currently issued twice per week – is envisaged for future system applications.

As a final remark, the current system is based on warning thresholds with fixed probability levels, corresponding to selected return periods. Actual flood risk also depends on the vulnerability of each area. For instance, in little populated areas or in regions with prominent flood defense works, the 100 yr discharge may cause limited economic damage. Conversely in densely populated areas with poor flood protection measures, peak discharges with relatively low return period can cause severe damage. The coupling of hazard and vulnerability maps would be extremely beneficial for this system, in order to rank warnings according to the potential economic damage that floods can cause as well as to the corresponding affected population.

## References


Alfieri, L., Salamon, P., Pappenberger, F., Wetterhall, F., and Thielen, J.: Operational early warning systems for water-related hazards in Europe, Environ. Sci. Policy, 21, 35–49, 2012a.

Alfieri, L., Thielen, J., and Pappenberger, F.: Ensemble hydro-meteorological simulation for flash flood early detection in southern Switzerland, J. Hydrol., 424–425, 143–153, doi:10.1016/j.jhydrol.2011.12.038, 2012b.

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., and Betts, A. K.: A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System, J. Hydrometeor., 10, 623–643, doi:10.1175/2008JHM1068.1, 2009.

Balsamo, G., Boussetta, S., Dutra, E., Beljaars, A., Viterbo, P. and Van de Hurk, B.: Evolution of land surface processes in the IFS, ECMWF Newsletter, 127, 17–22, 2011a.

Balsamo, G., Pappenberger, F., Dutra, E., Viterbo, P., and van den Hurk, B.: A revised land hydrology in the ECMWF model: a step towards daily water flux prediction in a fully-closed water cycle, Hydrol. Process., 25, 1046–1054, 2011b.

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D., Dutra, E., Pappenberger, F., de Rosnay, P., Muñoz Sabater, J., Stockdale, T., Vitart, F.: ERA-Interim/Land: A global land-surface reanalysis based on ERA-Interim meteorological forcing, ERA Report Series, ECMWF, Shinfield Park, Reading, 2012.

Bartholomé, E. and Belward, A. S.: GLC2000: A new approach to global land cover mapping from earth observation data, Int. J. Remote Sens., 26, 1959–1977, 2005.

Bartholmes, J. C., Thielen, J., Ramos, M. H., and Gentilini, S.: 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, 2009.

Buizza, R., Hollingsworth, A., Lalaurette, F., and Ghelli, A.: Probabilistic Predictions of Precipitation Using the ECMWF Ensemble Prediction System, Weather Forecast., 14, 168–189, doi:10.1175/1520-0434(1999)014<0168:PPOPUT>2.0.CO;2, 1999.

Candogan Yossef, N., van Beek, L. P. H., Kwadijk, J. C. J., and Bierkens, M. F. P.: Skill assessment of a global hydrological model in reproducing flow extremes, Hydrol. Earth Syst. Sci. Discuss., 8, 3469–3505, doi:10.5194/hessd-8-3469-2011, 2011.

Chow, V. T., Maidment, D. R., and Mays, L. W.: Applied hydrology, McGraw-Hill Science/Engineering/Math., 1988.







Cloke, H. L. and Pappenberger, F.: Ensemble flood forecasting: A review, J. Hydrol., 375, 613–626, 2009.

CRED, (first): EM-DAT: The OFDA/CRED International Disaster Database – www.emdat.be, Université Catholique de Louvain, Brussels (Belgium), 2011.

de Roo, A. P. J., Gouweleeuw, B., Thielen, J., Bartholmes, J., Bongioannini-Cerlini, P., Todini, E., Bates, P. D., Horritt, M., Hunter, N., Beven, K., Pappenberger, F., Heise, E., Rivin, G., Hils, M., Hollingsworth, A., Holst, B., Kwadijk, J., Reggiani, P., Van Dijk, M., Sattler, K., Sprokkereef, E.: Development of a European flood forecasting system, International Journal of River Basin Management, 1, 49–59, 2003.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge–Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteor. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

Dutra, E., Balsamo, G., Viterbo, P., Miranda, P. M. A., Beljaars, A., Schar, C., and Elder, K.: An improved snow scheme for the ECMWF land surface model: Description and offline validation, J. Hydrometeorol., 11, 899–916, 2010.

European Commission: Council Regulation (EC) No 2012/2002 of 11 November 2002 establishing the European Union Solidarity Fund, 2002.

de Groeve, T.: Flood monitoring and mapping using passive microwave remote sensing in Namibia, Geomatics, Natural Hazards and Risk, 1, 19–35, 2010.

Guha-Sapir, D., Vos, F., Below, R., and Ponserre, S.: Annual Disaster Statistical Review 2011: The Numbers and Trends, CRED, Brussels, 2012.

Hagen, E. and Lu, X.: Let us create flood hazard maps for developing countries, Nat. Hazards, 58, 841–843, doi:10.1007/s11069-011-9750-7, 2011.

He, Y., Wetterhall, F., Bao, H., Cloke, H., Li, Z., Pappenberger, F., Hu, Y., Manful, D., and Huang, Y.: Ensemble forecasting using TIGGE for the July–September 2008 floods in the Upper Huai catchment: a case study, Atmos. Sci. Lett., 11, 132–138, doi:10.1002/asl.270, 2010.

Hersbach, H.: Decomposition of the continuous ranked probability score for ensemble prediction systems, Weather Forecast., 15, 559–570, 2000.

Hopson, T. M. and Webster, P. J.: A 1–10-Day Ensemble Forecasting Scheme for the Major River Basins of Bangladesh: Forecasting Severe Floods of 2003–07, J. Hydrometeor., 11, 618–641, doi:10.1175/2009JHM1006.1, 2010.

Huffman, G. J., Adler, R. F., Bolvin, D. T., and Gu, G.: Improving the global precipitation record: GPCP version 2.1, Geophys. Res. Lett., 36, L17808, doi:10.1029/2009GL040000, 2009.

IFRC: The Red Cross Red Crescent approach to disaster and crisis management, Position paper, Geneva, 2011.

Lehner, B., Verdin, K., and Jarvis, A.: New Global Hydrography Derived From Spaceborne Elevation Data, Eos Trans. AGU, 89, 93–94, doi:10.1029/2008EO100001, 2008.

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., and Merchant, J. W.: Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data, Int. J. Remote Sens., 21, 1303–1330, 2000.

Marzban, C.: The ROC curve and the area under it as performance measures, Weather Forecast., 19, 1106–1114, 2004.

Miller, M., Buizza, R., Haseler, J., Hortal, M., Janssen, P., and Untch, A.: Increased resolution in the ECMWF deterministic and ensemble prediction systems, ECMWF Newsletter, 124, 10–16, 2010.

Paiva, R. C. D., Collischonn, W., Bonnet, M. P., and de Gonçalves, L. G. G.: On the sources of hydrological prediction uncertainty in the Amazon, Hydrol. Earth Syst. Sci., 16, 3127–3137, doi:10.5194/hess-16-3127-2012, 2012.

Pappenberger, F., Cloke, H. L., Balsamo, G., Ngo-Duc, T., and Oki, T.: Global runoff routing with the hydrological component of the ECMWF NWP system, Int. J. Climatol., 30, 2155–2174, 2010a.

Pappenberger, F., Thielen, J., and Del Medico, M.: The impact of weather forecast improvements on large scale hydrology: analysing a decade of forecasts of the European Flood Alert System, Hydrol. Process., 25, 1091–1113, doi:10.1002/hyp.7772, 2010b.

Pappenberger, F., Dutra, E., Wetterhall, F., and Cloke, H.: Deriving global flood hazard maps of fluvial floods through a physical model cascade, Hydrol. Earth Syst. Sci. Discuss., 9, 6615–6647, doi:10.5194/hessd-9-6615-2012, 2012.

Prinos, P., Kortenhaus, A., Swerpel, B., and Jiménez, J. A.: Review of Flood Hazard Mapping, http://www.floodsite.net/html/partner_area/project_docs/T03_07_01_Review_Hazard_Mapping_V4_3_P01.pdf, 2008.





Proud, S. R., Fensholt, R., Rasmussen, L. V., and Sandholt, I.: Rapid response flood detection using the MSG geostationary satellite, Int. J. Appl. Earth Obs., 13, 536–544, 2011.

Rao, C. X. and Maurer, E. P.: A simplified model for predicting daily transmission losses in a stream channel, J. Am. Water Resour. As., 32, 1139–1145, 1996.

Sperna Weiland, F. C., van Beek, L. P. H., Kwadijk, J. C. J., and Bierkens, M. F. P.: The ability of a GCM-forced hydrological model to reproduce global discharge variability, Hydrol. Earth Syst. Sci., 14, 1595–1621, doi:10.5194/hess-14-1595-2010, 2010.

Thielen, J., Bartholmes, J., Ramos, M.-H., and de Roo, A.: The European Flood Alert System – Part 1: Concept and development, Hydrol. Earth Syst. Sci., 13, 125–140, doi:10.5194/hess-13-125-2009, 2009a.

Thielen, J., Bogner, K., Pappenberger, F., Kalas, M., del Medico, M., and de Roo, A.: Monthly-, medium-, and short-range flood warning: Testing the limits of predictability, Meteorol. Appl., 16, 77–90, 2009b.

Thiemig, V., Pappenberger, F., Thielen, J., Gadain, H., de Roo, A., Bodis, K., Del Medico, M., and Muthusi, F.: Ensemble flood forecasting in Africa: A feasibility study in the Juba-Shabelle river basin, Atmos. Sci. Lett., 11, 123–131, 2010.

Trenberth, K. E., Dai, A., Rasmussen, R. M., and Parsons, D. B.: The changing character of precipitation, B. Am. Meteorol. Soc., 84, 1161, 1205–1217, 2003.

UN/ISDR: Living with risk: A global review of disaster reduction initiatives, ISDR Secretariat, Geneva, Switzerland, 2002.

van der Knijff, J. M., Younis, J., and de Roo, A. P. J.: LISFLOOD: A GIS-based distributed model for river basin scale water balance and flood simulation, Int. J. Geogr. Inf. Sci., 24, 189–212, 2010.

Voisin, N., Schaake, J. C., and Lettenmaier, D. P.: Calibration and Downscaling Methods for Quantitative Ensemble Precipitation Forecasts, Weather Forecast., 25, 1603–1627, doi:10.1175/2010WAF2222367.1, 2010.

Voisin, N., Pappenberger, F., Lettenmaier, D. P., Buizza, R., and Schaake, J. C.: Application of a Medium-Range Global Hydrologic Probabilistic Forecast Scheme to the Ohio River Basin, Weather Forecast., 26, 425–446, doi:10.1175/WAF-D-10-05032.1, 2011.

Wang, J., Hong, Y., Li, L., Gourley, J. J., Khan, S. I., Yilmaz, K. K., Adler, R. F., Policelli, F. S., Habib, S., Irwn, D., Limaye, A. S., Korme, T., Okello, L.: The coupled routing and excess storage (CREST) distributed hydrological model, Hydrolog. Sci. J., 56, 84–98, doi:10.1080/02626667.2010.543087, 2011.

Westerhoff, R. S., Kleuskens, M. P. H., Winsemius, H. C., Huizinga, J. H., and Brakenridge, G. H.: Automated and Systematic Water mapping in a Near-real-time Global Flood Observatory based on SAR data, Hydrol. Earth Syst. Sci., in review, 2012.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences: An Introduction, electronic version., Elsevier, San Diego, CA, http://cdsweb.cern.ch/record/992087, 2006.

Winsemius, H. C., Van Beek, L. P. H., Jongman, B., Ward, P. J., and Bouwman, A.: A framework for global river flood risk assessments, Hydrol. Earth Syst. Sci. Discuss, 9, 9611–9659, doi:10.5194/hessd-9-9611-2012, 2012.

Wu, H., Adler, R. F., Hong, Y., Tian, Y., and Policelli, F.: Evaluation of Global Flood Detection Using Satellite-Based Rainfall and a Hydrologic Model, J. Hydrometeorol., 13, 1268–1284, doi:10.1175/JHM-D-11-087.1, 2012.





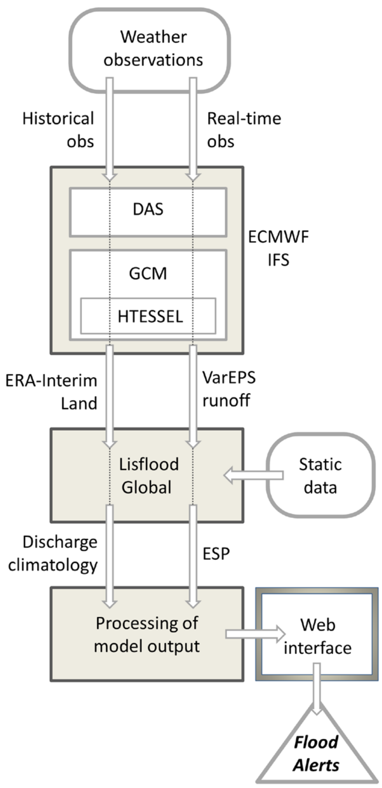


**Fig. 1.** Overview of the GloFAS structure.



Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

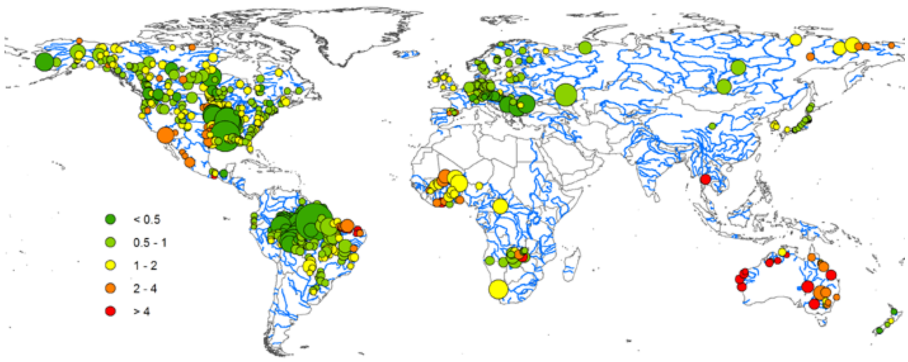


**Fig. 2.** Coefficient of variation of the estimation residuals for the 620 stations considered. Circle size is proportional to the upstream area of the river station.





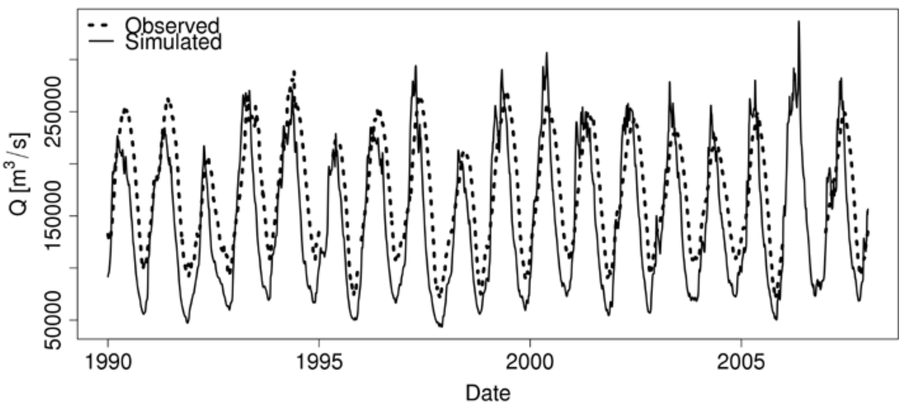


**Fig. 3.** Comparison between observed and simulated daily average discharge in the Amazon River at Obidos, linigrafo, Brazil.

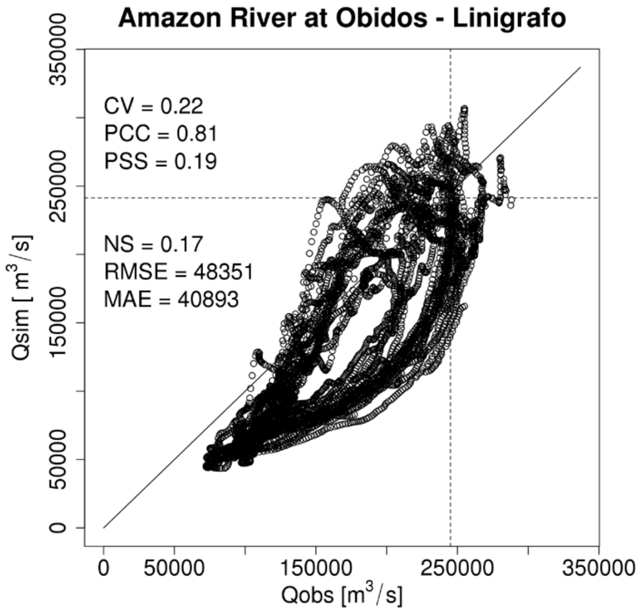


**Fig. 4.** Scatter plot of observed and simulated daily average discharge (1990–2007) in the Amazon River at Obidos, linigrafo, Brazil. 90th percentiles used for threshold exceedance analysis are shown with dashed lines while skill scores are shown on the left side.





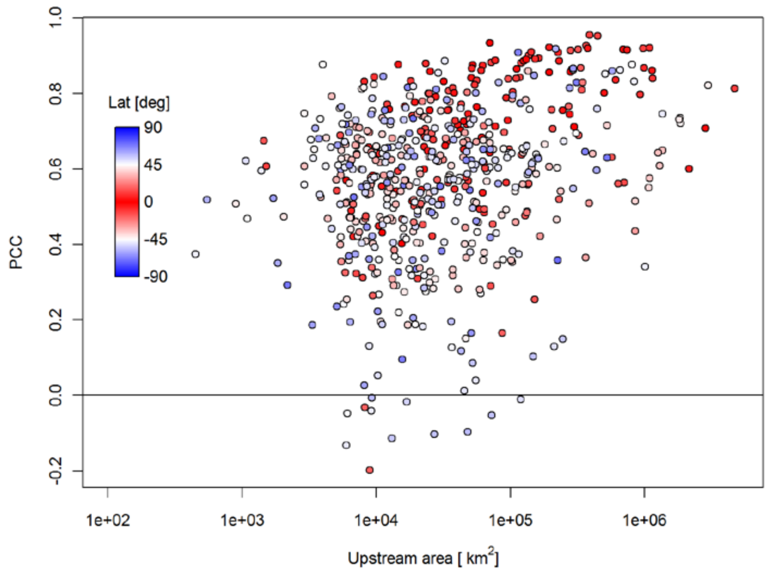


**Fig. 5.** Pearson correlation coefficient of simulated versus observed discharge for the 620 stations considered plotted against the corresponding upstream area. Circle color depends on the latitude of each river station.

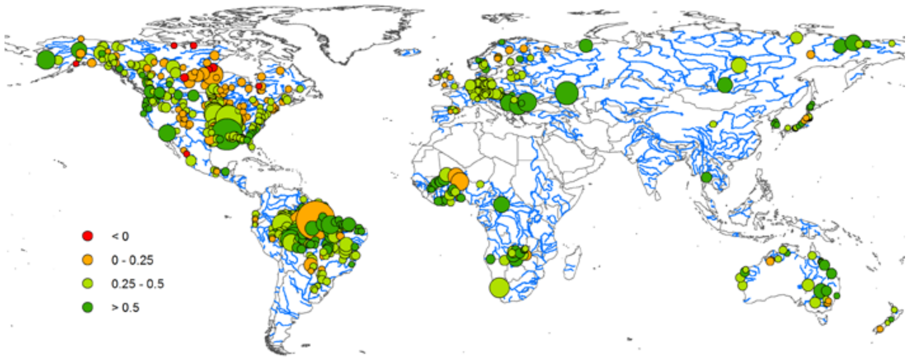


**Fig. 6.** Peirce's skill score of simulated versus observed discharge for the 620 stations considered. Circle size is proportional to the upstream area of the river station.



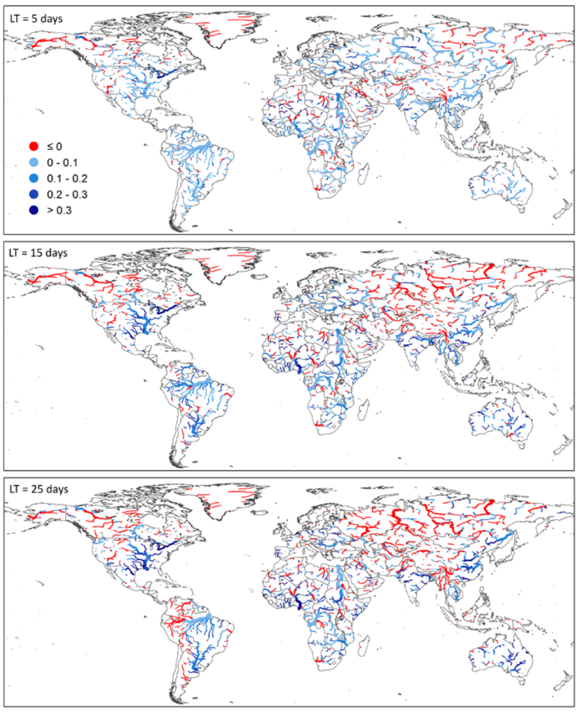


**Fig. 7.** CRPSS maps of ESP for 2009–2010 against simulated corrected discharge climatology. Panels refer to lead time of 5, 15, and 25 days (top to bottom).





Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

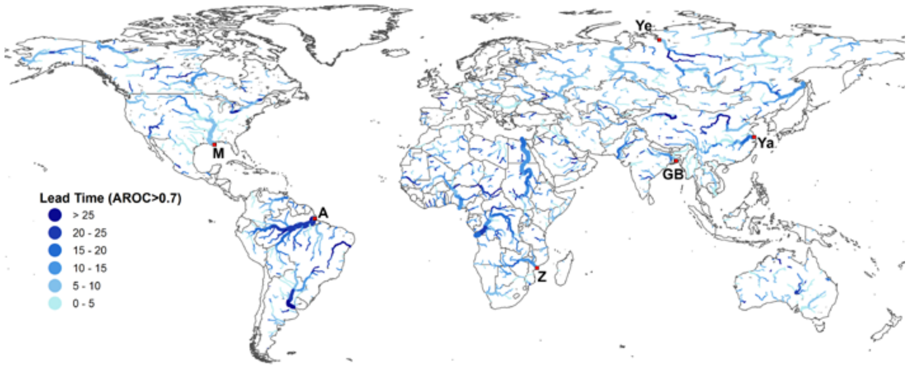


**Fig. 8.** Forecast lead time, in days, for which ESP are skilful (AROC > 0.7).





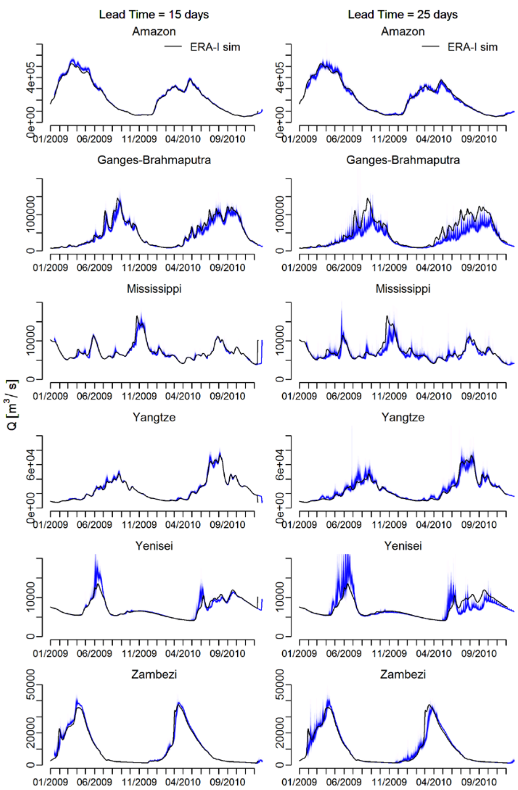


**Fig. 9.** ESP (blue shades) and corrected discharge climatology (ERA-I sim) at the outlet of six major river basins (see red markers in Fig. 7), for lead time of 15 (left column) and 25 days (right column).

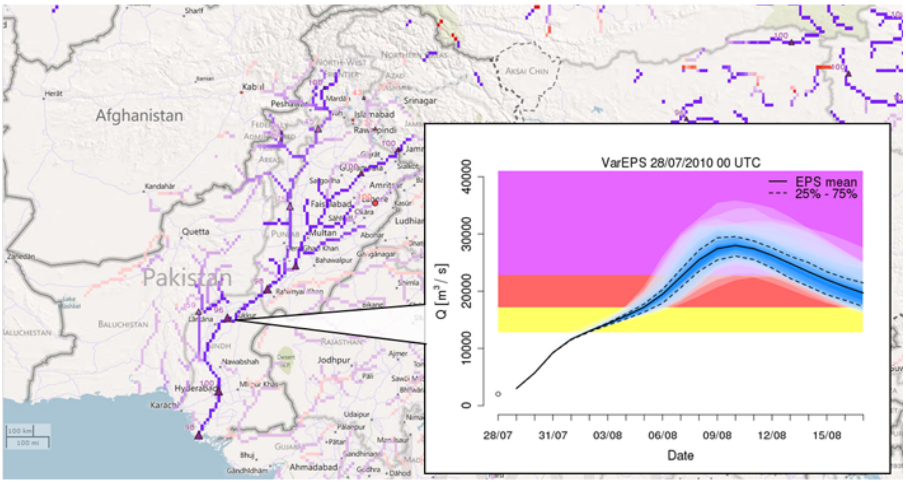


**Fig. 10.** 20-day ESP on 28 July 2010 for a dynamic reporting point in the Indus River near Sukkur, in Pakistan. The probability of severe threshold exceedance is shown with purple shadings.





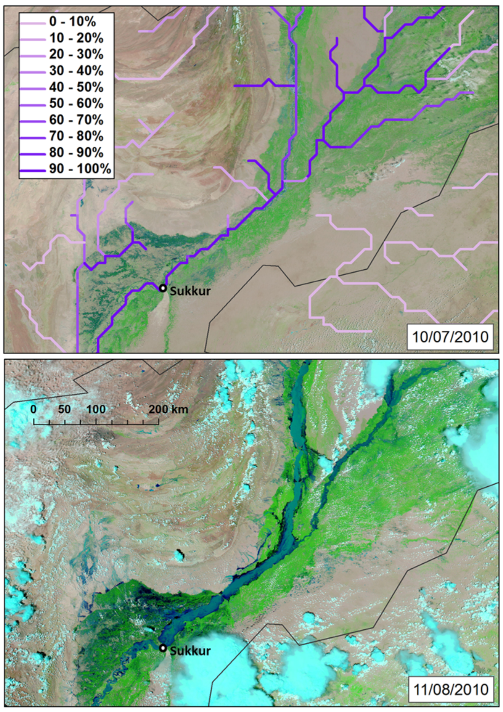


**Fig. 11.** Satellite images of the Indus River on 10 July 2010 (top) and on 11 August 2010 (bottom). Top panel also shows, with purple shadings, the maximum probability of exceeding the severe threshold in a 20-day forecast range (forecast on 28 July 2010).