



The extreme runoff index for flood early warning in Europe

L. Alfieri^{1,2}, F. Pappenberger¹, and F. Wetterhall¹

¹European Centre for Medium-Range Weather Forecasts, Reading, UK

²European Commission – Joint Research Centre, Ispra, Italy

Correspondence to: L. Alfieri (lorenzo.alfieri@jrc.ec.europa.eu)

Received: 21 November 2013 – Published in Nat. Hazards Earth Syst. Sci. Discuss.: 16 December 2013

Revised: 3 April 2014 – Accepted: 30 April 2014 – Published: 17 June 2014

Abstract. Systems for the early detection of floods over continental and global domains have a key role in providing a quick overview of areas at risk, raise the awareness and prompt higher detail analyses as the events approach. However, the reliability of these systems is prone to spatial inhomogeneity, depending on the quality of the underlying input data and local calibration.

This work proposes a simple approach for flood early warning based on ensemble numerical predictions of surface runoff provided by weather forecasting centers. The system is based on a novel indicator, referred to as an extreme runoff index (ERI), which is calculated from the input data through a statistical analysis. It is designed for use in large or poorly gauged domains, as no local knowledge or in situ observations are needed for its setup. Daily runs over 32 months are evaluated against calibrated hydrological simulations for all of Europe. Results show skillful flood early warning capabilities up to a 10-day lead time. A dedicated analysis is performed to investigate the optimal timing of forecasts to maximize the detection of extreme events. A case study for the central European floods of June 2013 is presented and forecasts are compared to the output of a hydro-meteorological ensemble model.

Pappenberger, 2009), of uncertainty analyses (Renard et al., 2010; Zappa et al., 2011), and the considerable research work devoted to improving the conveyance of probabilistic information to the end users (e.g., Demeritt et al., 2013; Pappenberger et al., 2013).

Most flood early warning systems operate at national level and require a wealth of input data and local information. Data assimilation and post-processing techniques are used to reduce the predictive uncertainty at river stations where observed water levels or discharges are collected (see van Andel et al., 2013). On the other hand, the large data requirement limits the current implementation of early warning systems at continental scale to just a few cases, the European Flood Awareness System (EFAS, see e.g., Thielen et al., 2009) being a prominent example for Europe. In poorly gauged areas, a simplified option to monitor and forecast floods is by linking them to extreme rainfall occurrences (e.g., Lalaurette, 2003; Hurford et al., 2012; Ahn and Il Choi, 2013). This assumption is widely accepted for surface water flooding events and flash floods due to short and intense rainfall events in small-size catchments. However, in larger river basins, other hydrological processes considerably influence the runoff dynamics and cannot be neglected in the early detection of flood events. The Flash Flood Guidance (FFG, see e.g., Ntelekos et al., 2006) was designed to provide a simple approach for early detection of flash floods in poorly gauged catchments, by including the effect of soil moisture conditions. Its success is demonstrated by its widespread application (see Gourley et al., 2012). A number of similar methods based on rainfall and soil moisture (e.g., Norbiato et al., 2009; Javelle et al., 2010; Van Steenberg and Willems, 2013) or on runoff (Raynaud et al., 2014) threshold exceedances have been proposed in recent years. Many of these supported the findings that simplified approaches for flood

1 Introduction

The impact on society from river floods and flash floods has steadily increased over the past decades at a global scale (CRED, 2013). Probabilistic approaches to tackling the issue of flood forecasting and early warning are becoming common practice in operational hydro-meteorological applications. Such a transition has been fostered by the increased availability of ensemble weather predictions (see Cloke and

early warning often provide as accurate results as those of physically based models, particularly when transferred to ungauged river basins.

Alfieri et al. (2011) proposed the European Precipitation Index based on simulated Climatology (EPIC) to monitor the European domain for upcoming severe storms possibly leading to flash floods. The main assumption of the EPIC is that statistically, extreme cumulated precipitation on small-size catchments is a good predictor for flash floods, independent from other hydrological processes taking place in the real world. A flash flood warning system based on EPIC is currently run in the EFAS system, which uses a probabilistic approach based on COSMO-LEPS (Marsigli et al., 2005) ensemble numerical weather prediction (NWP). The system has proved to be successful in spotting a number of flash floods across Europe (Alfieri and Thielen, 2012; see recent results in <http://www.efas.eu/efas-bulletins.html>), showing its complementarity to the hydro-meteorological forecasts run within EFAS for larger river basins.

The aim of this work is to test the feasibility and the performance of a warning system based on a concept similar to EPIC, in order to predict extreme streamflow events inducing river floods in a wide range of basin sizes. Such a system is based on the hereby defined extreme runoff index (ERI), which is calculated on forecasts of surface runoff, a variable produced by the land surface scheme of several operational weather prediction models. The basic idea of this approach is to use extreme cumulated surface runoff at the basin scale as predictor for flood occurrences. Different from the EPIC, the ERI includes hydrological processes such as snowmelt, evapotranspiration and the effect of soil moisture, among others, so all types of weather-driven floods can be detected by such an approach. The performance of ERI in flood early warning in Europe is assessed over about 32 months of daily simulation. Results are discussed and complemented by a case study for the central European flood that occurred in June 2013, where the ERI and EFAS hydro-meteorological forecasts are compared to each other.

2 Data and methods

2.1 Input data

2.1.1 Operational ensemble forecasts

Input data used to run ERI consist of surface runoff (sro) forecasts taken from the output of a NWP model. In the current setup, data used are taken from the integrated forecast system (IFS, Miller et al., 2010) of the European Centre for Medium-Range Weather Forecasts (ECMWF). Among the models run within the IFS, forecasts from a 51-member ensemble NWP are used, referred to as ENS. ENS is currently set up at global scale on a Gaussian reduced grid of T639 spectral resolution, corresponding to about 32 km horizontal

resolution, with a forecast lead time (LT) up to 10 days and a time step of 3–6 h, depending on the lead time. After day 10, the model run is extended up to day 15 (day 32 twice per week) at a coarser horizontal resolution of about 65 km. The surface runoff is an output variable of HTESSEL (see e.g., Balsamo et al., 2011), the ECMWF land surface scheme, which is coupled in the operational model runs with the atmospheric circulation model. For this work, 10-day surface runoff forecasts are extracted from daily runs at 00:00 UTC, for a time window of 2 years and 8 months starting on 1 December 2010, comprising a total of 988 ENS forecasts. This is the longest period that could be simulated with the proposed approach, as before December 2010 surface and sub-surface runoff were computed as a single cumulated variable (i.e., runoff) by the land surface scheme.

2.1.2 Reference climatology

To estimate the extremity of the forecast surface runoff, climatological values are needed for statistical comparison, consisting of a long data series of surface runoff for the same computation domain. At ECMWF, every Thursday the IFS is rerun in hindcast mode for the same day of the previous 20 years, using the latest operational model version of ENS. On such a basis, a 20-year climatology of surface runoff was constructed by taking the unperturbed run of ENS of the latest $N = 20$ years from the hindcast data set, taking one forecasts run per week. The first 7 days of each forecast surface runoff were extracted and merged into a continuous time series at each grid point of the domain, following the approach described by Alfieri and Thielen (2012).

2.2 The extreme runoff index

The extreme runoff index (ERI) is defined as

$$\text{ERI}(t) = \max_{\forall di} \left(\frac{\text{Usro}(di, t)}{\frac{1}{N} \sum_{yi=1}^N \max(\text{Usro}(di))_{yi}} \right);$$

$$0.6t_C \leq di \leq 1.2t_C, \quad (1)$$

where

$$\text{Usro}(di, t) = \sum_{t-di}^t \sum_A \text{sro} \quad (2)$$

and t_C is the basin time of concentration.

$\text{Usro}(di)$ is the upstream cumulated surface runoff at each grid point, that is, the double summation of surface runoff over the upstream area (A) and over a certain duration di preceding the considered time t .

Different from the EPIC index, the ERI is designed to forecast floods in a wide range of basin sizes. The upper limit is reached when the river routing, not considered within this

approach, starts to have a substantial role in the timing and dampening of the flood wave. To this end, di is set proportionally to the basin time of concentration (t_C), which was estimated for every grid point of the river network using the empirical formula by Giandotti (1934) based on geomorphologic parameters. In practice, the basic assumption is that flood events can occur at a generic point in the river network when the cumulated surface runoff is extreme (in statistical terms) over durations which are of the same magnitude of its lag time and its time of concentration. Such choice is in agreement with the rational method (see e.g., Chow et al., 1988) and with the findings from Fiorentino et al. (1987) and from Viglione and Blöschl (2009). In addition, Eq. (1) assumes the relation between basin lag time (t_L) and time of concentration (t_C) used by the Natural Resources Conservation Service (NRCS), $t_L = 0.6t_C$ (e.g., McCuen, 1989).

It is worth noting that a formulation similar to Eq. (1) was proposed by Raynaud et al. (2014) to forecast flash floods in Europe, where di is set to the fixed values {6, 12, 24h} as in EPIC, and the upstream runoff is estimated by multiplying the upstream precipitation by a variable runoff coefficient derived from different components of a background hydrological simulations.

The ERI is a dimensionless index aimed at flood early warning and comparable with EPIC or with normalized discharge, as shown by Alfieri and Thielen (2012). In operational forecasts the procedure was adapted to an ensemble approach, where the return period of ERI is shown for each ensemble member. The procedure is described in detail in the following:

1. For computational reasons, a fixed number of durations di is chosen for cumulating upstream surface runoff; $di \in \{6, 12, 18, 24, 30, 36, 48, 60, 72, 96, 120, 144\}$ h.
2. For each duration in di , a Gumbel extreme value distribution is fitted to the annual maxima of $Usro(di)$ derived from the 20-year climatology, using the method of L moments (see Hosking, 1990), deriving for each grid point a scale parameter $\alpha(di)$ and a location parameter $\xi(di)$ of the obtained distribution.
3. In operational 10-day ERI forecasts, each point of the river network is assigned a subset of durations $dj \in di$, among those that fulfill the criterion $0.6t_C \leq dj \leq 1.2t_C$. For instance, if $t_C = 60$ h, $dj = \{36, 48, 60, 72\}$ h.
4. For each duration in dj , the return period of $Usro(dj, t)$ is calculated from the formulation of the Gumbel distribution:

$$T(t, dj) = \frac{1}{1 - \exp\left(-\exp\left(-\frac{Usro(t, dj) - \xi(dj)}{\alpha(dj)}\right)\right)}. \quad (3)$$

5. For each 6 h time step t within the 10-day forecast, the maximum T among the selected dj is selected.

6. Points 4 and 5 are iterated over all 51 ensemble members.
7. For each time step, the probability of exceeding a warning threshold, corresponding to a selected return period, is calculated by summing the number of members above the threshold and dividing by the ensemble size.

A schematic view of the calculation of ERI forecasts is shown in Fig. 1. Note that when $t < \min(dj)$, the duration of the accumulation includes time steps which are before the start of the forecast. In such cases, the $Usro$ is calculated by filling previous time steps with the most recent 24 h forecasts of each antecedent day.

The above-described approach was set up on the same computational grid of EFAS, which covers all of Europe with a $5 \text{ km} \times 5 \text{ km}$ grid, including 29 349 grid points in the modeled river network, with upstream area larger than 1000 km^2 . Also, the largest rivers with $t_C > 144$ h are not considered in the calculation. This mostly occurs in large river basins with upstream areas larger than $100\,000 \text{ km}^2$.

3 Evaluation of ERI forecasts

The approach described in Sect. 2.2 was set up for operational daily run in hindcast mode over 2 years and 8 months starting on 1 December 2010, using the input data described in Sect. 2.1. The evaluation of results of the ERI is performed through a threefold approach, focused on (1) the evaluation of performance in detecting extreme events, (2) a statistical description of alerts produced by the ERI, and (3) a case study which compares ERI and EFAS flood forecasts with a reference hydrological simulation. They are described in the following sub-sections.

3.1 Performance in threshold exceedance prediction

The performance of ERI in flood early warning is tested by comparing ERI ensemble forecasts with the EFAS water balance (WB), which is a hydrological simulation of the whole European domain, run using spatially interpolated meteorological observations to obtain a continuous field. A description of the EFAS-WB and its underlying data can be found in Alfieri et al. (2013).

The suitability of the ERI for potential use in flood early warning is tested by comparing its skill in predicting discharges above threshold in Europe. The chosen threshold is the 2-year return period at each grid point of the river network, which is a suitable tradeoff between a relatively extreme value possibly leading to flooding, and having some simulated threshold exceedances in the simulation period. As the ERI is based on an ensemble system, forecast exceedances (Pfc) are expressed as probabilities, while simulated exceedances (Psim) for validation are taken from the EFAS-WB and are expressed as dichotomous values {0, 1}.

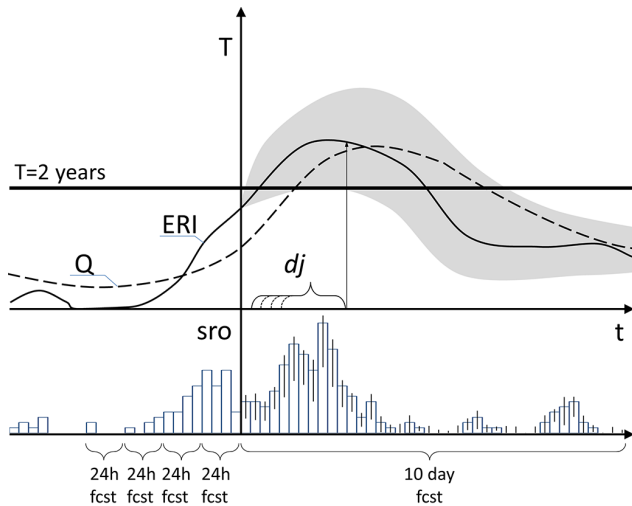


Figure 1. Schematic view of ERI forecasts and comparison with river discharge (Q). Sample input sro is shown at the bottom.

The verification is based on the Brier skill score (BSS, see e.g., Wilks, 2006):

$$\text{BSS} = 1 - \frac{\text{BS}}{\text{BS}_{\text{ref}}}; \quad (4)$$

where

$$\text{BS} = \frac{1}{n} \sum_{k=1}^n (\text{Psim}_k - \text{Pfc}_k)^2. \quad (5)$$

Equation (5) defines the Brier score, which is calculated on the $n = 988$ time steps, for each daily lead time. Since EFAS-WB is calculated at 06:00 UTC, while ERI forecasts are run at 00:00 UTC, lead times used in the verification range from 6 to 222 h (i.e., 9 days and 6 h). BS_{ref} is a reference BS calculated by assuming the climatological probability of exceeding the 2-year threshold, calculated on a 21-year time series of the EFAS-WB.

In addition, the same data set of forecast and simulated threshold exceedances was used to calculate the probability of detection (POD) and the false alarm rate (FAR) of the ERI by choosing five different probability thresholds (i.e., $pt = 10, 30, 50, 70, 90\%$) for ensemble forecasts above the 2-year return period.

3.2 Statistics of alert points

The second step of the evaluation approach was to collect and analyze some statistics of all grid points exceeding a critical flood threshold among forecast runs in the 32-month time span. By setting a probability threshold of 15 % of exceeding the 2-year return period, about 38 100 grid points with an ERI forecast above threshold were detected. Each point is characterized by the maximum probability of exceeding the 2, 5, and 20-year return period, and the corresponding time horizon to the forecast peak.

In addition, the system is designed to produce output images in a similar fashion as in EFAS, by producing for each forecast:

- Three maps of maximum probability of exceeding the three warning thresholds of 2, 5, and 20-year return period, over the forecast range.
- Point forecasts showing the ensemble prediction of ERI over the forecast range for selected reporting points. Reporting points are selected to give an adequate coverage of the areas at risk for every forecast. In detail, points are chosen among those (1) with a probability larger than 60 % of exceeding the 2-year return period, (2) with a minimum upstream area of 1000 km², and (3) by keeping a minimum distance of 100 km from each other, along the river network, in case of long river reaches above threshold.

Such criteria were derived iteratively to optimize the visualization of results; therefore they are independent from the evaluation approach.

3.3 Case study

Example figures of the output images are shown and commented based on a case study of the severe floods which hit a large portion of Central Europe in early June 2013. Results of the ERI are compared to the corresponding EFAS forecasts for the same event and to the simulated threshold exceedances. This will be the third step of the proposed evaluation approach.

4 Results and discussion

4.1 Performance in threshold exceedance prediction

Forecast threshold exceedances of the ERI are compared to the proxy simulations extracted from the EFAS-WB. A visual example of such comparison is shown in Fig. 2 for the Danube River in Linz, Austria, for lead times of 1, 4, 7, and 10 days. Figure 2 shows three simulated events above the 2-year threshold, though the second and third exceedances actually correspond to the same event, in June 2013. The ERI predicted the second event with high probability (i.e., $P(T > 2) = 100\%$ for $LT = 1$ and 4 days, $P(T > 2) \approx 80\%$ for $LT = 7$ days) though it missed the first event for all lead times. Some low probabilities of exceedance were also predicted in summer 2011 and 2012, though with no simulated event above threshold.

The average BSS is plotted against the forecast lead time in Fig. 3 with a dashed line. As shown in Fig. 3, the BSS is unskillful for all lead times, and approaches the zero line towards the longest forecast ranges. Indeed, such a comparison assumes that the timing of the ERI is set at the end of the last time step of Usro accumulation. Such an assumption is

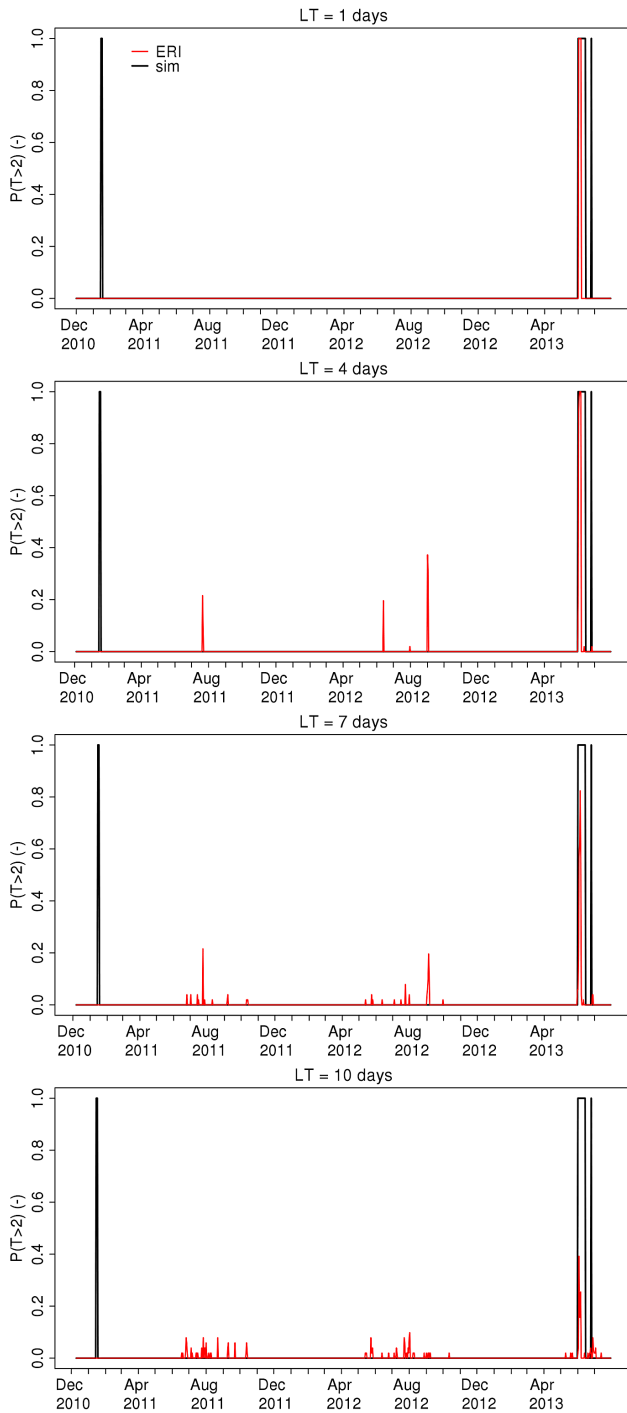


Figure 2. Forecast (ERI) and simulated (sim) exceedance of the 2-year discharge return period for the Danube River at Linz, Austria. From top to bottom, LT = 1, 4, 7, and 10 days.

plausible, to some extent, considering that (1) the durations of accumulation are constrained by $0.6t_C \leq di \leq 1.2t_C$ and (2) the basin lag time ($t_L \approx 0.6t_C$) is also defined as the time shift between the center of mass of the effective hyetograph (here comparable to surface runoff) and that of the hydro-

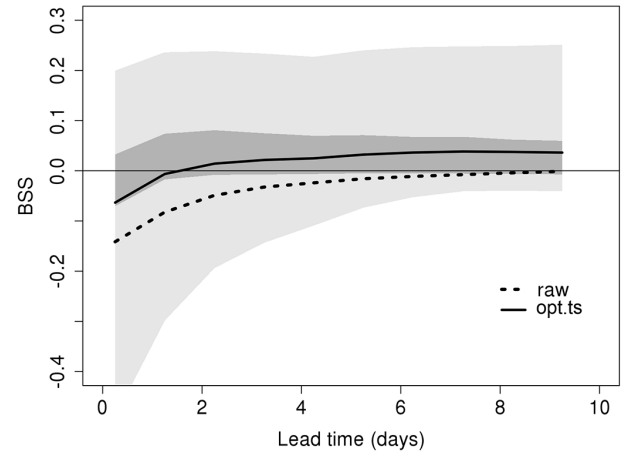


Figure 3. BSS of the ERI vs. lead time for the raw forecasts (average BSS with a dashed line) and considering the optimal time shift (average BSS with a solid line; shadings indicate the 5–95 % range, in light grey, and the 25–75 % range, in dark grey).

graph. However, the timing of events detected by the ERI cannot be defined precisely a priori, as no routing nor delay component is included in its definition. The issue of matching observed and simulated peaks in model verification has growth of interest with the spreading of hydrological forecasts in the past few years. Recent contribution to the topic was given by Zappa et al. (2013) and by Ewen (2011). The BSS was recalculated for several configurations, where the ERI was shifted in time to search for the optimal time shift to match it with the simulated threshold exceedance. Time shifts TS are tested, with 6 h spacing, in the range

$$-1 \text{ day} - 20\% \text{ LT} \leq T_S(\text{LT}) \leq 1 \text{ day} + 20\% \text{ LT}.$$

BSS derived with optimal time shifts are shown in Fig. 3 with grey shadings and a solid line. For $\text{LT} \geq 2$ days, about 75 % of BSS values are skillful compared to a climatological forecast. Figure 4a shows a map of the BSS with optimal time shift for each simulated grid point, averaged among all lead times. Figure 4b shows the corresponding average time shifts in hours. One can note that positive (i.e., skillful) BSS values are associated with positive shifts, meaning that the optimal timing of ERI predictions corresponds to a shift forward of about 0–21 h, with an increasing trend with the lead time. The average time shift among all points and lead times is 7 h, though it rises to 16 h if calculated only on grid points with positive average BSS. In 14 % of points, no exceedance of the 2-year return period was simulated in the considered time window, making the application and interpretation of the BSS more difficult. In these points the optimization of the timing of the forecasts often resulted in null or negative time shifts, due to the difficulty in matching forecast threshold exceedances with no simulated ones. Similarly, Fig. 5 displays the BSS of ERI (considering the optimal time shift) for four different lead times of 1, 4, 7,

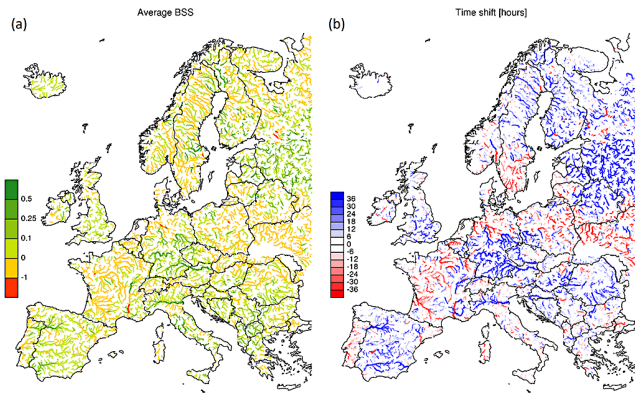


Figure 4. Average BSS among the considered lead times (a) and corresponding time shift of the ERI against the simulated threshold exceedances (b).

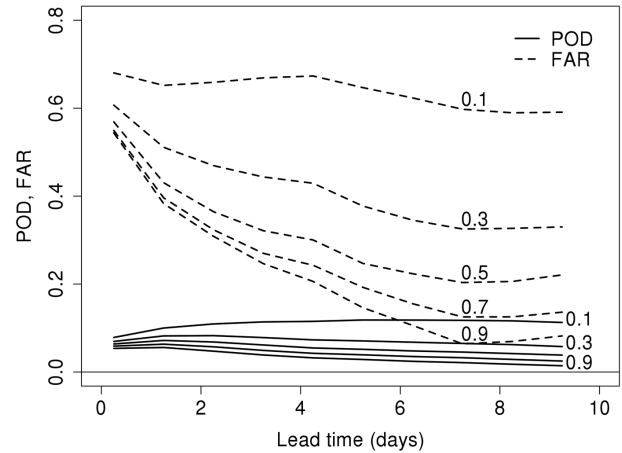


Figure 6. POD and FAR versus lead time, for different probability thresholds between 10 and 90 %.

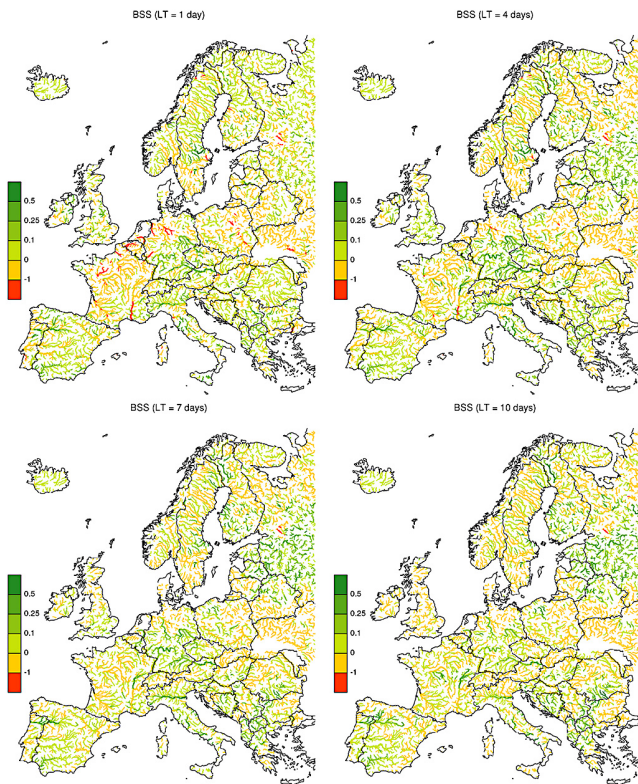


Figure 5. BSS of the ERI for lead times of 1, 4, 7, and 10 days.

and 10 days. It shows a general convergence towards zero for the BSS with the lead time. At $LT = 1$ day, the ensemble spread is narrower, thus events above threshold are either detected with high probability or completely missed. The same behavior was found by Pappenberger et al. (2010) in evaluating ensemble streamflow predictions. For increasing lead times, the ensemble spread gets wider and more events are predicted with lower probability of exceedance.

Figure 6 shows the POD and FAR calculated on the same data set of forecasts and simulations, averaged among all points and displayed as a function of the lead time. Such skill measures are based on binary outcomes of forecasts and simulations at each time step; hence ensemble forecasts of ERI were turned to dichotomous information, depending on the probability of exceeding the 2-year return period. Figure 6 shows a higher sensitivity of the FAR depending on the probability threshold (pt), due to the increasing spread of the ensemble with the lead time. The POD is rather constant with the lead time and mostly below 0.1, which suggests substantial differences between the duration of simulated discharge above threshold and forecast threshold exceedance from the ERI. Such figures are likely to underestimate the true potential in early warning of ERI, as they are calculated on each time step rather than on an event basis (not computed in this work). Indeed the ERI, as the EPIC, has the tendency to decay below warning values faster than discharge, as the approach is based on a statistical comparison of the input surface runoff and does not account for the routing of the flood wave and in turn for the correct timing of the runoff. For example, if calculated on an “event basis” and with $pt = 30\%$, the POD at the station in Fig. 2 would become between 33 and 50 % (the latter if one considers that the last two peaks above threshold are part of the same event) for all lead times. Similarly, the FAR would become $FAR_{Linz} = \{0, 50, 0, 0\} \%$ for lead times of 1, 4, 7, 10 days.

4.2 Statistics of alert points

The cumulative distribution functions (cdf) of the probability of exceeding return periods of 2, 5, and 20 years for the set of forecasts above threshold (i.e., only for the event peak) are shown in the three panels of Fig. 7. These are shown with grey shades for each forecast lead time between 6 and 240 h. Contour lines are plotted at selected quantiles. By

definition of the selection criteria (see Sect. 3.2), all points in Fig. 7a have $P(T > 2) \geq 15\%$ (see black area corresponding to quantile 0). For higher quantiles, high probabilities of exceeding the 2-year threshold are mostly detected for short lead times. It is interesting to note in Fig. 7a the diurnal cycle of the surface runoff (particularly for quantiles 0.75 and 0.9), which induces higher values for lead times corresponding to 12:00–18:00 UTC, where the influence of the snowmelt and of convective precipitation (see Bechtold et al., 2013) is more pronounced. Another peculiar feature shown in Fig. 7 is the high sharpness in forecasting extreme values. Indeed, the contour line of the highest quantile in all three panels does not show a significant trend with the lead time, indicating some forecasts reach the 100% probability of exceeding the three thresholds, even for lead times as long as 240 h. Finally, the black column in Fig. 7a for $LT = 6$ h supports the idea of under-dispersed forecasts. In practice, at the shortest lead times the ensemble spread is comparatively narrow, so that if the 15th percentile exceeds a warning threshold, the full ensemble is likely to exceed it too. For comparison, a dashed line is shown in Fig. 7b to indicate the current criterion to send flash flood alerts in EFAS, based on EPIC (i.e., $P(T > 5) \geq 60\%$). Assuming that the same criterion could be used to issue flood warnings on the basis of ERI, a subset of 2091 forecasts above threshold would be detected in the selected 32-month period. Such points, hereafter referred to “flood alerts” are shown on a map in Fig. 8, where the circle size is proportional to the maximum lead time for which the event peak was spotted. The probability density function (pdf) and the cdf of the lead time of the flood alerts are shown in Fig. 9, together with those of the corresponding upstream area (A). Figure 9 indicates that nearly 60% of flood alerts were produced for lead times up to 12 h and upstream areas smaller than 5000 km². However, 192 flood alerts are associated to a lead time of 5 or more days. Flood alerts in Figs. 8 and 9 could not be verified against observed events, as this would require the availability of observed discharge and the corresponding thresholds virtually in every European river; therefore, we limit the verification approach to the analysis shown in Sect. 4.1.

4.3 Case study – the central European floods of June 2013

Between the end of May and the beginning of June 2013, a low pressure system brought moist air from the east and the northeast of Europe, generating large rainfall accumulations in southern Germany and western Austria. In addition, orographic enhancement of precipitation on the northern side of the Alps played a prominent role. A number of rivers, mostly within the Danube, Rhine, and Elbe river basins, exceeded warning thresholds and several cities suffered from damage and service disruption caused by the floods. Further details on the flood and on the underlying atmospheric processes are described by Blöschl et al. (2013) and by Grams et al. (2014).

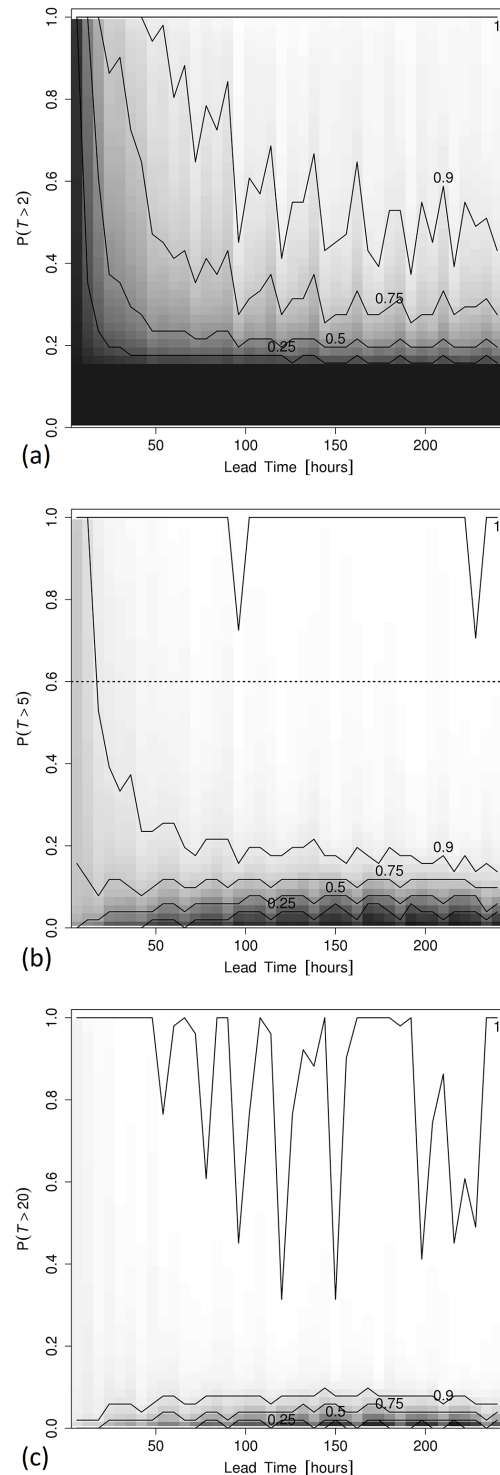


Figure 7. Cdf of the probability of the ERI exceeding the 2, 5, and 20-year return period (event peak), with contour lines at significant probability levels.

A visual comparison between ERI and EFAS forecasts is performed for this flood event and is shown in Fig. 10, in panels a and c, respectively. Results from the EFAS water

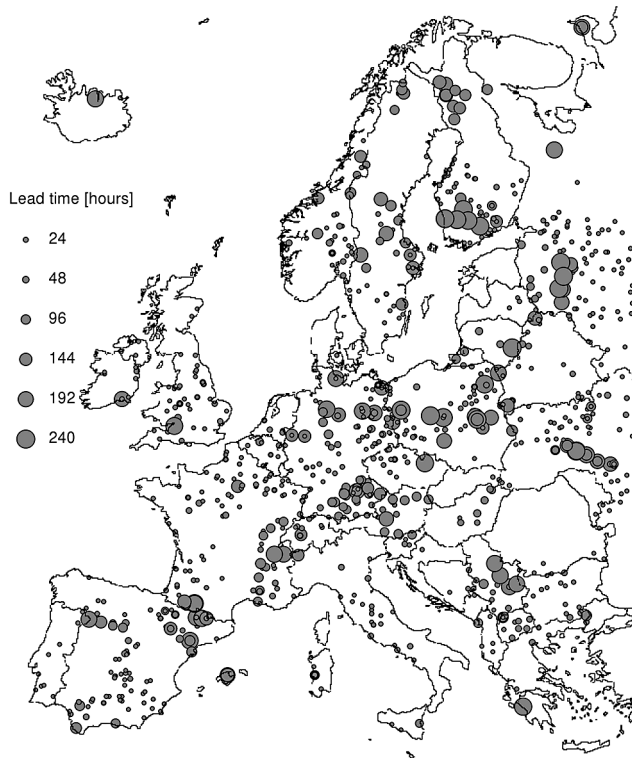


Figure 8. Location of flood alerts predicted in the 32-month period. Circle size is proportional to the lead time to the event peak.

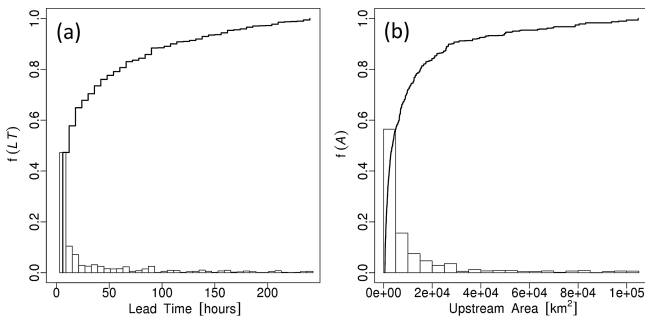


Figure 9. Pdf and cdf of lead time (a) and upstream area (b) of flood alerts detected by the ERI.

balance are shown in Fig. 10b. The three panels in Fig. 10 show the forecast and simulated exceedance of the 5-year return period for the three models. On this occasion, ERI predicted reasonably well the river reaches at risk of threshold exceedance. Results from the EFAS hydrological simulations, run with the same ECMWF ensemble model as input, produced a similar pattern but with lower magnitude, indicating maximum probabilities of threshold exceedance around 30 % (see Fig. 10c). Figure 11 shows the ensemble prediction of ERI (left) and the EFAS multi-model (right) for a point on the Danube by the city of Linz, in Austria. The comparison of the two panels indicates higher severity of the ERI, with

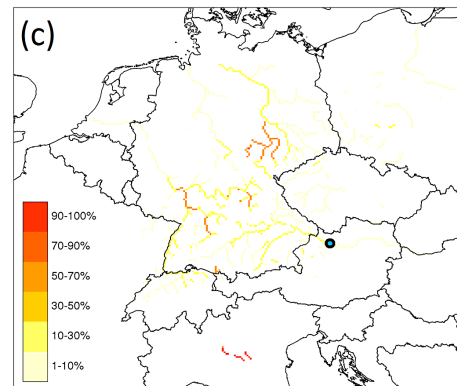
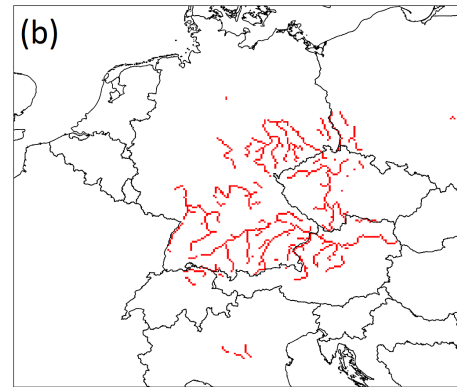
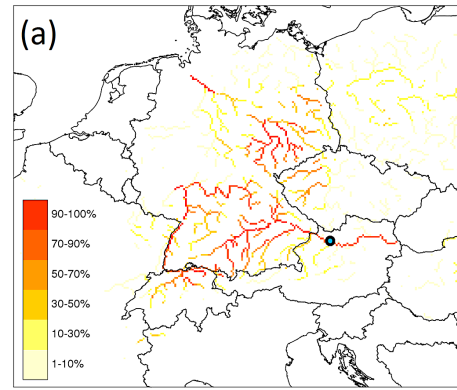


Figure 10. Maximum exceedance of the 5-year return period between 30 May 2013 and 9 June 2013 for (a) ERI forecasts, (b) EFAS-WB, and (c) EFAS forecasts. The location of Linz is shown with a blue circle.

the ensemble mean reaching the 50-year return period, while most of the EFAS ensemble lay between 2 and 5 years (yellow area) at the time of the forecast peak. Such difference stresses the potential of using a consistent reforecast data set to calculate warning threshold, as was done for the ERI. On the other hand, current EFAS thresholds are derived from statistical analysis on the EFAS-WB, which in turn is based on interpolated meteorological observations as input. Indeed,

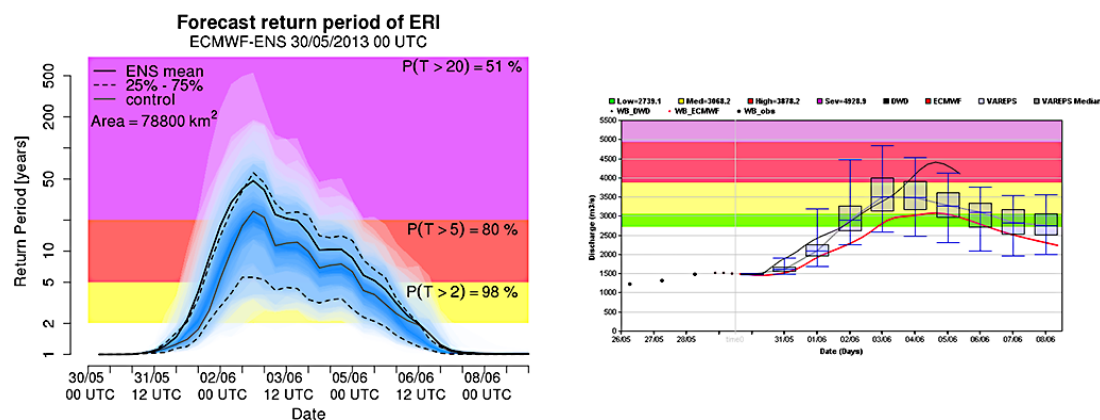


Figure 11. Comparison between 10-day ensemble forecasts of the ERI (left) and the EFAS multi-model (right) for Linz.

recent work on the evaluation of EFAS forecasts in Europe pointed out some underestimation of the forecast runoff in mountain areas such as in the Alps and the Pyrenees (Alfieri et al., 2014). Further, the event peak of the ERI is anticipated for about 12–24 h compared to the hydrological forecasts of EFAS, supporting the findings of Sect. 4.1 of the need for a positive time shift to optimize the timing of ERI forecasts.

5 Discussion and concluding remarks

In this work we present a non-parametric approach for ensemble flood early warning for a wide range of basin sizes, based exclusively on the output of a state-of-the-art global circulation model. We defined the extreme runoff index (ERI), which is designed to detect extreme accumulations of surface runoff over critical flood durations for each section of the river network. Its strength in detecting extremes is given by the use of a coherent 20-year climatology of the same input parameter (i.e., surface runoff), so that anomalous forecasts are identified and their severity quantified in statistical terms. In addition, the reforecast data set is updated in parallel with changes in the circulation model, so that warning thresholds can be recalculated and maintain their consistency with operational ensemble forecasts. The work follows and complements the positive findings of the European Precipitation Index based on simulated Climatology (EPIC, Alfieri and Thielen, 2012), currently used in the context of the European Flood Awareness System to issue flash flood warnings. The main advances of the proposed approach are:

- The ERI is based on the output of a land surface scheme of a global circulation model, thus it considers all the hydrological processes involved in the generation of surface runoff. It is an appropriate indicator to predict river floods for a wide range of conditions, including soil saturation and snowmelt-driven floods, yet preserving the capabilities in detecting floods driven by extreme precipitation over short durations.

- The range of basin sizes monitored by the ERI is increased, compared to the EPIC, thanks to a procedure that considers a variable range of durations to detect extreme events, which depends on the response time of the basin. Theoretical boundaries of the minimum basin size monitored by the ERI is related to the resolution of the input data and the consequent ability of the circulation model to represent correctly the anomaly of an extreme event, compared to climatological conditions. Following the discussions in Alfieri and Thielen (2012) and Sangati and Borga (2009), the authors recommend the use of ERI forecasts in river basins with areas larger than 1000 km², which is of the same magnitude of the grid resolution of the input data. The upper limit is less clear to define, as it is conditioned by the increasing effect of the river routing with the basin size, the timing of flood peaks in different tributaries of the same basin, the dampening of the flood wave in its travel downstream and due to floodplains, the interplay between surface, subsurface runoff and the groundwater. In the current approach, the upper limit of basin size is of the order of 10⁵ km², and it is bounded by a maximum accumulation period of surface runoff of 6 days.

In the presented setup, the ERI uses the same computation domain and grid resolution of the EFAS. Also, input forecasts are derived from the same circulation model. The advantage of this is twofold. First, ERI forecasts can be compared to those of a distributed hydrological model and reasons for potential mismatch can be investigated, to help address further improvements of both systems. Second, a background simulation of the actual river state on the same domain (i.e., EFAS-WB) was available and suitable for this work as proxy truth to verify the performance of the ERI in predicting the exceedance of discharge warning thresholds. Such a unique data set for all of Europe enabled a verification approach (1) on a large domain covering a wide range of climates and basin sizes and (2) based on assessing the system behavior for extreme events, rather than just for relatively

high flows (e.g., 90th, 95th percentile) as often found in the literature. This aims to address some of the common weaknesses in the verification of ensemble flood forecasting systems as listed by Cloke and Pappenberger (2009). Results in Sect. 4 suggest a positive skill of the ERI in flood early warning, stressing the need for longer simulation periods to achieve a consistent spatial overview in such a large domain. Indeed, some river reaches within the simulated domain had no exceedance of the 2-year threshold used for validation in the considered 32-month time window. In 81 % of grid points where ERI provided skillful forecasts (i.e., BSS > 0), extreme events were found to be shifted forward in time to optimize the timing of their detection, in comparison to the initial assumption of matching the ERI with the end of the accumulation period as in the EPIC. The implication is an average increase of the forecast lead time (e.g., in Figs. 3 and 6), all skill scores being equal.

In a first step, ERI was set up for Europe and can now be seen as a complementary tool to EPIC and EFAS hydrological simulations, particularly for those river reaches where no hydrological parameter can be calibrated due to lack of observed discharge. However, additional development work in this area could lead to two important achievements:

- The same system can be set up in any other part of the world or even at a global scale, where computer resources are available. Indeed, the dynamic input data currently used (ECMWF-ENS and the corresponding reforecast data set) are available globally and simply need to be complemented with few static maps such as drainage direction and upstream area, among others. It is a relatively simple system for flood early warning with strong potential in developing countries and in ungauged river basins, able to give a quick overview of areas at risk of extreme streamflow conditions in the coming days.
- The implementation of the ERI on higher-resolution forecasts from limited-area models is likely to bring significant advances in flash flood forecasting and early warning, especially in its ability to detect flood events where the snowmelt component and the initial soil wetness play key roles in the runoff production. This would be possible by applying land surface schemes to those models with a consistent reforecast climatology available for use (e.g., COSMO-LEPS), so that surface runoff can be derived and used as input for the ERI.

Acknowledgements. The EFAS team at the Joint Research Centre of the European Commission is gratefully acknowledged for the provision of EFAS-WB data and of ensemble streamflow forecasts over Europe used in the evaluation of ERI forecasts.

Edited by: B. Merz

Reviewed by: M. Zappa and one anonymous referee

References

- Ahn, J. H. and Il Choi, H.: A New Flood Index for Use in Evaluation of Local Flood Severity: A Case Study of Small Ungauged Catchments in Korea, *J. Am. Water Resour. Assoc.*, 49, 1–14, doi:10.1111/jawr.12025, 2013.
- Alfieri, L. and Thielen, J.: A European precipitation index for extreme rain-storm and flash flood early warning, *Meteorol. Appl.*, online first, doi:10.1002/met.1328, 2012.
- Alfieri, L., Velasco, D., and Thielen, J.: Flash flood detection through a multi-stage probabilistic warning system for heavy precipitation events, *Adv. Geosci.*, 29, 69–75, doi:10.5194/adgeo-29-69-2011, 2011.
- Alfieri, L., Salamon, P., Bianchi, A., Neal, J., Bates, P., and Feyen, L.: Advances in pan-European flood hazard mapping, *Hydrol. Process.*, online first, doi:10.1002/hyp.9947, 2013.
- Alfieri, L., Pappenberger, F., Wetterhall, F., Haiden, T., Richardson, D., and Salamon, P.: Evaluation of ensemble streamflow predictions in Europe, *J. Hydrol.*, submitted, 2014.
- 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, 2011.
- Bechtold, P., Semane, N., Lopez, P., Chaboureaud, J.-P., Beljaars, A., and Bormann, N.: Breakthrough in forecasting equilibrium and non-equilibrium convection, *ECMWF Newsletter*, 136, 15–22, 2013.
- Blöschl, G., Nester, T., Komma, J., Parajka, J., and Perdigão, R. A. P.: The June 2013 flood in the Upper Danube Basin, and comparisons with the 2002, 1954 and 1899 floods, *Hydrol. Earth Syst. Sci.*, 17, 5197–5212, doi:10.5194/hess-17-5197-2013, 2013.
- 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 – EM-DAT: The OFDA/CRED International Disaster Database, Université Catholique de Louvain, Brussels, Belgium, available at: www.emdat.be, last access: 6 June 2014, 2013.
- Demeritt, D., Nobert, S., Cloke, H. L., and Pappenberger, F.: The European Flood Alert System and the communication, perception, and use of ensemble predictions for operational flood risk management, *Hydrol. Process.*, 27, 147–157, doi:10.1002/hyp.9419, 2013.
- European Flood Awareness System – Bimonthly Bulletin, available at: <http://www.efas.eu/efas-bulletins.html>, last access: 6 June 2014.
- Ewen, J.: Hydrograph matching method for measuring model performance, *J. Hydrol.*, 408, 178–187, doi:10.1016/j.jhydrol.2011.07.038, 2011.
- Fiorentino, M., Rossi, F., and Villani, P.: Effect of the basin geomorphoclimatic characteristics on the mean annual flood reduction curve, in: *Proc. Of 18th Annual Pittsburgh modelling and Simulation Conference*, Pittsburgh, 5, 1777–1784, 1987.
- Giandotti, M.: Previsione delle piene e delle magre dei corsi d'acqua, *Memorie e studi idrografici*, 8, 107 pp., 1934 (in Italian).
- Gourley, J. J., Erlingis, J. M., Hong, Y., and Wells, E. B.: Evaluation of Tools Used for Monitoring and Forecasting Flash

- Floods in the United States, *Weather Forecast.*, 27, 158–173, doi:10.1175/WAF-D-10-05043.1, 2012.
- Grams, C. M., Binder, H., Pfahl, S., Piaget, N., and Wernli, H.: Atmospheric processes triggering the Central European floods in June 2013, *Nat. Hazards Earth Syst. Sci. Discuss.*, 2, 427–458, doi:10.5194/nhessd-2-427-2014, 2014.
- Hosking, J. R. M.: L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics, *J. Roy. Stat. Soc. B Met.*, 52, 105–124, 1990.
- Hurfurd, A., Priest, S., Parker, D., and Lumbroso, D.: The effectiveness of Extreme Rainfall Alerts in predicting surface water flooding in England and Wales, *Int. J. Climatol.*, 32, 1768–1774, doi:10.1002/joc.2391, 2012.
- Javelle, P., Fouchier, C., Arnaud, P., and Lavabre, J.: Flash flood warning at ungauged locations using radar rainfall and antecedent soil moisture estimations, *J. Hydrol.*, 394, 267–274, 2010.
- Lalurette, F.: Early detection of abnormal weather conditions using probabilistic extreme forecast index, *Q. J. Roy. Meteor. Soc.*, 129, 3037–3057, 2003.
- Marsigli, C., Boccanera, F., Montani, A., and Paccagnella, T.: The COSMO-LEPS mesoscale ensemble system: validation of the methodology and verification, *Nonlin. Processes Geophys.*, 12, 527–536, doi:10.5194/npg-12-527-2005, 2005.
- McCuen, R. H.: *Hydrologic analysis and design*, Prentice Hall PTR., 1989.
- 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.
- Norbiato, D., Borga, M., and Dinale, R.: Flash flood warning in ungauged basins by use of the flash flood guidance and model-based runoff thresholds, *Meteorol. Appl.*, 16, 65–75, doi:10.1002/met.126, 2009.
- Ntelekos, A. A., Georgakakos, K. P., and Krajewski, W. F.: On the uncertainties of flash flood guidance: Toward probabilistic forecasting of flash floods, *J. Hydrometeorol.*, 7, 896–915, 2006.
- 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, 2010.
- Pappenberger, F., Stephens, E., Thielen, J., Salamon, P., Demeritt, D., van Andel, S. J., Wetterhall, F., and Alfieri, L.: Visualising probabilistic flood forecast information: expert preferences and perceptions of best practice in uncertainty communication, *Hydrol. Process.*, 27, 132–146, 2013.
- Raynaud, D., Thielen, J., Salamon, P., Burek, P., Anquetin, S., and Alfieri, L.: A dynamic runoff coefficient to improve flash flood early warning in Europe: validation on the 2013 Central European floods in Germany, *Meteorol. Appl.*, accepted, 2014.
- Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resour. Res.*, 46, W05521, doi:10.1029/2009WR008328, 2010.
- Sangati, M. and Borga, M.: Influence of rainfall spatial resolution on flash flood modelling, *Meteorol. Appl.*, 9, 575–584, 2009.
- 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, 2009..
- van Andel, S. J., Weerts, A., Schaake, J., and Bogner, K.: Post-processing hydrological ensemble predictions intercomparison experiment, *Hydrol. Process.*, 27, 158–161, doi:10.1002/hyp.9595, 2013.
- Van Steenbergen, N. and Willems, P.: Increasing river flood preparedness by real-time warning based on wetness state conditions, *J. Hydrol.*, 489, 227–237, doi:10.1016/j.jhydrol.2013.03.015, 2013.
- Viglione, A. and Blöschl, G.: On the role of storm duration in the mapping of rainfall to flood return periods, *Hydrol. Earth Syst. Sci.*, 13, 205–216, doi:10.5194/hess-13-205-2009, 2009.
- Wilks, D. S.: *Statistical Methods in the Atmospheric Sciences: An Introduction*, electronic version, Elsevier, San Diego, CA, 2006.
- Zappa, M., Jaun, S., Germann, U., Walser, A., and Fundel, F.: Superposition of three sources of uncertainties in operational flood forecasting chains, *Atmos. Res.*, 100, 246–262, 2011.
- Zappa, M., Fundel, F., and Jaun, S.: A “Peak-Box” approach for supporting interpretation and verification of operational ensemble peak-flow forecasts, *Hydrol. Process.*, 27, 117–131, doi:10.1002/hyp.9521, 2013.