Discharge Interval method for uncertain flood forecasts using a flood model chain: city of Kulmbach

Md Nazmul Azim Beg, Jorge Leandro, Punit Bhola, Iris Konnerth, Winfried Willems, Rita F. Carvalho and Markus Disse

ABSTRACT

Real-time flood forecasting can help authorities in providing reliable warnings to the public. Ensemble prediction systems (EPS) have been progressively used for operational flood forecasting by European hydrometeorological agencies in recent years. This process, however, is non-deterministic such that uncertainty sources need to be considered before issuing forecasts. In this study, a new methodology for flood forecasting named Discharge Interval method is proposed. This method uses at least one historical event hindcast data, run in several ensembles and selects a pair of best ensemble discharge results for every certain discharge level. Later, the method uses the same parameter settings of the chosen ensemble discharge pair to forecast any certain flood discharge level. The methodology was implemented within the FloodEvac tool. The tool can handle calibration/validation of the hydrological model (LARSIM) and produces real-time flood forecasts with the associated uncertainty of the flood discharges. The proposed methodology is computationally efficient and suitable for real-time forecasts with uncertainty. The results using the Discharge Interval method were found comparable to the 90th percentile forecasted discharge range obtained with the Ensemble method.

Key words | calibration, forecasting, hydrological modelling, uncertainty, validation

INTRODUCTION

The economic loss within the European Union due to flood issue exceeded 60 billion euros from 1998 to 2009 with 1,126 fatalities (EEA 2010; Kauffeldt et al. 2016) making flood resilience one of the prominent issues for many cities. The loss increased in the past decade as a result of climate change, increasing city population and increase in per capita wealth (EEA 2010, 2017). Improved disaster risk management through early warning information is one of the critical procedures to reduce flood losses. Significant success in flood forecasts lies in the accuracy to predict the state of future atmospheric conditions. Yet, numerical weather forecasts are still inaccurate due to the limitation in mathematically representing the non-linear and complex atmospheric physics and also because of the limited resolution of simulated atmospheric dynamics (Lorenz 1969; Buizza et al. 1999; Kauffeldt et al. 2016). Uncertainty may also arise from the inherent issues with the structures of a hydrological model (Renard et al. 2010). Combining these reasons, hydrological forecast models contain uncertainty to a great extent (Beven & Binley 1992; Boyle et al. 2000; Refsgaard et al. 2007; Wani et al. 2017). Assessing uncertainty in the model results is an integrated part for hydrologic modelling and considered necessary in research and practice, especially when models are used for water management issues (Beven 2002; Refsgaard et al. 2005, 2007; Vandenberghhe et al. 2007; Todini 2009; Barbetta et al. 2017).
Different methods have been developed for uncertainty quantification in flood forecasting. Some methods focus on the model input uncertainty such as meteorological input forcing and model initial states, e.g., van Andel et al. (2003) and Li et al. (2009), respectively, and others focus on the hydrological model parameters such as Benke et al. (2008) as well as the hydrological model itself (Coccia & Todini 2011; Deletic et al. 2012).

Among different uncertainty estimation approaches, Bayesian inference methods are very popular as they can utilise multiple parameter values within model structure limitations. One example is the Generalized Likelihood Uncertainty Estimation (GLUE) methodology, described by Beven & Binley (1992). A large number of Monte Carlo simulations are required in GLUE, yet accepting parameter set criteria is subjective and based on a user-defined threshold (Dotto et al. 2011). The results can be sensitive to the choice of the threshold value (Freni et al. 2008). The Bayesian Model Averaging (BMA) method is considered a better approach as it optimises the model posterior density by estimating different weights (Raftery et al. 2005; Vuigt & Robinson 2007). BMA applies predictive probability density function considering weights of discrete bias corrected forecasts. This method reflects the relative contributions to the predictive skill of the model by discouraging ensembles with lower weights, which can be useful in reducing computation costs of running large numbers of ensembles (Raftery et al. 2005). Hydrological Ensemble Prediction System (widely known as HEPS or only EPS: Ensemble Prediction System) is one of the most practised methodologies to predict river flow, which is mainly based on the BMA method. In this process, the system generates an ensemble of river flow forecasts for the same forecast period considering a range of probabilistic assessments of future river flow instead of only one projection. Several case studies are published in the literature indicating its decent performance in forecasting flood, particularly in the case of issuing flood alerts with more confidence (Roulin 2007; Regimbeau et al. 2007; Cloke & Pappenberger 2009; Laurent et al. 2010; Brown 2014). The advantage of EPS forecasts is that it can be instantly used for flood forecasting.

With the advancement of radar technology, it is possible to obtain spatial rainfall data (Codo & Rico-Ramirez 2018). Recent improvements of radar rainfall data accuracy and resolution have shown possibilities (Pedersen et al. 2010) to replace point rain gauge data by spatially variable rainfall forecasts in the near future. However, rain gauge data are still considered better regarding measurement accuracy (Muthusamy et al. 2017). One of the significant challenges of using rain gauge data is to convert the point measurements to spatially distributed data as most of the popular hydrological models are lumped catchment area models and for which the point measurements must be upscaled for the whole catchment area. The uncertainty contribution of the spatial distribution error becomes important when point rainfall data are needed to be interpolated to an area. Geospatial simulations such as Kriging is also a better option for spatial rainfall interpolations as this method considers spatial dependence structures of the data (Mair & Fares 2011; Ly et al. 2013). Kriging with external drifts showed effective and reliable ways to improve the quality of spatial rainfall distribution (Berndt et al. 2014; Cecinati et al. 2017).

In the FloodEvac project, a real-time flood forecasting tool was developed which can forecast flood discharges and flood extents with the inclusion of uncertainties (Disse et al. 2017). The tool can be utilised using Ensemble-based prediction to consider both model input and model parameter uncertainties with flood discharge forecast. The current study explains the hydrological forecast efficiencies and proposes a new alternate methodology to reduce forecast computational time by optimising the ensembles of the forecast. Reporting past performance of the forecast systems is given the highest priority by the hydrologists, researchers and end users to evaluate the forecast performance of a hydrological model (Wetterhall et al. 2013). For this reason, the proposed methodology is validated by hindcasting a few historic flood events.

**FLOODEVAC TOOL AND LARSIM MODEL**

The FloodEvac tool, developed under the FloodEvac project, allows the simulation of the rainfall–runoff process and includes uncertainty from different sources (Leandro
The tool can be run in simulation or forecast modus. The former is suitable for reproducing specific flood events or for the simulation of long time series (e.g., yearly) while the latter is suitable for real-time flood forecasting. The model chain includes a rainfall uncertainty module, an uncertainty and calibration module for the hydrological model, and a link to several hydraulic models.

In the rainfall module, rainfall data can be introduced in three different ways: using (1) hourly observed/forecast rainfall from German Meteorological Services (DWD), generated/collected at each rain gauge location; (2) generated rainfall based on historical data; or (3) generated rainfall based on synthetic data. Uncertainty can then be added to catchment rainfall based on the sequential conditional simulation (Seo et al. 2014).

LARSIM (Large Area Runoff Simulation Model) is a conceptual hydrological model used in the tool. The catchment parameters can be subdivided either in gridded or irregular sub-catchments. The hydrological processes are calculated in a series of sub-catchments connected by flood routing elements in a predetermined sequence. Each sub-catchment has its own properties including elevation, length of watercourses and schematic river cross sections with roughness coefficients. The channel flow routing is done using translation-retention method. It simulates the hydrologic processes for one sub-catchment for a defined period. The resulting output hydrograph is the input information for the next element according to the general model structure rules. The model structure can be both grids or hydrological sub-catchments based. It considers a soil module with storage capacities in calculating the flood routing consisting of three storages at upper, middle and lower soil. All three may contribute to the discharge components, modelled as a linear storage system (Figure 1(a)).

LARSIM is already in use at the Upper Main catchment area for operational flood forecasting, operated continuously by the Flood Forecast Centre of Bavarian Water Authority. The model has been calibrated for historical flood events.

![Figure 1](https://iwaponline.com/jh/article-pdf/21/5/925/602829/jh0210925.pdf)

*Figure 1* | (a) LARSIM hydrological model structure, adapted from Disse et al. (2017) and (b) flowchart of FloodEvac tool with Discharge Interval method, adapted from Leandro et al. (2017).
(Laurent et al. 2010) where each sub-catchment has its own set of 34 calibrated parameters. The current work uses the same initial setup as the operational forecast model which is referred to as ‘base model’ for the rest of the work.

In forecast modus, the FloodEvac tool can generate ensembles by sampling LARSIM parameters, using a beta or a normal probability distribution function. The beta function can produce skewed distribution and hence considers the asymmetric uncertainty parameter intervals around the calibrated parameter set when generating the ensembles.

CATCHMENT AREA AND DATASETS

The current study focuses on the city of Kulmbach located in the catchment area of the Upper Main in Bavaria. The total catchment area is 4,244 km² (Figure 2). A total of 77 rain gauges are available in the catchment area. Historical data are accessible at each rain gauge points from 2005 to 2015 at a temporal resolution of 1 hour. There are also 55 discharge measuring stations in this area. The discharge data are also available at hourly temporal resolution, and collected by Bavarian Hydrological Services. The Kulmbach city contains 92.8 km² with a population of nearly 26,000 (Bhola et al. 2019). Part of the White Main river passes through the city accumulating river flow from Schorgast and upper White Main. The Red Main river also connects Kulmbach at the downstream side of the city at the west. Flood waves coming from the first two rivers are more significant than the third as they are connected at the upstream part of the city and both flows sum up together at the city entrance. There are three discharge measuring gauges at these three rivers, namely, Kauerndorf, Ködnitz and Unterzettlitz, respectively. All three are located just outside the city perimeter. Flood discharge forecast of these three gauges is considered in the current study.

![Figure 2](http://iwaponline.com/jh/article-pdf/21/5/925/602829/jh0210925.pdf)
Between 2009 and 2014, four different flood events were seen in Kulmbach (Figure 3). These flood events occurred during January 2011 (referred to as Jan 2011 event), January 2012 (referred to as Jan 2012 event), December 2012 (referred to as Dec 2012) and May 2013. Within these four events, Jan 2011 was the highest and had a hundred-year return period (Bhola et al. 2019). The recorded discharge was 92 m$^3$/s and 74 m$^3$/s at Kauerndorf and Ködnitz, respectively. However, at Unterzettlitz, the flood discharge was recorded as 252 m$^3$/s. The May 2013 event was second highest, where the recorded highest flows were around 50 m$^3$/s at Kauerndorf, 100 m$^3$/s at Ködnitz and 185 m$^3$/s at Unterzettlitz. During the Dec 2012 flood, Ködnitz gauge recorded close to 65 m$^3$/s flood discharge which was higher than that of Kauerndorf flow, close to 40 m$^3$/s. The last event of Jan 2012 was slightly less significant in terms of observed highest flood discharge.

A hydrological model result may have contributions from different uncertainty sources. However, in this work, the uncertainty estimation in the forecast was done at two stages: (a) rainfall (input) uncertainty and (b) parameter uncertainty. The actual DWD measured hourly rainfall at rain gauge locations was available to this study and is therefore used here as forecast input data. The current work hindcasts the mentioned four flood events using multi-ensemble-based ‘Ensemble method’ and a proposed new method named ‘Discharge Interval method’. The latter method needs a selection of best ensemble member pairs which requires one flood event result utilising multi-ensemble-based hindcast along with a well-calibrated base model. The flood event of Dec 2012 is used to exemplify the forecasting tool (in Figure 3). This event was considered for the selection of pairs of the Discharge Interval method, as it is advised not to use a very high flood event for calibration purposes of a hydrological model (Laurent et al. 2010).

**METHODS**

**Generation of rainfall and rainfall uncertainty**

The rainfall uncertainty module checks observed or forecast rainfall data at the available stations and distributes the data...
within the whole catchment considering sequential conditional geospatial simulation. Previously, Wilks (1998) extended a Markov-chain based approach for precipitation occurrence combined with exponential distribution for nonzero amounts to multiple sites, successfully reproducing means, variances and interstation correlations. The dependence structures may be captured in a more complex manner by copula-based approaches such as Bárdossy & Pegram (2009). Studies in order to simulate precipitation in complex spatial terrains are based on multifractals (Deidda 1999), multiplicative cascades (Gupta & Waymire 1993; Nykanen & Harris 2005) or Gaussian random fields (Shah et al. 1996).

In this work, rainfall is not considered as normally distributed and continuous; on the contrary, the intermittency (alternation of rainfall and no rainfall) and the positive skewness of the distribution of nonzero rainfall are considered within the geospatial simulation, which is a similar approach to Wilks (1998). Therefore, a discrete–continuous mixed distribution is considered, using two variants. The discrete part of the distribution is empirically recorded via the proportion of zeros in the total sample, and the continuous part is mapped on the three-parametric gamma distribution as well as by a nonparametric kernel density. The whole geostatistical simulation is implemented using two different R-packages, namely, gstat (Gomez-Hernandez & Journal 1993; Pebesma 2004) and RandomFields (Schalather et al. 2015).

The LARSIM model takes input from distributed rainfall data for the catchment at a spatial resolution of 1 km × 1 km. Preliminary tests showed that the variation in the spatial interpolation of the rainfall distribution has less effect on discharge forecasts when compared to the effects of variation in hydrological parameters. Therefore, and for this work, only 10 rainfall simulation sets were considered to estimate the uncertainty of the spatial rainfall distribution (Figure 4).

**Generation of parameter uncertainty**

An analysis of parametric uncertainty was performed in LARSIM base model to derive the most sensitive parameters of the model regarding flood discharge at the upstream gauges of Kulmbach, i.e., Ködnitz and Kauerndorf. Eight parameters were identified as most sensitive including EQD, beta, BSF, EQB, EQI, EQD2, A2 and Dmax (Härder 2017). In this work, only these eight parameters were considered for uncertainty analysis out of the 34 available parameters in the LARSIM model. Table 1 shows the description and value range of these considered parameters. Later, the hydrological model was set for auto calibration, applying historical observed flow data of five years (2009–2013) at Ködnitz and Kauerndorf gauges. This step assigned a set of 34 model parameters at each of the 81 sub-catchments. In a next step, flood discharge from the original operational model (Laurent et al. 2010; Haag et al. 2016), currently in use by the Flood Forecast Centre at the Bavarian Water Authorities (base model) was compared with the one obtained using the automatic calibrated model. Similar results were obtained between the two models, which indicated that the hydrological response of the study area remained unchanged from 2009 to 2013 considering the base model calibration. Hence, the base model parameters were kept unchanged and considered as the calibrated parameters in this work.

In the end, the ensemble was generated randomly from a given parameter range, as described by Härder (2017). Different normal or beta probability distribution curves were prepared for each sub-catchment such that the curves consider the base model parameter as the median value and empirical parameter range as the extreme values. Later, one set of parameters was chosen randomly for each of the sub-catchments considering the eight parameters mentioned in Table 1. The remaining 26 parameters were kept unchanged as the base model.

**Forecast of flood discharge using FloodEvac Ensemble method**

Before the hindcasting process can start, a warm-up period is required. The model is simulated using observed precipitation and temperature data for one year of the warm-up period until 49 hours before the forecast initialisation time. At this stage, the model uses the base model parameters. The model results are stored in an ‘initial state file’ at the end of the one-year warm-up simulation time. As such, we ensure that the internal model state condition of the basins is as close as possible to the real conditions. Later,
this initialisation state is used to simulate each forecast run of the ensemble. However, in the case of starting the simulation from an ‘initial state file’, it is recommended to start the flood forecasting for at least 49 hours ahead of the initial time (Ludwig & Bremicker 2006), to allow the model sufficient additional warm-up period before producing reliable
results. Therefore, each forecast simulation was run for 61 hours. In this process, the tool collects 49 hours of observed hourly rainfall followed by 12 hours of forecast rainfall data. These 61 hours of rainfall data are passed through the rainfall distribution uncertainty module, and 10 different rainfall uncertainty datasets are prepared. Later, 50 different parameter sets are produced using the parameter uncertainty module. These 50 parameter sets are combined with the 10 rainfall uncertainty cases, repeating one rainfall scenario with every five parameter sets in sequential order, thus, making 50 sets of hydrological models for the Upper Main catchment. These 50 models are simulated, and the results of discharge ensembles are stored. In this way, the runtime required for producing a forecast for the next 12 hours is around 25 minutes on a desktop with 3 GHz Intel processors and 16 GB of RAM, running in parallel mode using three cores. Later, the first 49 hours of simulation results are deleted, and the last 12 hours of forecast data are stored.

New weather forecast/updated rain gauge data become available at every hour in this case. The tool can receive real-time observed rainfall and discharge data at every hour. For this reason, the forecast process repeats at every hour. In the next forecast process, the model simulates a new ensemble of 50 simulation runs of 61 hours simulation each, of which the first 48 hours are a repetition of the previous simulation; the 49th-hour data use the latest available real-time observed data and the last 12 hours data use the newly available weather forecasts. At this stage, the tool regenerates 10 sets of rainfall simulations using the same pattern and the model uses the same parameter set that was generated at the first stage.

**Forecast of flood discharge using the Discharge Interval method**

In the case study, the weather forecasting frequency is 1 hour. In the case of flood forecast mode, the new results would become outdated at the end of an hour as updated weather forecasts become available. However, the Ensemble method takes a significant fraction of an hour in flood discharge forecasting. For real-time forecasts, it is intended to reduce the forecast time requirement. In this section, we propose an alternative option to produce similar uncertainty bands of forecast within a much shorter period. This methodology is termed as ‘Discharge Interval method’ hereinafter.

The Discharge Interval method is proposed to optimise the Ensemble method in order to improve the computational time required for the forecast model. We chose different parameters for different flood intervals in order to properly forecast all the flood levels with the intention to forecast the possible maximum and minimum flood discharges instead of forecasting all the ensemble members. However, this method requires prior knowledge of the hydrological response in the catchment area before the actual forecast, which can be obtained by hindcasting a flood event using the Ensemble method. This event is considered as the best pair selection event.

In the proposed method, the possible observed discharge is divided into predefined discharge intervals ($\Delta Q$),

<table>
<thead>
<tr>
<th>Parameter name</th>
<th>Unit</th>
<th>Description</th>
<th>Recommended empirical value range (Härder 2017)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQB</td>
<td>[-]</td>
<td>Gauging size for the retaining constant of the basis discharge storage</td>
<td>10,000–50,000</td>
</tr>
<tr>
<td>EQ1</td>
<td>[-]</td>
<td>Calibration variable for the retaining constant of the interflow storage</td>
<td>1,000–5,000</td>
</tr>
<tr>
<td>EQD</td>
<td>[-]</td>
<td>Calibration variable for the retaining constant of the slow runoff storage</td>
<td>500–1,000</td>
</tr>
<tr>
<td>EQD2</td>
<td>[-]</td>
<td>Calibration variable for the retaining constant of the fast runoff storage</td>
<td>10–500</td>
</tr>
<tr>
<td>A2</td>
<td>[mm/h]</td>
<td>Threshold value, if reached the surface runoff will be assigned to the fast runoff storage</td>
<td>0.5–4.0</td>
</tr>
<tr>
<td>BSF</td>
<td>[-]</td>
<td>Exponent of the soil moisture saturation area function for adjustment of the share of runoff as a function of the soil storage load</td>
<td>0.01–0.5</td>
</tr>
<tr>
<td>beta</td>
<td>[1/d]</td>
<td>Drainage index for the deep seepage (to the basis discharge storage)</td>
<td>0.000002–0.1</td>
</tr>
<tr>
<td>Dmax</td>
<td>[-]</td>
<td>Index for lateral drainage to the interflow storage in the area of large grain sizes</td>
<td>0–10</td>
</tr>
</tbody>
</table>
based on the selection event. A pair of flood discharge ensembles is selected for each interval in such a way so that the ensemble pair can bracket the maximum numbers of observed discharge data within them. At a later stage, only the selected best pair ensemble parameter sets are used to forecast the possible maximum and minimum discharge for any other events within the same catchment. The pairs are expected to give a maximum and a minimum limit of flood discharge forecast. Larger intervals will lead to very few best pairs. Very small intervals will lead to many best pairs, which loses the benefits against using the ensemble in the first place. Figure 1(b) shows a flowchart of the whole process.

UNCERTAINTY ANALYSIS OF THE ENSEMBLE METHOD

The uncertainty in the Ensemble method forecast is assessed in this section. The flood forecast results of Dec 2012 event at gauge Ködnitz is used as an example. The event had two peaks, on 24 and 28 December, respectively. The peak discharges were recorded as 64.2 m$^3$/s and 58.4 m$^3$/s on 24 and 28 December, respectively. Figure 5 shows the results of all the ensemble simulations for the flood discharge forecast at gauge Ködnitz.

As in the hindcast, the rainfall data are available at every hour; the Ensemble method forecast was done using 50 ensemble members to limit the computation time required. The forecast quality might have been improved if more uncertainty runs were used. However, it can be seen from Figure 5 that the ensemble of these 50 members could effectively bracket the observed discharge data at both Kauerndorf and Ködnitz.

The quality of the forecast data is assessed from each hourly lead of the forecast. Figure 6 shows the scatter plot of ensemble forecasts and corresponding observed discharge data. The inconsistency between forecast and observed data increases with increasing forecast lead time, as indicated by their increasing mean error (BIAS) and root mean square error (RMSE) values. It is apparent that the model is excellent for forecasting flood up to 4 hours in advance as it shows a narrow spread at the scatter plot with small BIAS (around $-0.25$ m$^3$/s) and the RMSE is below 3 m$^3$/s. The uncertainty in forecast within 5 or 6 hours lead time is slightly worse as these values increase. However, the scatter plots at a lead time of 7 to 12 hours are even wider with higher RMSE, which is an indication

![Figure 5](image_url)
of higher uncertainty. The shape of the scatter spread as well as the BIAS and the RMSE values are almost consistent at 10 to 12 hours lead time, which indicates that at a lead time of more than 9 hours, the forecast uncertainty becomes stable and stops increasing.

It can be seen from different forecast leads, especially at 7 to 12 hours, that LARSIM underestimated the high flood peaks. According to Ludwig & Bremicker (2006), the cause could be the lack of high temporal resolutions or quality of input data (rainfall measurements).

In assessing the uncertainty of the basin response to uncertain flood discharge forecast, confidence intervals are calculated. According to the procedure described, for each time point of the forecast horizon, the discharges were forecasted 12 times using a 50 member ensemble of uncertainty runs. In this way, each temporal result is predicted 600

![Figure 6](image-url)
times, which is used for the statistical assessments for each time step.

A confidence interval chart is shown for all the simulated forecast data (Figure 7). We calculate the confidence intervals of the whole forecast data in two parts: considering a forecast lead time up to 6 hours and 12 hours. In calculating the confidence intervals at a specific date and time, all the forecasted data at that time were considered for calculating the corresponding percentiles. Comparing with the observed flood discharge of the area, it can be seen that the model can forecast the rising limb of the flood peak reasonably well. Both rising limbs of the observed flood discharge lie within the 90th percentiles of the simulated results at both 6 hours and 12 hours lead forecasts. The peak discharge predicted in the model simulation is slightly earlier than the actual flood peak time. However, the uncertainty interval is found higher at the flood peak as the range of 90th percentile band at 12 hour lead is 42 and 78 m$^3$/s, respectively. The similar uncertainty band is found considerably lower at 6 hours lead, varied between 48 and 65 m$^3$/s. The falling limb of the flood discharge is found within the 98% confidence interval of the simulated discharges.

**IMPROVING THE FORECAST COMPUTATIONAL EFFICIENCY: DISCHARGE INTERVAL METHOD**

**Defining the best pairs**

The Discharge Interval method requires one ensemble-based flood hindcast result to optimise the forecast hydrological model parameters. As mentioned earlier, data at Ködnitz gauge point from Dec 2012 event were chosen for this purpose.

In Ködnitz river gauge, the highest warning level (warning level 4) is equal to 69.1 m$^3$/s (Bavarian State Office for the Environment 2018). Considering this issue, the possible observed discharge is divided into seven segments with an interval of 10 m$^3$/s starting from 0 discharge. Each possible combination of ensemble member pair is investigated considering the number of observations. For this part, all the

---

**Figure 7** | Confidence intervals in flood forecasting at Ködnitz for Dec 2012 event (adapted from Beg et al. 2018).
ensemble results, each having 50 members of 12 hours of temporal forecast data were checked, and a list of all the possible member pairs were listed. In this case, the current data had 50 members in each ensemble which creates a total of \((50!/(2! \times 48!)) = 1,225\) possible pair combinations.

In the next step, the tool checks which pair of ensemble members can bracket the maximum numbers of observation discharge of a certain interval within them. A pair of members containing the maximum numbers of observation is selected and termed as ‘best pair’ for that discharge interval. Later, the best pairs were applied to the other three events, e.g. whenever the latest observed discharge is within that certain interval, the forecast model would use only the best pair parameter sets to forecast the next 12 hours of flood discharge. The best pairs found for the current model setup are listed in Table 2.

For example, on 23 December 2012 at 19:00 (Figure 8), before starting the forecast for the next 12 hours, the forecast tool checks the latest available observed flood discharge at Ködnitz gauge and finds that the observed flow at that time was 45.2 m³/s. Then the tool forecasts the next 12 hours of discharge using parameters of ensemble member number 10 and 16. After 1 hour, on 23 December 2012 at 20:00, the forecast tool repeats its forecast. At that time, the latest available observed forecast was 52.1 m³/s, and the forecast tool chooses a new pair of ensemble members to forecast the next 12 hours of flood, which are member number 32 and 45. However, it should be noted that the Upper Main river catchment area contains 81 sub-catchments and each of the ensemble member set contains different parameters at each sub-catchment level. As such, ensemble member 1 contains 34 parameters which are different at each of the 81 sub-catchments. Out of these 34, the first eight parameters, as mentioned in Table 1, are randomly selected and the remaining 26 have the same value as those in the base model. Ensemble member 2 also has 34 different parameters at each of the sub-catchments. However, eight of these parameters at each sub-catchment are different from those assigned at ensemble member 1 and the remaining 26 are the same. In this way, the whole combination of parameters is directly related to better agreement of discharge observation at a certain interval.

**Validation of the Discharge Interval method**

To check the effectiveness of the proposed Discharge Interval method in the above section, we applied the methodology to hindcasting, using three other flood event scenarios. The chosen three additional flood scenarios were (Figure 3): January 2011, January 2012 and May 2013. The methodology was also applied to the gauge locations of Kauerndorf, Ködnitz and Unterzettlitz, as these were located just at the upstream of the city of Kulmbach.

For these cases, the hindcast was run using both Ensemble method and Discharge Interval method, utilising the same combination as described earlier. The seven ensemble member pairs chosen from the Dec 2012 event were used in the Discharge Interval method to check if they can still be able to forecast the discharge appropriately at these three gauges. The results are shown in Figure 9.

During the hindcast of January 2011, the observed discharge showed two peaks with a maximum discharge close to 80 m³/s at both Kauerndorf and Ködnitz and 180 m³/s at Unterzettlitz. The flood forecast predicted a similar flood peak at Kauerndorf and Unterzettlitz, however, much higher peak discharge at Ködnitz (160 m³/s). The seven pairs of selected ensemble members showed qualitatively similar flood prediction to the 90th percentile of the Ensemble method. During the first flood peak between 6 and 12 January 2011, flood discharges were precisely forecasted using Discharge Interval method at each hour. The rising limb, peak and the falling limb were accurately forecasted for this case. At the second peak discharge, the Discharge Interval method forecast showed a much faster rising and falling limb with a higher peak prediction than the observed.

**Table 2 | Observed best pair sequence in the work**

<table>
<thead>
<tr>
<th>Discharge interval (m³/s)</th>
<th>Ensemble member best pair</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 &lt; Q ≤ 10</td>
<td>36 and 46</td>
</tr>
<tr>
<td>10 &lt; Q ≤ 20</td>
<td>17 and 34</td>
</tr>
<tr>
<td>20 &lt; Q ≤ 30</td>
<td>10 and 36</td>
</tr>
<tr>
<td>30 &lt; Q ≤ 40</td>
<td>16 and 26</td>
</tr>
<tr>
<td>40 &lt; Q ≤ 50</td>
<td>10 and 16</td>
</tr>
<tr>
<td>50 &lt; Q ≤ 60</td>
<td>32 and 45</td>
</tr>
<tr>
<td>Q &gt; 60</td>
<td>2 and 45</td>
</tr>
</tbody>
</table>


Downloaded from http://iwaponline.com/jh/article-pdf/21/5/925/602829/jh0210925.pdf by guest
In the second flood event scenario of January 2012, Discharge Interval method could not forecast the flood event with enough accuracy at any of the three gauge points. The first flood hydrograph has a duration of 2 days, considering its time of rising, peak and retention time. The second peak discharge in this event had a duration of nearly 4 days. The Ensemble-based forecast model as well as the Discharge Interval forecast model showed, to some extent, better performance at forecasting the second peak than the first. However, the highest discharge value was underpredicted at both peak flows at Kauerndorf and Ködnitz but overestimated at Unterzettlitz. The forecast of the falling limb of this flood event showed better accuracy at all the three gauges.

In May 2013, the observed discharge peak was found to be up to 50 m$^3$/s, 120 m$^3$/s and 180 m$^3$/s at the three gauges, respectively. The entire flood duration was for 11 days. Like the other events, it also contained two peak discharges. Both forecast models underestimated the first peak but overestimated the second at both Kauerndorf and Ködnitz, but overestimated at Unterzettlitz. Although, in all the cases, the Discharge Interval showed similar characteristics forecast range to that of the 90% interval of the Ensemble method.

In the case a different flood event was used as a selection event of best pair, there could have been a marginal change in the selection of parameter sets at different discharge intervals. However, as the results already showed a good agreement with the validation of three additional events at three gauges, totally different parameter sets are not expected. In any case, by increasing the ensemble members, e.g., 100 runs, it might produce a different set of parameters, which may slightly improve our results. A drastic increase is not expected since the validation already showed similarity to that of the Ensemble method.

Assessing the forecast quality

To check the forecast quality, a likelihood table is presented considering the forecasted discharge using both Ensemble method as well as our proposed Discharge Interval method. For this purpose, data at Ködnitz gauge was considered (Table 3). The table was formed considering yes/no forecasts and yes/no observed discharge consideration after dividing the forecasted discharge range into several subgroups/discharge interval thresholds. Hindcasted discharge data and observed discharge at each hour were checked if they fall within a given threshold bin and a data table was prepared considering a ‘yes’ if the data are below the given threshold and a ‘no’ if the data are over the threshold. When both forecast and observed data show ‘yes’, that is considered as a hit (a) and when both show ‘no’, that is considered as another success of the forecast.
After the likelihood table construction, hit ratio (also known as the probability of detection–yes or PODy) and false alarm ratio (also known as the probability of false detection or POFD) are counted using the following equations for every discharge threshold considered:

\[
\text{Hit ratio} = \frac{a}{a + b} \\
\text{False alarm ratio} = \frac{b}{b + d}
\]  

The results are represented as Receiver Operating Characteristics (ROC) (Mason & Graham 2002; Marzban 2004; Liguori et al. 2012; Codo & Rico-Ramirez 2018), which describe the occurrence of hit and false alarm rate for a defined event, helping in crucial decision-making issues (Mason & Graham 2002).

The area under the curve is a good indicator of the forecast quality and applied in many studies to assess the goodness of numerical forecasts (Murphy & Winkler 1987;
Theshold discharge

The Ensemble method and the Discharge Interval method both showed almost similar forecast quality (see Figure 10). ROC curve band created by the Discharge Interval method was narrower, which also indicates a narrower flood forecast range. However, out of the four events, the January 2011 event had the worst forecast quality in both methods. Forecasts of the other three events showed a better performance. All the forecast data showed less hit ratio (PODy) with higher false alarm (POFD) when forecasting small discharges. At forecasting higher discharges, the PODy was much closer to 1 for January 2012, December 2012 and May 2013 events. During December 2012 and May 2013 events, the 5th and 95th percentile ensembles’ ROC curves intersect each other. This may happen in some cases based on the observed data and the interval chosen in ROC curve construction. For example, for a particular discharge interval, if observation and the 95th percentile fall in the same interval whereas the 5th percentile remains below, the 95th percentile would contribute to higher ROC area. It may change for another case where the 5th percentile and the observation may fall in the same category keeping the 95th percentile outside. In that case, the 5th percentile will contribute to higher ROC area. For these reasons, the ROC curves may intersect each other. Either of the percentiles may not necessarily be better than the other. The same condition applies to the upper and the lower bound forecasts of the Discharge Interval method. In Table 4, all the average area under the ROC curve is presented. For the Ensemble method, the average area from 5th and 95th percentile are calculated and for Discharge Interval method, the averaged area between upper bound ROC and the lower bound ROC is considered.

Comparing the two forecast bands, we can see that the ROC curve areas of both forecast methods are adjacent to each other. The flood events of January 2012, December 2012 and May 2013 forecasts showed high predictability and had ROC curve area between 0.83 and 0.95. The January 2011 event had the least forecast area under the curve. These results also show that the predictability of discharge decreases slightly with lead in both forecast methods.

Table 4 also shows the fact that the forecast data provided by the Discharge Interval method are similar to that of the Ensemble method. The forecast band provided by this method can be comparable to the 90th percentile interval of the Ensemble method. This indicates that the Discharge Interval method forecasts the discharges considerably faster as well as with reliable accuracy.

Forecasting using Discharge Interval method takes less than 4 minutes to run a 12-hour forecast simulation set using a desktop with 3 GHz Intel processors and 16 GB of RAM. This indicates a seven times speed-up when compared to the simulation of Ensemble method with 50 members.

**CONCLUSIONS**

In this work, a new flood discharge forecast methodology has been proposed to produce reliable real-time flood forecast.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Likelihood table for calculating a forecast category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threshold discharge $\leq x$</td>
<td>Observed discharge $\leq x$</td>
</tr>
<tr>
<td>Forecast discharge $\leq x$</td>
<td>a</td>
</tr>
<tr>
<td>Forecast discharge $&gt; x$</td>
<td>c</td>
</tr>
</tbody>
</table>

**Zhang & Casey 2000; Mullen & Buizza 2000**. A good forecast is indicated by a higher area under the curve, which is closer to 1. If a forecast can predict a predefined event perfectly, the corresponding event would calculate a PODy = 1 and POFD = 0, and the point would be located at the left top corner of the curve. The first diagonal line represents the line of ‘No skill’. Whenever a point is located below the diagonal line, it is termed as worse scenario than no prediction. The computational method is not parametric without suggesting any hypothesis regarding the probability distribution of the forecast. The ROC curve characteristic, however, is dependent on the number of discrete values used in creating the curve (Mason & Graham 2002).
This is based on optimised Bayesian Model Averaged method and termed as Discharge Interval method. This method uses previously hindcasted flood discharge uncertainty ensembles and chooses a pair of ensemble members which can bracket maximum observations within a given discharge range. Later it uses the parameter sets of the selected ensemble members to forecast a flood of the same range. The pair creates an upper and lower limit of the flood discharge forecast.

The new method was implemented within the Flood-Evac tool developed for real-time flood forecasting. The tool considers both input and parameter uncertainty in simulating flood scenario ensembles. It creates spatial rainfall distributions using geospatial simulations and the uncertainty of parameters considering multi-Ensemble based simulations and provides the combinations of scenarios to a hydrological model to generate flood discharge forecasts.
ensembles. The tool was applied to hindcast flood at the city of Kulmbach located at the catchment of the Upper Main river.

In this case study, the tool effectively hindcasted four historical flood events and was found useful to predict flood discharge with 12 hours of lead time using both Ensemble method and Discharge Interval method. Uncertainty analysis was done from the Ensemble method results. Analysis showed that the uncertainty in the forecasted discharge increases with the forecast lead time. The forecast showed a narrow scatter plot spread for up to 4 hours lead time. The forecast within 5 to 6 hours lead showed a slightly wider uncertainty band which would give approximately 7.5 hours of preparation time for taking the necessary preventive actions by the city council.

The forecast quality of the proposed Discharge Interval method was compared with results from Ensemble-based forecast simulations. The comparison was done using ROC curves. ROC curves of the Ensemble method and Discharge Interval method showed that the latter methodology could deliver qualitatively similar forecast data when compared with the results provided by the former. Still, the time required for forecasting was considerably small. However, this method is unable to produce confidence interval band estimations in the forecast scenarios.

The proposed Discharge Interval method is very fast indicating an increase of speed of seven times as it runs only a pair of simulations for the forecast. However, the method is dependent on the proper hydrological response of the training event/ensemble member selection event. The method considers that the hydrological response of the catchment remains unchanged during the other forecasts. For this reason, it is recommended to retrain the model after a high flood event.

ACKNOWLEDGEMENTS

The development of the FloodEvac tool is funded by the German Federal Ministry of Education and Research (BMBF, FKZ 13N13196 (TU Munich)). The first, second and sixth authors would also like to acknowledge the QUICS project funded by the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 607000 and Project UID/MAR/04292/2019, financed by MEC (Portuguese Ministry of Education and Science) as well as the ESF (European Social Fund), under the programs POPH/QREN (Human Potential Operational). Very special thanks to the Bavarian Water Authority and Bavarian Environment Agency in Hof for providing us with the quality data to conduct the research.

REFERENCES


