

Assessing uncertainties in flood forecasts for decision making: prototype of an operational flood management system integrating ensemble predictions

J. Dietrich^{1,*}, A. H. Schumann¹, M. Redetzky², J. Walther², M. Denhard³, Y. Wang¹, B. Pfütznner⁴, and U. Büttner⁵

¹Inst. of Hydrology, Water Resources Management and Environmental Engineering, Ruhr Univ. Bochum, Bochum, Germany

²DHI-WASY GmbH, Dresden, Germany

³Deutscher Wetterdienst DWD (German National Weather Service), Offenbach, Germany

⁴Büro für Angewandte Hydrologie, Berlin, Germany

⁵Sächsisches Landesamt für Umwelt, Landwirtschaft und Geologie, Dresden, Germany

*now at: Institute of Water Resources Management, Hydrology and Agricultural Hydraulic Engineering, Leibniz University, Hanover, Germany

Received: 3 March 2009 – Revised: 30 July 2009 – Accepted: 14 August 2009 – Published: 31 August 2009

Abstract. Ensemble forecasts aim at framing the uncertainties of the potential future development of the hydro-meteorological situation. A probabilistic evaluation can be used to communicate forecast uncertainty to decision makers. Here an operational system for ensemble based flood forecasting is presented, which combines forecasts from the European COSMO-LEPS, SRNWP-PEPS and COSMO-DE prediction systems. A multi-model lagged average super-ensemble is generated by recombining members from different runs of these meteorological forecast systems. A subset of the super-ensemble is selected based on a priori model weights, which are obtained from ensemble calibration. Flood forecasts are simulated by the conceptual rainfall-runoff-model ArcEGMO. Parameter uncertainty of the model is represented by a parameter ensemble, which is a priori generated from a comprehensive uncertainty analysis during model calibration. The use of a computationally efficient hydrological model within a flood management system allows us to compute the hydro-meteorological model chain for all members of the sub-ensemble. The model chain is not re-computed before new ensemble forecasts are available, but the probabilistic assessment of the output is updated when new information from deterministic short range forecasts or from assimilation of measured data becomes available. For hydraulic modelling, with the desired result of a probabilistic inundation map with high spatial resolution, a replacement model can help to overcome computational lim-

itations. A prototype of the developed framework has been applied for a case study in the Mulde river basin. However these techniques, in particular the probabilistic assessment and the derivation of decision rules are still in their infancy. Further research is necessary and promising.

1 Introduction

Due to the intrinsic uncertainty of meteorological forecasts, flood forecasts are also affected by uncertainty. Furthermore, as hydrological models are used for transformation of rainfall into runoff, their structural parameter uncertainty should be considered in the forecasts as well. Inaccurate human interaction and technical problems may also affect the output of a flood forecast chain. Thus uncertainty is an issue of concern when dealing with flood forecasts. On the one hand, an underestimation or even missing of a flood warning may hinder the affected people from preparing for a flood. As a consequence damage may increase or even casualties may occur. On the other hand, “crying wolf” too often may encourage people to ignore warnings in the future.

During the last decades, modelling evolved from a deterministic towards a probabilistic paradigm. Nowadays forecasters have information about uncertainty. The imperfection of forecasts is more and more accepted and the communication of uncertainties does not automatically make the users losing confidence in the forecast. Numerous approaches for uncertainty estimation and probabilistic evaluation were developed. Interdisciplinary studies dealing with the probabilistic assessment of the flood forecast chain were published



Correspondence to: J. Dietrich
(dietrich@iww.uni-hannover.de)

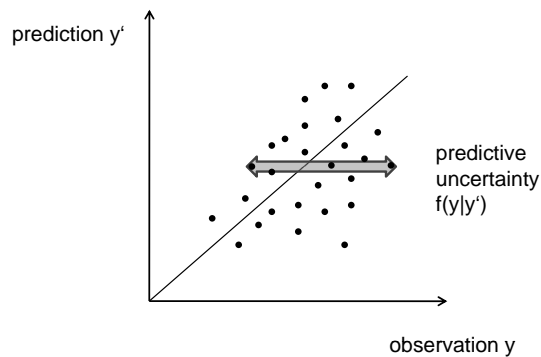


Fig. 1. Predictive uncertainty, describing the differences of observations from a prediction.

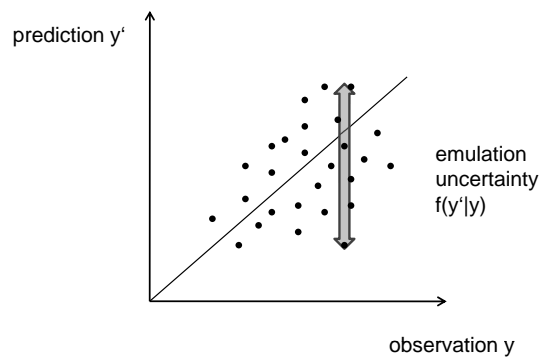


Fig. 2. Emulation uncertainty (Todini, 2009).

e.g. by Krzysztofowicz (2002), Apel et al. (2004) and Papenberger et al. (2005).

The predictive uncertainty characterises differences between observed values and forecasted values (Fig. 1). The forecasts are derived from a single model with defined parameters under event-specific initial and boundary conditions. Forecasts of many different events can be used to characterise the forecast uncertainty empirically.

Ensemble techniques were developed in order to frame the uncertainty with a relatively low number of simulations (Anderson, 1996; Kalnay, 2002; Toth et al., 2003). Within the context of operational forecasting, this makes ensemble techniques superior to many other ways of generating a probability distribution instead of a single solution. Thus ensembles specify the differences between several forecasts (Fig. 2). Different types of ensembles can be classified according to the generating mechanisms (for meteorological as well as for hydrological applications):

- single system ensembles: variation of initial and boundary conditions, different model components, e.g. convection schemes (physically based ensembles), variation of model parameters;

- multiple systems or multi-model ensembles (“poor man ensembles”): combination of simulations from different models (e.g. Georgakakos et al., 2004; Ajami et al., 2007);
- lagged average ensembles: combination of current forecasts with forecasts from earlier model runs (Hoffman and Kalnay, 1983).

Ensembles characterise the so-called “emulation uncertainty” (Todini, 2009). In informatics, the term “emulation” describes the imitation of the behaviour of a computer or other electronic system with the help of another type of computer or system. The differences between forecasts within an ensemble can be used to discuss the effects of the mechanisms used to construct this ensemble.

However, the probabilities within ensembles are not aleatoric ones, because many epistemic uncertainties are included in their estimation. Frequentists use Monte Carlo simulation to account for uncertainty associated with the parameters of a probability model that Bayesian methods handle natively. Ensembles are based on a lower number of data points and make many assumptions about the models and other characteristics. It should be considered that the existing ensemble forecasting systems do not completely represent the uncertainties of models. Hence a probabilistic evaluation of the outcome is needed. Let’s say there may be more uncertainty in uncertainty propagation itself. Thus it is challenging to develop ensemble generation mechanisms which do not only result in a bunch of model outcomes, but also represent the probabilistic assessment of the variables under consideration. The basic assumption is that each result has the same probability. Neither the total variation (which would result from a large amount of possible combinations of uncertain aspects of modelling) nor a differentiation in more or less probable forecasts can be represented in this way.

Nowadays meteorological ensemble prediction systems (EPS) are operational at global scale (e.g. ECMWF, MSC, NCEP, Buizza et al., 2005) and regional scale. The COSMOLEPS (Molteni et al., 2001) is a limited area physically based EPS for the medium range (3 to 5 days lead time). It was developed within the COSMO (Consortium for Small-scale Modelling) to improve the predictability of extreme weather events in Central and Southern Europe. The added value of the system resides in joining the skill of the ECMWF EPS to depict the possible evolution scenarios with the capability of the COSMO limited area model to improve the descriptions of local meteorological processes. Further regional EPS in Europe are MOGREPS, NORLAMEPS, ARPEGE/ALADIN and PEARP. The short-range SRNWP-PEPS (Denhard and Trepte, 2006) combines up to 23 deterministic forecasts from 21 national meteorological services with a lead time of two days. It can be seen from case studies and probabilistic verification for Germany (Trepte et al., 2006) that this ensemble is a valuable tool for severe weather forecasting. A major benefit of this multi-model EPS is the possibility to compare

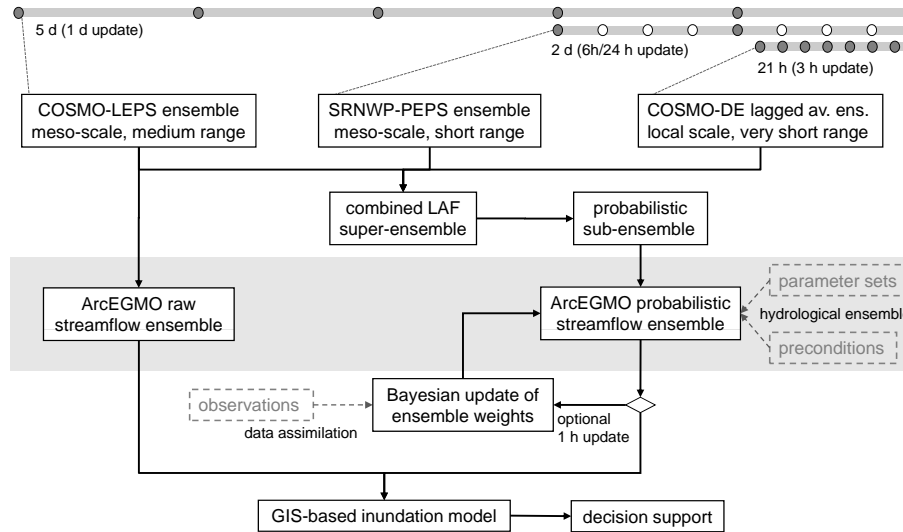


Fig. 3. Flow chart of the flood forecast chain.

the behaviour of all operational European limited area models.

The development of hydrological applications of ensemble forecasts has started in the late 1990-ies and is subject of ongoing research (e.g. Verbunt et al., 2006; Komma et al., 2007; Reed et al., 2007; Diomedea et al., 2008). The participatory HEPEX project (Hydrological Ensemble Prediction Experiment, Schaake et al., 2007) integrates meteorologists, hydrologists and users in order to promote the development of ensemble stream flow forecast systems. In Europe, the probabilistic Flood Alert System (EFAS) is under development (Thielen et al., 2009). EFAS aims to provide flood information for the medium to long-range at large scale river basins being relevant for decisions at national or EU level.

For the setup and near real-time operation of an uncertainty aware flood management system, a compromise between predictive capability of the models, computational efficiency and cognitive burden for the flood managers is still unavoidable. On the one hand, data flow and control activities must be automated to the highest achievable level. On the other hand, the complex nature of the problem requires options for flood managers and decision makers to take control over the simulation process, e.g. when sources of information are identified as unreliable or when parts of the model chain fail. This contribution presents a framework for ensemble based flood forecasting, which is based on hydrological forecasts driven by operational meteorological EPS with different spatial resolution and different lead times. The hydrological models are controlled in an adaptive way, mainly depending on the lead time of the forecast, the expected magnitude of the flood event and the availability of measured data.

In the following section we introduce the respective workflow and a corresponding software prototype for an operational flood management system (OFMS). In Sect. 3

we demonstrate the combination of ensemble members from different meteorological forecast systems and different model runs to generate a multi-model super-ensemble with weighted members. Furthermore we introduce an approach for the a priori framing of parameter uncertainty with a hydrological ensemble. Results from a case study in the Mulde river basin including hindcast simulations of historic flood events were shown by Dietrich et al. (2008) and Dietrich et al. (2009). Here we focus on methodological developments.

2 Methods

2.1 Concept and workflow of an operational flood management system

The operational flood management system (OFMS) presented in this paper is designed for ensemble-based flood forecasting in meso- to macro-scale river basins (100 km² to 10 000 km² contributing catchment area). The OFMS integrates meteorological forecasts from different operational limited area prediction systems and supports the flood manager in managing uncertainty. Typically the OFMS is installed in a regional or national flood management centre.

The OFMS combines meteorological medium-range forecasts from COSMO-LEPS (3 to 5 days lead time), short-range forecasts from SRNWP-PEPS (1 to 2 days lead time) and very short-range forecasts (<1 d lead time) in an adaptive manner (Fig. 3). For the latter the convection resolving, deterministic COSMO-DE model with 21 h lead time and horizontal resolution of 2.8 km is used (Doms and Förstner, 2004). The COSMO-DE model forecasts from different initialization times are combined to generate a lagged average forecast (LAF) ensemble, which proved to be an

improvement over the use of the latest deterministic model run in the Mulde case study (Dietrich et al., 2008; Dietrich et al., 2009). The lead time of the OFMS could be extended to 10 or more days by way of using predictions from global systems, but the spatial resolution of these systems restricts their applicability in meso-scale flood forecasting.

Each of the three meteorological prediction systems has individual advantages for flood forecasters. Medium-range forecasts provide the basis for decisions about reservoir management and early warnings previous to a potential extreme flood event. Short to very short-range forecasts are used for issuing flood alerts and planning of local flood defence measures. One would expect that the shorter range forecasts are more accurate than the medium range forecasts. As it has been shown in the Mulde case study mentioned above, this expectation does not always hold.

In a real time situation, the attention of flood managers is focused on a variety of aspects and the computational and cognitive resources are limited. There is not time to speculate which result would be the best this time. Decision makers need detailed information about relevant criteria like peak time, peak discharge and possible inundation areas. An approach for combining the maximum information from a priori knowledge with the actual forecasts is the generation of a calibrated super-ensemble for the short-range, which also includes information from medium range forecasts and from earlier model runs. Because such a multi-model lagged average super-ensemble has too many members for real time computation of stream flow forecasts, a reduced sub-ensemble is generated as a representative subset of the members of the super-ensemble. The sub-ensemble aims at producing a high resolution probabilistic weather scenario as input for flood forecast models (Fig. 3, detailed description in the following section).

A hydrological rainfall-runoff model simulates ensemble forecasts of stream flow at several points of interest like gauges and vulnerable sites. We decided to choose the conceptual hydrological model ArcEGMO (Becker et al., 2002) as the default OFMS component for transformation of quantitative weather forecasts into stream flow forecasts for the Mulde case study. Conceptual rainfall-runoff models describe the complex natural hydrological processes in a simplified manner. These models are widely used for the meso- and macro-scale due to their significant advantages compared to physical models regarding parameter estimation and computational efficiency (Carpenter and Georgakakos, 2006; Smith et al., 2004).

The OFMS simulates stream flow forecasts for all members and all initialization times of the meteorological prediction systems, using uncalibrated “raw” ensembles (assuming that a bias correction has been performed by the data providers before, if necessary). Based on the generation rules for the meteorological sub-ensemble, the simulations can be recombined and weighted to build a stream flow ensemble. Here, the hydrological model is computed with a default pa-

rameter set, which proved to be efficient for historic flood events. If we assume that the chosen model structure has sufficient predictive capabilities and input uncertainty is expressed by the meteorological sub-ensemble, two additional sources of hydrological uncertainty are regarded: the initial state of the model and the selection of model parameters (Sect. 2.3). While the m-member super-ensemble is simulated with a single “best” set of hydrological model parameters, the smaller n-member sub-ensemble is simulated with p sets of parameters building up an n*p-member probabilistic stream flow ensemble.

A Bayesian inference mechanism adjusts the weights of the ensemble members when new observed rainfall and discharge data become available during the event. The Bayesian updating procedures can, but not necessarily must reduce the number of members. We aim at successively reducing the number of parameter ensemble members for a proper relation between meteorological and hydrological contributions to total (knowable) uncertainty. Ideally, this procedure sequentially reduces uncertainty by using new information when available. Note that we do not recalibrate the parameters online.

2.2 Meteorological super-ensemble and sub-ensemble

The meteorological ensemble prediction systems COSMO-LEPS, SRNWP-PEPS and COSMO-DE/LAF try to frame the future development of the weather situation by variation of initial states and boundary conditions and/or modifications of the model physics. Per se all forecasts have the same probability. However the probabilistic assessment of ensemble members can be updated by statistical post-processing when observations become available. Nevertheless there is a need to aggregate this information a) to reduce the number of simulations of flood forecast models (or other model chains driven by meteorological input) and b) to compute a kind of best estimation (e.g. the median or a weighted average), which can be used like a deterministic forecast.

The Bayesian model averaging method (BMA, Raftery et al., 2005) assigns a weight and an error distribution to each model by means of iterative optimization. The predicted probability distribution of the complete ensemble is computed by overlay of the error distributions of the single models for the respective forecasted variables. There is a training period necessary for calibrating the BMA. The longest possible training period is still very short (two years) and includes only a few heavy rainfall events. Unfortunately there are no re-forecasts available for the three systems under consideration. For local extreme rainfall events a training of BMA seems not to be possible at all because of the unique nature of such events. There is a danger of over fitting the calibration parameters of the BMA, in particular in the context of flood forecasting.

Thus we developed an alternative ensemble post-processing method. An eventually missing spread of the

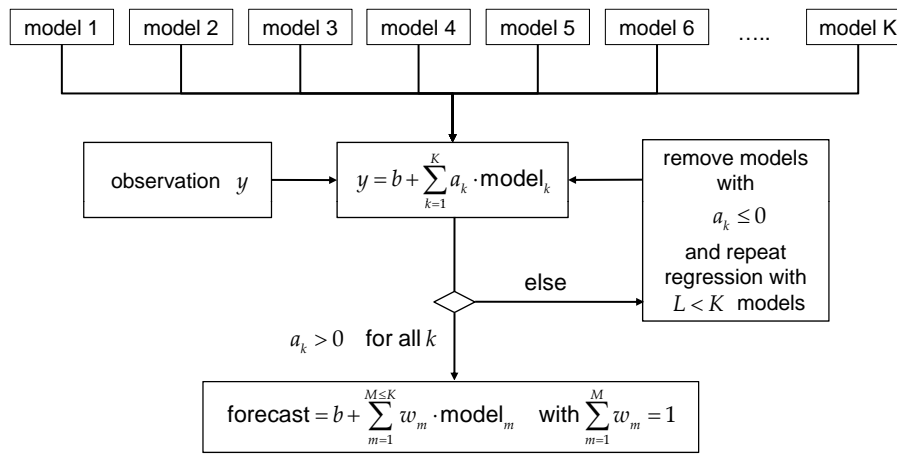


Fig. 4. Workflow of the sub-ensemble calibration procedure.

ensemble (observations are not fully framed) is not compensated by stamping an error distribution on the ensemble but by integrating previous COSMO-LEPS model runs into the ensemble. Even though old forecasts have in general less skill the COSMO-LEPS system is designed to give optimal spread estimates in the medium range (forecast day 3–5). This information can be used to improve short range probability forecasts.

The older COSMO-LEPS model runs normally have a higher root mean square error than the short range forecasts, but their higher spread helps to frame reality, which is an improvement of the basic ensemble, if proper weights are assigned to the lagged forecasts. This is done by an iterative multiple linear regression approach (IMLR) between observations and forecasts. Starting from the full regression with all super-ensemble members only the members with a regression coefficient greater than zero survive the first iteration. This procedure is repeated with the remaining members until all regression coefficients are positive.

The result is a weighting scheme for the remaining members including a bias correction. In Fig. 4 (bottom), f is the weighted mean of the M selected forecasts from the super-ensemble of size K . The term b is the bias correction and the positive coefficients w are normalized such that they sum up to one.

2.3 Combined hydro-meteorological ensembles

The ArcEGMO rainfall-runoff-model was chosen for the OFMS because of its computational efficiency and its process oriented conceptualisation. On the one hand the first estimation of the model parameters can be performed by GIS analyses, which is beneficial when experts perform this task and experience from previous work is available. On the other hand the model has more parameters than less process based conceptual models have. In ArcEGMO, there is an interaction between parameters. Also some parameters are less sen-

sitive in most but not all situations. The simulated runoff is an overlay of several processes, which can not be calibrated separately in case of missing observations for the different runoff components. In the mathematical sense the model is over parameterized. Thus calibration results in a large number of equifinal sets of parameters (i.e. sets of parameters with different values produce model outputs with similar efficiency). Furthermore, the number of flood events available for calibration is limited. The event specific response of the hydrological system may not have been sufficiently observed. Thus parameter uncertainty is an issue of concern.

For operational use of the model near real time, the following main steps are performed:

1. compute initial state from continuous simulation (daily or hourly time steps)
2. update initial state for the preceding seven days (hourly time step)
3. simulate raw ensembles to build up the super-ensemble flood forecast
4. simulate parameter ensembles driven by the members of the reduced meteorological sub-ensemble

Uncertainty in initial conditions (e.g. soil moisture, fast groundwater storage) is reduced by an iterative trial and error system state update procedure, which adjusts the storage representing soil moisture and upper groundwater until the hydrograph of the seven days preceding the expected rainfall is best represented by the model. This initial state update violates continuity, but it is important to tackle uncertainty from prior simulation of the continuous model without performing a recalibration during real-time application of the system. Alternatively, an automated state update is possible with the Ensemble Kalman Filter (Evensen, 2003), which is also implemented in the OFMS. Only in specific situations

(mainly very heavy rainfall at the beginning of the event) the model is very sensitive even against the updated system state. Here an additional state ensemble can be used (this has not been done in the case study).

For framing parameter uncertainty of the hydrological model we propose an ensemble approach. Hydrological parameter ensembles are generated by combination of model parameter sets, which proved to be efficient for simulating flood events in the calibration and validation periods. For the derivation of the generating rules of the parameter ensemble, we combine optimization, stochastic methods and expert knowledge.

In the first step we used multi-criteria optimization (Yapo et al., 1998; Vrugt et al., 2003) to find efficient parameter sets for single historic flood events. The selection and weighting of objective functions allows finding a large set of numerically efficient solutions. Hereby it is possible to differentiate between characteristic types of rainfall events. Among the numerically efficient solutions might be some parameter sets, which have a less realistic composition of runoff components. Here a plausibility check is performed by model experts in order to exclude such parameter sets. After these analyses the non dominating process parameters can be fixed and the dominating parameters can be limited within plausible ranges.

A large number of parameter sets is generated by variation of the dominating parameters with a Monte Carlo experiment. Structural uncertainty of the model produces mixed quality of model efficiency for the different flood events. Thus the parameter ensemble contains a priori defined sets of parameters, which are efficient for different types of past flood events. For a multi-criteria selection of the parameter sets to be integrated in the ensemble we used compromise programming. The ensemble members are weighted according to the a priori expected type of event. If no decision about the type of event is possible, these parameter sets are equally weighted. The parameter ensemble concept is designed for real-time application with the possibility to update the weights of the ensemble members when new information becomes available. The restricted number of possible simulations in real time (we already have to simulate the meteorological ensembles!) restricts the number of members of the parameter ensemble.

In operational forecasting it is possible to assimilate observed stream flow data. These data can be useful for the reduction of model uncertainty. Four general types of update procedure are established (Refsgaard, 1997):

1. correction of input data,
2. update of internal state variables,
3. recalibration of model parameters,
4. direct correction of model output.

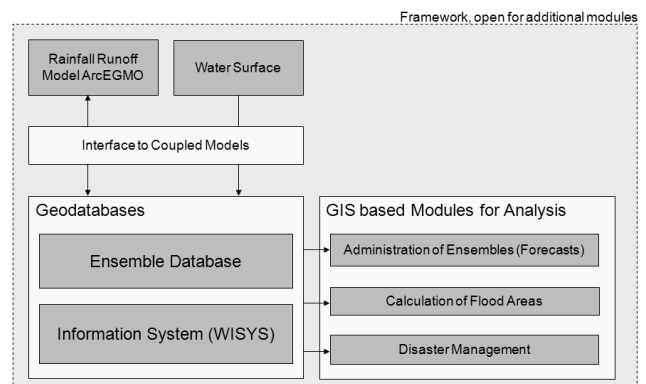


Fig. 5. Methodological concept and components of the operational flood management system (OFMS).

By using meteorological ensembles from external providers, the first approach is not recommended (the meteorologist should use data assimilation and apply a bias correction instead, if required). Within the event, the correction of internal state variables violates mass continuity and may produce unreliable model results in the near future. The online recalibration of model parameters during the event is critical for similar reasons. Thus we apply approach 4. Based on the method described above we generate a parameter ensemble with 20 sets of parameters, which cover the range of observed events. With a single driving meteorological input, we perform 20 simulations of the rainfall-runoff model, which 20-folds the computational demand of the ensemble simulations.

The resulting stream flow ensemble is evaluated against observed stream flow. A likelihood value according to the principle of Bayesian inference (Box and Tiao, 1992) is computed for the time series of each ensemble member (respectively this is the likelihood for each parameter set). Thus the model output produced by the 20 sets of parameters can be weighted according to their likelihood values. From the ensemble members with higher likelihood values one can construct an uncertainty band. Additionally it is possible to compute a weighted average, which can be communicated like a deterministic forecast. After new rainfall observations are available, new runoff simulations are performed with the 20 sets of parameters. After new stream flow observations are available, the likelihood values of these simulations are evaluated for the generation of new weights of the members of the stream flow ensemble.

2.4 OFMS prototype implementation

The software prototype of an operational flood management system (OFMS) integrates data, models and GIS-based tools (Fig. 5). The core components of the OFMS are based on the ESRI ArcGIS software architecture, particularly the DHI-WASY “GeoFES” spatial decision support system for

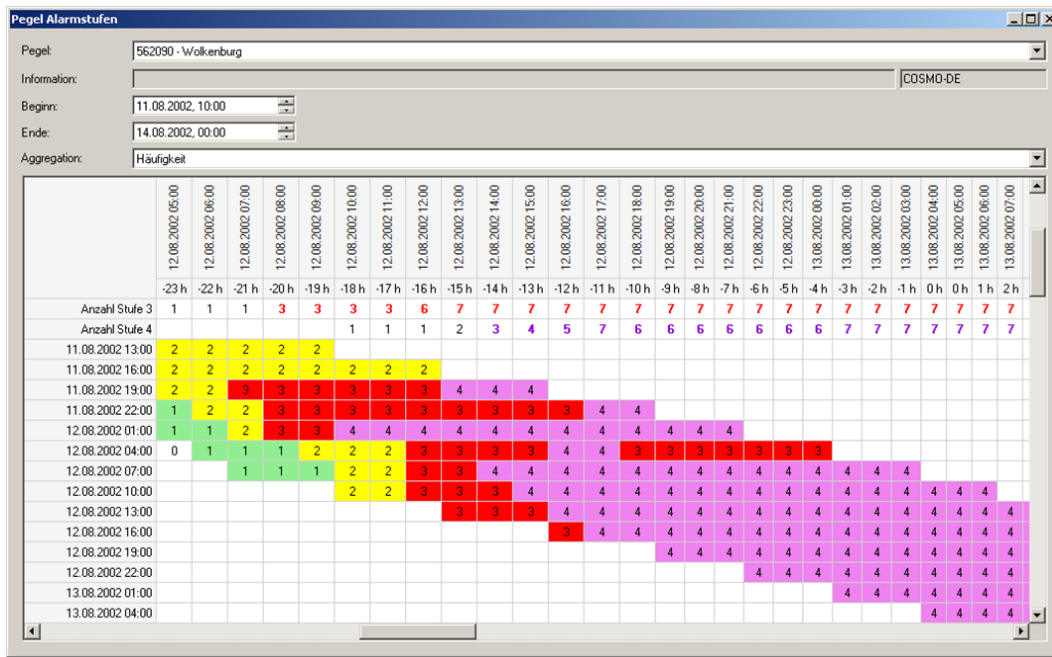


Fig. 6. Persistence chart as presented by the OFMS user interface.

emergency management. The OFMS interface assists the user in the intelligent control of hydrological and hydraulic simulations as well as spatial analyses and the visualization of model output. The ArcEGMO model code was not completely integrated in the OFMS for reason of computational efficiency, but the existing ArcEGMO software was redesigned to work as a set of three dynamic link libraries controlled by the OFMS interface. This allows the simulation of forecasts from an a priori defined system state without re-initialization, but with different climatic inputs. Based on the meteorological ensembles, which are transferred from the operational forecasts database to the hydrological model, the OFMS starts the required number of model runs and manages analysis, aggregation, visualization and archiving of results. The OFMS GIS tools support further analyses like the computation of flood areas with a replacement model for hydraulic simulation (Sect. 3.3) and risk analyses.

In the OFMS, required data for flood management is maintained in two geodatabases: the information system and the ensemble database. Based on the UML object model of the WISYS software (Becker et al., 2006), the information system comprises spatial base data and further geodata originating mainly from the flood protection concepts of the respective rivers. Dynamic data for the model applications and their results is maintained in the ensemble database. Amongst other things, it includes discharge values calculated with the ArcEGMO simulation model, derived water levels and flood plains as well as the measured precipitation and discharges which are retrievable from respective databases (of the State Flood Center of Saxony in the Mulde case study).

To connect the simulation model ArcEGMO to the geodatabases, an interface was developed. It guarantees the data pre-processing and data provision necessary for the model, the model initialization and the storage of the simulation results in the ensemble database. Data pre-processing and provision is carried out in two steps. Firstly, elementary areas are defined by intersecting land use, soil data and catchment area boundaries under consideration of elevation. These elementary areas are combined to hydrological response units and then intersected with the grid cells of the respective meteorological ensemble prediction systems. The area proportion of the EPS grid cells in the hydrological response units is then calculated and saved. The initialization of ArcEGMO and the storage of the simulation results occur for all simulations selected in the user interface.

A post-processing tool converts simulated discharges to water levels. It uses the data tables stored in the information system that, for gauge sites, correspond to the valid stage-discharge relation. The predicted water levels are linked to the water levels of the flood alert level of the respective gauges. This enables a forecast of the flood alert level at the gauges for all simulated time steps. As mentioned before, it is still difficult to make a probabilistic judgement of ensemble members. The OFMS prototype offers frequency analysis of the ensemble forecasts. Alert persistence charts (Fig. 6) can be generated for the gauges, aggregating the ample results of an ensemble forecast and reducing them to the information essential for the user. From the alert persistence chart the following information can be deduced:

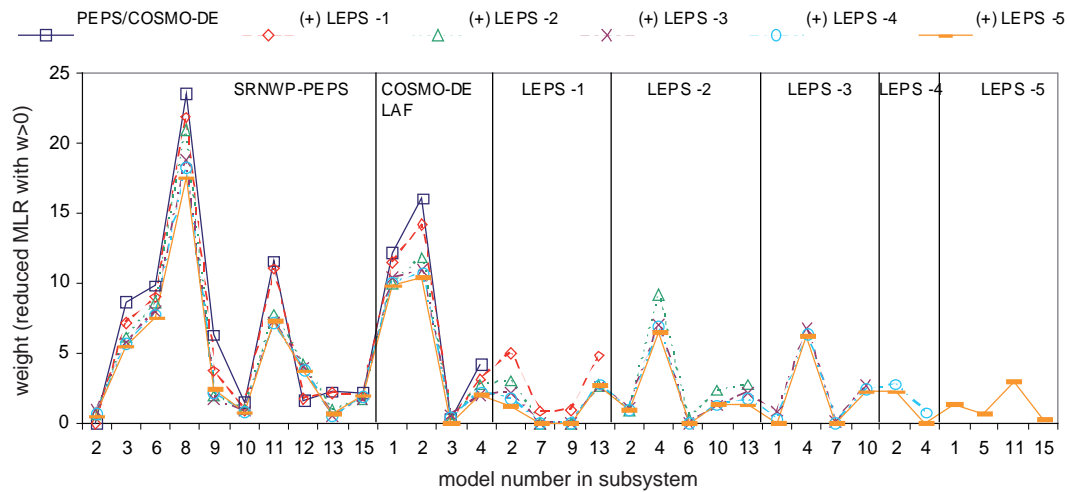


Fig. 7. Weights of the selected members from different super-ensembles by the iterative regression method (IMLR) described in the text. All super-ensembles contain the SRNWP-PEPS (PEPS) forecasts initialized at 00:00 UTC and the COSMO-DE lagged average forecasts from 03:00 and 00:00 UTC on the day of the validation as well as 21:00 and 18:00 UTC from the day before. The model numbers (1–4) in the COSMO-DE LAF ensemble increase with increasing lag-time. Time-lagged COSMO-LEPS forecasts are added (+) to this “basic” super-ensemble, where e.g. –1 indicates the LEPS run from the day before the validation time period.

- The duration (columns, coloured background) of the predicted alert level for each ensemble available for the query time steps (rows),
- for each time step for which the alert levels are determined, the number of ensembles exceeding the alert levels 3 and 4 respectively,
- the time between the time step for which the alert levels are determined and the query time step.

Alert persistence charts are very demonstrative because of their compact and temporal overlapping presentation of the available results at query time. The considerable uncertainty, especially persisting during the rise of a flood, and the stability (persistence) with increasing forecast duration are illustrated clearly. Alert persistence charts are, consequently, an adequate instrument for decision support in ensemble-based flood forecasting.

3 Experimental application

3.1 Meteorological super-ensemble/sub-ensemble demonstration for 2007/2008

For the period from May 2007 to April 2008 a multi-model super-ensemble of accumulated 12 h (06:00 to 18:00 UTC) rainfall forecast for Germany was generated. In the first step the 17 members of the SRNWP-PEPS forecast were combined with the 00:00 UTC COSMO-DE forecast and the median of the COSMO-LEPS forecast from the preceding day to build up a 19 member ensemble with equal weights. After

a bias correction of the single models, the COSMO-DE is one of the best performing models. The COSMO-LEPS median is often the best model regarding the outliers. Thus the 19 member super-ensemble is considered an improvement over the single systems.

To fully implement the idea of adding lagged average forecasts, in a second step, we combined a lagged average multi-model super-ensemble from 4 lagged COSMO-DE, runs (03:00, 00:00, 21:00, 18:00 UTC), up to 5 COSMO-LEPS runs of the preceding 5 days and the SRNWP-PEPS run of 00:00 UTC. The super-ensemble combines up to 101 forecasts (17 SRNWP-PEPS, 4 COSMO-DE and 80 COSMO-LEPS from 5 model runs with 16 members each).

For summer 2007, the ensemble could be reduced by application of the multiple linear regression approach to a sub-ensemble built up by 26 members (Fig. 7). With an independent validation dataset, the forecast quality of the sub-ensemble was better than with other post-processing methods (Schumann and Dietrich, 2009). The calibration of the ensemble improved the forecast in the case study.

3.2 Hydrological parameter ensemble demonstration for the 2002 extreme flood

The parameter ensemble updating procedure (Sect. 2.3) aims at the reduction of forecast uncertainty based on data assimilation. Meteorological uncertainty is propagated through the hydrological model and cannot be completely separated from hydrological uncertainty in real time. As driving meteorological input we used a) the observed precipitation from rain gauges up to the forecasting point and b) afterwards the COSMO-DE very short range forecast, which has a lead time

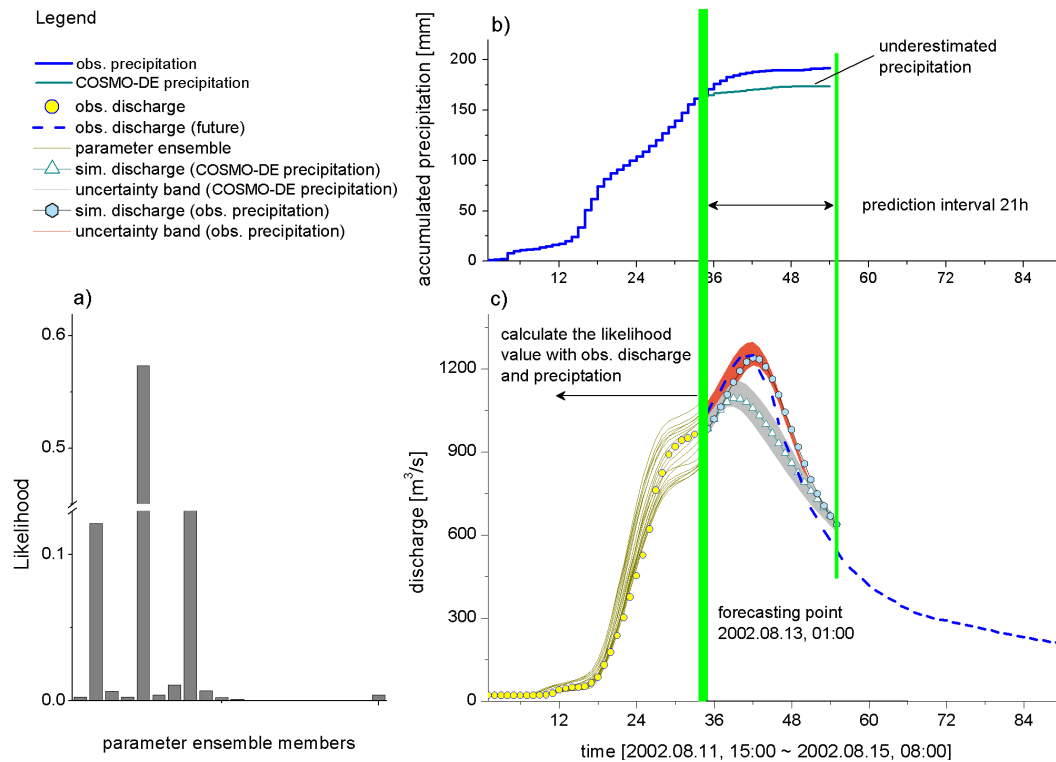


Fig. 8. Parameter ensemble updating procedure shown by example of the 2002 extreme flood event in the Mulde river (Wechselburg 1 gauge). Likelihoods for the parameter sets of the hydrological model are shown in (a), (b) compares accumulated precipitation from observation and deterministic forecast, (c) compares the bands of hydrological parameter uncertainty. Evolving likelihoods reduce the band width of forecasts. The differences resulting from utilisations of observed or forecasted precipitation are also shown in (c).

of 21 h and is initialized every 3 h. The selected forecasting point at 13/08/2002 01:00 UTC is at the end of the rainfall event, but 10 h before the flood peak passes the Wechselburg 1 gauge. The likelihood values of the members of the parameter ensemble are shown in Fig. 8a). It is obvious that the error in the stream flow forecast is mainly caused by the underestimation of precipitation in the forecast (Fig. 8b and c). The uncertainty of rainfall input dominates the uncertainty of the model parameters in this example.

The combination of the meteorological sub-ensemble and the hydrological parameter ensemble adds up to a stream flow ensemble with 520 members, which can be completely updated every 24 or 12 h (when all systems deliver new data) and partly updated every 3 h (when the COSMO-DE lagged average ensemble receives one new member). Due to the ArcEGMO model code optimization described above, the complete update computation of the stream flow ensemble takes about one hour at a personal computer workstation.

3.3 Replacement model for GIS-based inundation forecast

To be able to calculate flood plains and to carry out risk analysis, further data tables (water levels, discharge) are calculated for a sample of points in the watercourses and archived

in the information system. This is done with the help of the hydrological models developed in the OFMS. Simulated water levels are the primary data for the calculation of flood plains (Fig. 9).

The determination of flood plains using 1D-hydrological calculations consists of the interpolation of a water surface and the intersection with a digital elevation model (DEM). A triangulated irregular network (TIN) is created, based on a network of base points located in the potential flood plain. The distance between the base points and the resulting resolution of the TIN determine the degree of detail of the resulting flood plain. When defining the network base points, the DEM applied (being the base data) and the detail requirements of the calculation should be taken into consideration. Interpolating the values of the routed water course, a kilometre specification is assigned to all chosen base points in the flood plain. The generation of the base point network is executed only once.

As the previously selected sample points in the watercourse are situated along the routed watercourse, they also have a kilometre specification, as well as calculated water levels. Joining sample points and network base points through their kilometre specification, water levels can be interpolated and assigned to each base point. The result is a

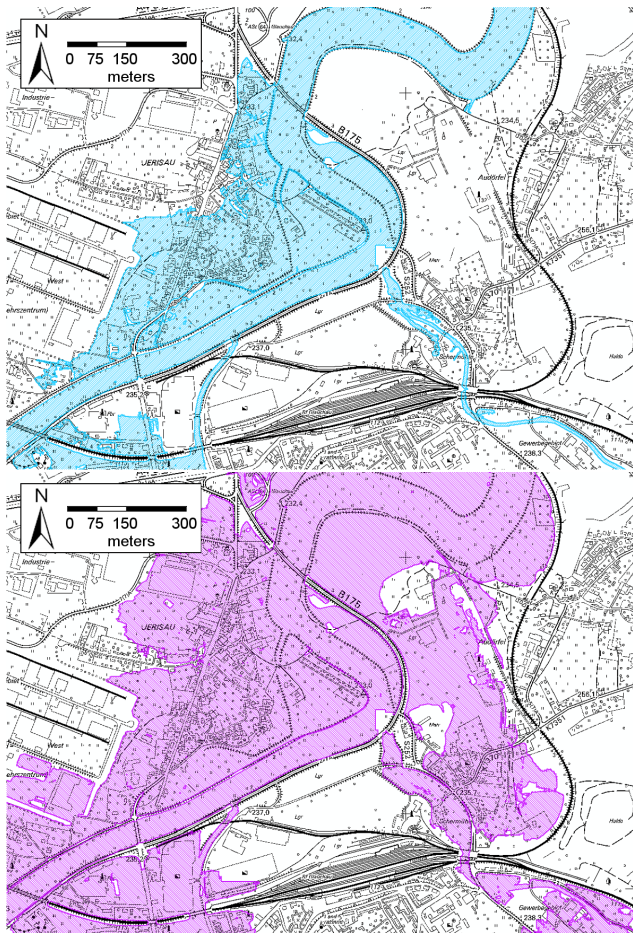


Fig. 9. Calculated flood plain in the area of the mouth of the Lungwitzbach into the Zwickauer Mulde in Glauchau on 12/08/2002, 6 a.m. (top) and 6 p.m. (bottom).

water surface as a TIN, which can then be converted into a GRID. Cells without a direct connection to the water course are identified and marked as “not flooded”. After intersecting this GRID with the DEM, an inundation model is now available as a first result. The contour of this inundation model, in turn, corresponds to the flood extent.

This tool also allows taking culvert constructions and barriers into account. However, within the OFMS this function is not made use of.

The computational demand of the calculation of flood plains is directly related to the resolution of the elevation model as well as to the number of base points used in the calculation. The number of sample points in the watercourses in the test system/prototype with stage-discharge relations adds up to 1600, with 18 000 base points in total. With a spatial resolution of 2×2 m, the calculation time amounts to about 10 min per forecast in the case study. When using a raster of 20×20 m, the calculation time can be reduced by a factor of 10. One must consider that a fine-tuning of the prototype was not conducted.

In the OFMS Mulde, one can choose between two modes: On one hand, it is possible to calculate the flood areas for a specific time step in the forecast period. If this procedure is repeated for various time steps, the forecasted temporal progression of the flood can be visualized. The second mode allows the determination of the maximum extent of the flood plain within the complete forecast period by calculating the maximum water level for each sample point in the watercourse. These two modes can be utilized in the OFMS according to the task in question. With the definition of the maximum flood plain extent during the forecast period a fast overview of all potentially flooded areas is possible. A time step oriented analysis can be applied when generating evacuation plans or assigning roads for provision delivery.

4 Conclusions and discussion

In this study and in the two preceding studies (Dietrich et al., 2008, 2009) we demonstrated the use of medium range to very short range ensembles in flood forecasting for meso-scale river basins. In most of the hindcasts the ensembles were able to frame uncertainty, even at the extreme 2002 flood event (Dietrich et al., 2008). The hindcasts have shown an improvement of the spatial and temporal resolution of flood forecasts compared to conventional forecasts. The extension of lead times is possible, but at the cost of higher uncertainty, which is in turn reflected in the ensembles.

A prototype of an operational flood management system (OFMS) was developed, which combines forecasts from three meteorological systems with different characteristics and different predictive capability. The OFMS supports flood managers and decision makers in managing uncertainty. The prototype demonstrates the applicability of stream flow ensemble forecasts in an operational environment, e.g. at local authorities.

A major drawback of the probabilistic assessment of ensembles is the limited availability of hindcasts (resp. re-forecasts), that means forecasts for already observed events, which have been computed with the operational forecast system afterwards. For the three systems under consideration there was only a two year period from the pre-operational setup of the models available. From this time series it is impossible to perform an a priori assessment of meteorological uncertainty in extreme weather situations. With a longer time series we suppose to be able to assign more reliable weights to the members of the combined super-ensemble, which are valid for the specific large scale weather situation. Despite these limitations we have developed an approach for the probabilistic assessment of the super-ensemble based on a one year time series to demonstrate the potential value of ensemble post-processing for flood forecasting. We are not yet able to validate the method under the conditions of heavy or extreme rainfall. Further work must be related to this issue. Here we see options to integrate likelihoods, derived

from observed data into real-time assessments of ensemble forecasts. For meteorological data this could be done e.g. with radar data to consider the spatial distribution and the movement of convective cells during extreme rainfall events.

In general the integration of observations in the flood management strategy is very important. Observations deliver evidences about the quality of the forecast. Further more they allow the adaptation of the (uncertain) probabilistic assessment of the forecast ensembles by processing all available information. This could support local human forecasters which have to judge unusual situations and modify forecasts based on their familiarity with models and the flood situation in the area of interest (Blöschl, 2008).

The hydrological parameter ensemble is a promising approach for regarding uncertainty of the hydrological model. However, the interaction between parameter uncertainty and input uncertainty cannot be completely resolved during calibration. One can expect that some of the optimized parameter sets perform an indirect bias correction of the input (e.g. by lowering the fast groundwater storage activation level of ArcEGMO). If input uncertainty is completely framed by the meteorological ensemble forecasts, the parameter ensemble generated from historic data (with uncertain rainfall input) may even over estimate the spread of the hydrological uncertainty. The probability distribution of the combined output may be unnecessarily flattened.

The GIS-based inundation model proved to be able to simulate peak inundation reasonably well. The dynamics of inundation is simplified compared to non-stationary hydraulic models (assuming that retention effects are negligible). On the one hand, this adds structural uncertainty. On the other hand, in the context of the OFMS presented here, the forecaster may be more interested in peak inundation at the lead times under consideration. The further development of adaptive schemes a) to improve the combination and assessment of forecasts and b) to reduce the number of ensemble members is promising. Future work should also deal with the spatial heterogeneity of the forecasted variables and of forecast uncertainty.

Acknowledgements. The authors thank their colleagues and project partners for continuing support and discussion. The research reported in this paper was carried out with support from the German Ministry for Education and Research (BMBF) under the initiative “Risk Management of Extreme Flood Events (RIMAX)”. The authors thankfully acknowledge funding. Data and practical experience have been contributed by the State Flood Center of Saxony (Dresden), the Saxonian Reservoir authority (Pirna), the Saxonian land survey authority and the German National Meteorological Service (DWD). The editor thanks the anonymous referees for assisting in evaluating this paper.

Edited by: M. Disse

Reviewed by: two anonymous referees

References

- Ajami, N. K., Duan, Q., and Sorooshian, S.: An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, *Water Resour. Res.*, 43, W01403, doi:10.1029/2005WR004745, 2007.
- Anderson, J. L.: A method for producing and evaluating probabilistic forecasts from ensemble model integrations, *J. Climate*, 9, 1518–1530, 1996.
- Apel, H., Thielen, A. H., Merz, B., and Blöschl, G.: Flood risk assessment and associated uncertainty, *Nat. Hazards Earth Syst. Sci.*, 4, 295–308, 2004, <http://www.nat-hazards-earth-syst-sci.net/4/295/2004/>.
- Becker, A., Klöcking, B., Lahmer, W., and Pfützner, B.: The hydrological modelling system ARC/EGMO, in: *Mathematical models of large watershed hydrology*, edited by: Singh, V. P. and Frevert, D. K., Water Resources Publications, Littleton/Colorado, 321–384, 2002.
- Becker, A., Michels, I., McCurdy, E., Düwel, H., Müller, R., and Timmermann, R.: Informationssystem, in: *Werkzeuge für das integrierte Flussgebietsmanagement. Ergebnisse der Fallstudie Werra – Konzepte für die nachhaltige Entwicklung einer Flusslandschaft*, edited by: Dietrich, J. and Schumann, A., No. 7, Weißensee-Verlag Berlin, 2006 (in German).
- Blöschl, G.: Flood warning – on the value of local information, *Intl. J. River Basin Management*, 6(1), 41–50, 2008.
- Box, G. E. P. and Tiao, G. C.: *Bayesian inference in statistical analysis*, Wiley Classics Library Edition, Wiley-Interscience Publication, John Wiley and Sons 1992.
- Buizza, R., Houtekamer, P. L., Toth, Z., Pellerin, G., Wei, M., and Zhu, Y.: A comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems, *Mon. Weather Rev.*, 133, 1076–1097, 2005.
- Carpenter, T. M. and Georgakakos, K. P.: Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales, *J. Hydrol.*, 329(1–2), 174–185, doi:10.1016/j.jhydrol.2006.02.013, 2006.
- Denhard, M. and Trepte, S.: Calibration of the European multi-model ensemble SRNWP-PEPS. Second THORPEX international science symposium, WMO/TD No. 1355, WWRP/THORPEX No. 7, 2006.
- Dietrich, J., Trepte, S., Wang, Y., Schumann, A. H., Voß, F., Hesser, F. B., and Denhard, M.: Combination of different types of ensembles for the adaptive simulation of probabilistic flood forecasts: hindcasts for the Mulde 2002 extreme event, *Nonlin. Processes Geophys.*, 15, 275–286, 2008, <http://www.nonlin-processes-geophys.net/15/275/2008/>.
- Dietrich, J., Denhard, M., and Schumann, A. H.: Can ensemble forecasts improve the reliability of flood alerts?, *Journal of Flood Risk Management*, accepted, doi:10.1111/j.1753-318X.2009.01039.x, 2009.
- Diomede, T., Davolio, S., Marsigli, C., Miglietta, M. M., Moscatello, A., Papetti, P., Paccagnella, T., Buzzi A., and Malguzzi, P.: Discharge prediction based on multi-model precipitation forecasts, *Meteorol. Atmos. Phys.*, 101, 245–265, 2008.
- Doms, G. and Förstner, J.: Development of a kilometre-scale NWP-system: LMK, edited by: Doms, G., Schättler, U., and Montani, A., *COSMO Newsletter No. 4*, 168–176, 2004.
- Evensen, G.: The Ensemble Kalman Filter: Theoretical formula-

- tion and practical implementation, *Ocean Dynam.*, 53, 343–367, 2003.
- Georgakakos, K. P., Seo, D.-J., Gupta, H., Schaake, J., and Butts, M. B.: Towards the characterization of streamflow simulation uncertainty through multimodel ensembles, *J. Hydrol.*, 298(1–4), 222–241, 2004.
- Hoffman, R. N. and Kalnay, E.: Lagged average forecasting, an alternative to Monte Carlo forecasting, *Tellus*, 35A, 100–118, 1983.
- Kalnay, E.: *Atmospheric modelling, data assimilation and predictability*, Cambridge University Press, 2002.
- Komma, J., Reszler, C., Blöschl, G., and Haiden, T.: Ensemble prediction of floods – catchment non-linearity and forecast probabilities, *Nat. Hazards Earth Syst. Sci.*, 7, 431–444, 2007, <http://www.nat-hazards-earth-syst-sci.net/7/431/2007/>.
- Krzysztofowicz, R.: Bayesian system for probabilistic river stage forecasting, *J. Hydrol.*, 268, 16–40, 2002.
- Molteni, F., Buizza, R., Marsigli, C., Montani, A., Nerozzi, F., and Paccagnella, T.: A strategy for high-resolution ensemble prediction, part I: definition of representative members and global model experiments, *Q. J. Roy. Meteor. Soc.*, 127, 2069–2094, 2001.
- Pappenberger, F., Beven, K. J., Hunter, N. M., Bates, P. D., Gouweleeuw, B. T., Thielen, J., and de Roo, A. P. J.: Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European Flood Forecasting System (EFFS), *Hydrol. Earth Syst. Sci.*, 9, 381–393, 2005, <http://www.hydrol-earth-syst-sci.net/9/381/2005/>.
- Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M.: Using Bayesian Model Averaging to calibrate forecast ensembles, *Mon. Weather Rev.*, 133, 1155–1174, 2005.
- Reed, S., Schaake, J., and Zhang, Z.: A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged locations, *J. Hydrol.*, 337, 402–420, 2007.
- Refsgaard, J. C.: Validation and intercomparison of different updating procedures for real-time forecasting, *Nord. Hydrol.*, 28, 65–84, 1997.
- Schaake, J. C., Hamill, T. M., Buizza, R., and Clark, M.: HEPEX: The Hydrological Ensemble Prediction Experiment, *Bull. Am. Meteorol. Soc.*, 88(10), 1541–1547, 2007.
- Schumann, A. H. and Dietrich, J. (Eds.): *Entwicklung integrierter Lösungen für das operationelle Hochwassermanagement am Beispiel der Mulde*, Schriftenreihe Hydrologie und Wasserwirtschaft No. 23, Ruhr-Universität Bochum, 2009 (in German).
- Smith, M. B., Seo, D.-J., Koren, V. I., Reed, S. M., Zhang, Z., Duan, Q., Moreda, F., and Cong, S.: The distributed model intercomparison project (DMIP): motivation and experiment design, *J. Hydrol.*, 298(1–4), 4–26, 2004.
- 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, 2009, <http://www.hydrol-earth-syst-sci.net/13/125/2009/>.
- Todini, E.: Predictive Uncertainty in Flood Forecasting and Emergency Management, Proc. 17th Congress of the Asia and Pacific Division of the IAHR, Auckland, New Zealand, February 24–27, 2010, accepted paper, personal communication, 2009.
- Toth, Z., Talagrand, O., Candille, G., and Zhu, Y.: Probability and ensemble forecasts, in: *Forecast verification: a practitioner’s guide in atmospheric science*, edited by: Joliffe, I. T. and Stephenson, D. B., John Wiley & Sons, 137–163, 2003.
- Trepte, S., Denhard, M., Göber, M., and Anger, B.: SRNWP-PEPS: some results of verification, Second THORPEX international science symposium, WMO/TD No. 1355, WWRP/THORPEX No. 7, 2006.
- Verbunt, M., Zappa, M., Gurtz, J., and Kaufmann, P.: Verification of a coupled hydrometeorological modelling approach for alpine tributaries in the Rhine basin, *J. Hydrol.*, 324, 224–238, 2006.
- Yapo, P. O., Gupta, H. V., and Sorooshian, S.: Multi-objective global optimization for hydrologic models, *J. Hydrol.*, 204(1–4), 83–97, 1998.
- Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W., and Sorooshian, S.: Effective and efficient algorithm for multi-objective optimization of hydrologic models, *Water Resour. Res.*, 39(8), 1214, doi:10.1029/2002WR001746, 2003.