

# Combination of different types of ensembles for the adaptive simulation of probabilistic flood forecasts: hindcasts for the Mulde 2002 extreme event

J. Dietrich<sup>1</sup>, S. Trepte<sup>2</sup>, Y. Wang<sup>1</sup>, A. H. Schumann<sup>1</sup>, F. Voß<sup>1,\*</sup>, F. B. Hesser<sup>1,\*\*</sup>, and M. Denhard<sup>2</sup>

<sup>1</sup>Institute of Hydrology, Water Resources Management and Environmental Engineering, Ruhr University Bochum, Bochum, Germany

<sup>2</sup>Deutscher Wetterdienst DWD (German National Weather Service), Offenbach, Germany

\* now at: Center for Environmental Systems Research, University of Kassel, Kassel, Germany

\*\* now at: Bundesanstalt für Wasserbau (Federal Waterways Engineering and Research Institute), Hamburg, Germany

Received: 24 August 2007 – Revised: 3 December 2007 – Accepted: 13 February 2008 – Published: 19 March 2008

**Abstract.** Flood forecasts are essential to issue reliable flood warnings and to initiate flood control measures on time. The accuracy and the lead time of the predictions for head waters primarily depend on the meteorological forecasts. Ensemble forecasts are a means of framing the uncertainty of the potential future development of the hydro-meteorological situation.

This contribution presents a flood management strategy based on probabilistic hydrological forecasts driven by operational meteorological ensemble prediction systems. The meteorological ensemble forecasts are transformed into discharge ensemble forecasts by a rainfall-runoff model. Exceedance probabilities for critical discharge values and probabilistic maps of inundation areas can be computed and presented to decision makers. These results can support decision makers in issuing flood alerts. The flood management system integrates ensemble forecasts 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.

The aforementioned flood forecast techniques have been applied to a case study. The Mulde River Basin (South-Eastern Germany, Czech Republic) has often been affected by severe flood events including local flash floods. Hindcasts for the large scale extreme flood in August 2002 have been computed using meteorological predictions from both the COSMO-LEPS ensemble prediction system and the deterministic COSMO-DE local model. The temporal evolution of a) the meteorological forecast uncertainty and b) the probability of exceeding flood alert levels is discussed. Results from the hindcast simulations demonstrate, that the sys-

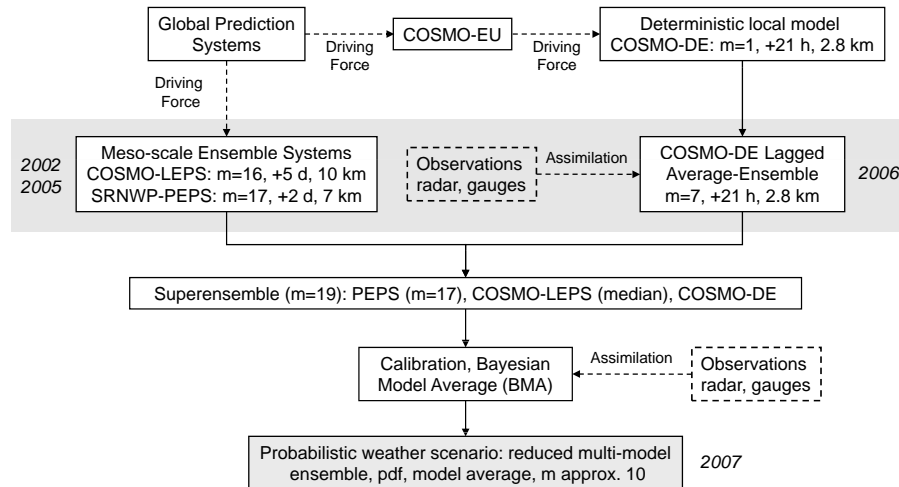
tems would have predicted a high probability of an extreme flood event, if they would already have been operational in 2002. COSMO-LEPS showed a reasonably good performance within a lead time of 2 to 3 days. Some of the deterministic very short-range forecast initializations were able to predict the dynamics of the event, but others underpredicted rainfall. Thus a lagged average ensemble approach is suggested. The findings from the case study support the often proposed added value of ensemble forecasts and their probabilistic evaluation for flood management decisions.

## 1 Introduction

Uncertainties in flood forecasting mainly result from incomplete knowledge of the further meteorological development and from uncertainties of hydrological and hydraulic modelling. Known and knowable sources of uncertainty are the availability and quality of input data, respectively initial and boundary conditions for the models, as well as model parameters and model structure. Inaccurate human interaction and technical problems may also affect the output of a flood prediction chain. The highly nonlinear behaviour of the atmospheric system and the land-atmosphere interaction adds unknowable sources of uncertainty. Thus a perfect weather forecast is impossible (Lorenz, 1963). Resulting from these uncertainties it is not possible to issue a perfect flood forecast.

During the last decades modelling and forecasting techniques evolved from a deterministic towards a probabilistic paradigm. Uncertainty estimation in forecasting aims at framing the possible future development, admitting and communicating the imperfection of the forecast. Exceedance probabilities for threshold values (e.g. critical discharge

Correspondence to: J. Dietrich  
(joerg.dietrich@rub.de)



**Fig. 1.** Meteorological ensembles from different sources used for an ensemble-based operational flood risk management strategy. The number of ensemble members is denoted by  $m$ , the other values indicate the lead time and the horizontal spatial resolution of the ensemble prediction systems.

levels causing inundation) can be provided to flood managers and decision makers.

Ensemble techniques are suitable methods for producing probabilistic forecasts (Anderson, 1996; Kalnay, 2002; Toth et al., 2003). In the context of flood management, ensembles are a group of alternative scenarios of possible future development of the hydro-meteorological situation. Different types of ensembles can be classified according to the generating mechanisms (for meteorological as well as for hydrological applications):

- single system ensembles: perturbation of initial and boundary conditions, different convection schemes (physically based ensembles), perturbation of model parameters;
- multiple systems or multi-model ensembles (“poor man ensembles”): combination of simulations from different models (e.g. Georgakakos et al., 2004);
- lagged average ensembles: combination of current forecasts with forecasts from earlier model runs (Hoffman and Kalnay, 1983).

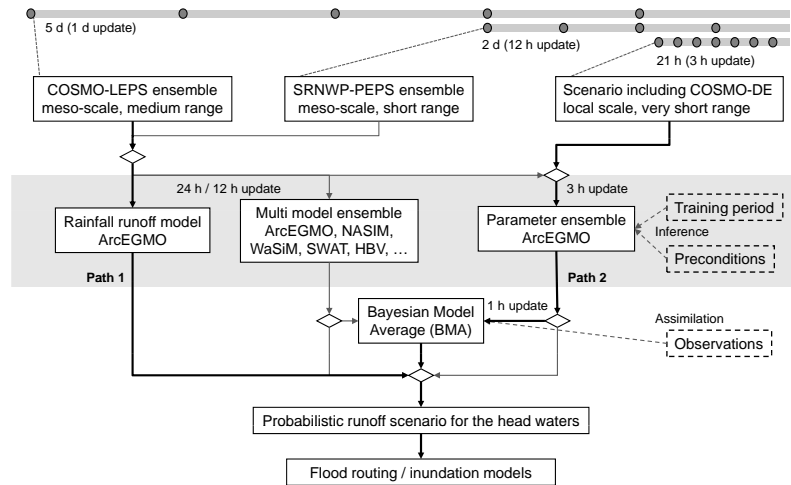
Meteorological ensemble prediction systems (EPS) have become operational. Buizza et al. (2005) compare three global EPS. The development of hydrological applications of ensemble forecasts has started in the late 1990-ies and is subject of ongoing research (e.g. de Roo et al., 2003; Mullusky et al., 2004; Gouweleeuw et al., 2005; Verbunt et al., 2006; Reed et al., 2007).

Beven and Binley (1992) published seminal work about uncertainty estimation in hydrological modelling. Most of the techniques applied in hydrologic modelling need hundreds or even thousands of model runs and often produce

a lot of information, which is not useful for decision oriented operational flood management. Ensemble techniques aiming at the representation of hydrological uncertainty with only a few (in the dimension of 10) members still have to be developed. Methods of combining the simulations of multiple models are subject of current research (e.g. Bayesian Model Averaging, Hoeting et al., 1999; Raftery et al., 2005; Sloughter et al., 2006). Recently Ajami et al. (2007) presented a study about the computation and combination of hydrological multi-model ensembles to account for model structure uncertainty. They list relevant sources in this field. Interdisciplinary studies dealing with the probabilistic assessment of the flood forecast chain have been published e.g. by Krzysztofowicz (2002), Apel et al. (2004) and Pappenberger et al. (2005).

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 streamflow forecast systems. In Europe, the probabilistic Flood Alert System (EFAS) is under development (Thielen et al., 2008). 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.

Probabilistic forecast systems require the computation of a large number of model runs within a time frame of a few hours. Here computational resources still constrain the possibilities of probabilistic forecasts in an operational real-time environment. However the application of numerical models on personal computers can profit from recent developments in parallelization. Flood forecasts for large river basins can now be simulated on workstations, which can be operated by local water management authorities.



**Fig. 2.** Simulation of hydrological ensembles within an adaptive flood management strategy. Model averaging and ensemble updating are optional branches in the flow of actions.

The aim of this paper is to present an ensemble based operational flood management strategy, which integrates methodological developments in ensemble techniques. A prototype of a corresponding flood management system demonstrates the adaptive coupling and control of forecast models. In the subsequent sections of this paper we summarize the main characteristics of the operational meteorological ensembles from different sources, which are used as input for a rainfall-runoff-model. Whereas the conceptualization of the presented flood management strategy covers the complete flood management chain from rainfall to risk, our case study presented in this paper deals with the probabilistic forecast of the rainfall-runoff process in head waters. As an example we present and discuss results of ensemble hindcasts for the 2002 extreme flood event in the Mulde river basin, which is one of the fastest reacting river systems in the Central European low mountain range.

## 2 Conceptualization of an ensemble-based operational flood forecast

A flood forecast system for operational application is a complex set of coupled models, sensors and databases. When designing such a system, a compromise between computational efficiency, availability of data, predictive capability of the models and the cognitive burden for the flood manager has to be negotiated. 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.

We have designed and partially tested an ensemble based strategy for medium to very short-range flood forecasts. In Figs. 1 and 2, we give an overview of the pathways to combine the components in an adaptive manner in order to manage uncertainty. Two main perspectives are adopted: on Fig. 1 in view of the components of the meteorological forecast and on Fig. 2 in view of the main activities in hydrological modelling and flood management.

The proposed strategy combines meteorological medium-range forecasts (3 to 5 days lead time), short-range forecasts (1 to 2 days lead time) and very short-range forecasts (<1 d lead time) from different operational prediction systems. From these systems a superensemble for the very short-range can be generated and calibrated with the Bayesian Model Average method to produce a high resolution probabilistic weather scenario (Fig. 1, detailed description of the prediction systems in Sect. 3). Hydrological and hydraulic models are driven by meteorological input scenarios provided by the different forecast models. Rainfall-runoff models can simulate ensemble forecasts of river flow at several points of interest like gauges and vulnerable sites. We decided to choose a conceptual hydrological model as the default component for transformation of climatic input into runoff. 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 computation time (Carpenter and Georgakakos, 2006; Smith et al., 2004; Ajami et al., 2004).

The adaptive ensemble simulation strategy shown in Fig. 2 has two default pathways, which can be followed for the entire catchment depending on the hydro-meteorological situation and the needs of flood managers. Path 1, shown on the left, is operated at lead times of 3 to 5 days. The lead time could be extended to 10 days using predictions

from global systems, but the spatial resolution of these systems restricts their applicability in meso-scale flood forecasting. Medium-range forecasts provide the basis for decisions about reservoir management and early warning previous to a potential extreme flood event. Here one hydrological model (ArcEGMO in the case study) is computed with a default parameter set, which proved to be efficient for historic flood events. Hydrological uncertainty is considered low compared to meteorological uncertainty. The latter is mainly represented by the spread of the ensemble. In case of a predicted extreme event, hydrological uncertainty becomes more crucial. Due to limited predictive capabilities of the conceptual model in extrapolation, a second branch of the workflow offers a multi-model computation for the head waters to compute an ensemble of runoff generation. A multi-model ensemble is computationally and cognitive very demanding. Thus it can only be implemented for a few sub-catchments in practice. For the two Mulde sub-catchments “Upper Zwickauer Mulde/Eibenstock reservoir” and “Schwarze Pockau” models from different research and consulting activities are available. These models range from more physically based (WaSiM-ETH), conceptual with strong physical meaning (SWAT, ArcEGMO) to more conceptual models (NASIM, HBV).

Short to very short-range forecasts (path 2 in Fig. 2) are used for issuing flood alerts and planning of tailored flood defence measures. Here the attention of flood managers is focused on a variety of aspects and the computational and cognitive resources are limited. Nevertheless decision makers need more detailed information about relevant criteria like peak time, peak discharge and possible inundation areas. Here only one conceptual model is used and updated in 1–3 hourly intervals including assimilation of observed rainfall and discharge. If we assume that the chosen model structure has sufficient predictive capabilities and input uncertainty is expressed by the meteorological ensemble, two additional sources of hydrological uncertainty are regarded: the initial state of the model and the selection of model parameters.

Uncertainty in initial conditions (e.g. soil moisture) 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. The updated state will only be kept for the event simulation to reduce error in flood forecast. After the event took place, the calibration of the continuous model is revised. The Ensemble Kalman filter (Evensen, 1994) will be implemented for automation of state updating.

For framing parameter uncertainty of the hydrological model we propose an ensemble approach. Hydrological pa-

rameter ensembles are generated by combination of model parameter sets, which proved to be efficient for simulating flood events in the calibration and test periods. We classified the historic flood events into four types of hydrological response, mainly depending on maximum intensity and total amount of rainfall as well as snow cover. A parameter ensemble combines up to 20 simulations using different parameter sets, which are chosen from the best performing sets for each group obtained by a priori Monte-Carlo simulations. These 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 combination of the different types of ensembles adds up to a discharge ensemble with about 100 members.

A Bayesian inference mechanism adjusts the weights when new 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.

Each of the two pathways can be simulated in two different modes: hindcast and forecast. Path 1 has been used for hindcast simulations as shown in the case study (Sects. 5 and 6), when the hydrological model can be calibrated on observations. The hindcast mode is also used for test and further improvement of the system including an update of the knowledge base of efficient parameter sets.

### 3 Meteorological ensemble prediction systems for Central Europe

Our flood management system aims to combine early warnings with dynamical refinement strategies of precipitation forecasts in the short and very short-range. For medium-range forecasting (3 to 5 days) and early warnings the COSMO-LEPS is used, which performs a dynamical downscaling of the global EPS operated by ECMWF (European Center for Medium-Range Weather Forecasting). The short-range SRNWP-PEPS combines 23 deterministic forecasts from 21 national meteorological services of which 17 cover the Mulde catchment in a poor man’s EPS. It concentrates the most sophisticated knowledge in limited area numerical modelling on the meso-scale and has a lead time of 48 h. For the incorporation of forecast refinements with very high horizontal resolution and improved quantitative precipitation forecasts we use the COSMO-DE model of the National German Weather Service (DWD). This model is among the new generation of convection resolving models and is running operationally every three hours with 2.8 km horizontal resolution.

### 3.1 COSMO-LEPS

The COSMO-LEPS is a limited area EPS, developed within the COSMO (Consortium for Small-scale Modelling) to improve the predictability of extreme weather events, especially when orographic and meso-scale-related processes play a crucial role. COSMO is a consortium involving Germany, Italy, Switzerland, Greece and Poland which aims to develop, improve and maintain the non-hydrostatic limited-area COSMO model. The LEPS methodology (Limited area Ensemble Prediction System) combines the members from the 51-member ECMWF global EPS into a small number of clusters. A representative member of each cluster provides initial and boundary conditions to run the COSMO model (Molteni et al., 2001; Montani et al., 2003). The added value of the system resides in joining the skill of a global ensemble system to depict the possible evolution scenarios with the capability of a limited area model to improve the descriptions of local meteorological processes. The current COSMO-LEPS suite uses 16 representative members driving 16 COSMO model integrations which generate probabilistic output with lead time 12:00 UTC +132 h. The horizontal resolution is approximately 10 km. Marsigli et al. (2005) show that the COSMO-LEPS system significantly enhances the skill of the ECMWF-EPS in predicting local amounts of precipitation.

### 3.2 SRNWP-PEPS

The EUMETNET Short-Range Numerical Weather Prediction Programme (SRNWP) incorporates the four modelling consortia HIRLAM, ALADIN, COSMO and the UK Met Office. The weather services associated to these consortia produce a reasonable variety of operational forecasts on different domains with different grid resolutions using different model parameterisations or releases and data assimilation techniques. Since 2005 the DWD has been joining all available high resolution numerical forecasts in a poor man's EPS (PEPS, Denhard and Trepte, 2006). This multi-model ensemble currently generates quasi-operational probabilistic forecasts for Europe using 23 different deterministic forecasts provided by 21 meteorological services. The single model forecasts are interpreted on a horizontal reference grid with a spacing of  $0.0625^\circ$  ( $\sim 7$  km) like the COSMO-EU model of DWD. Since the individual members have different spatial resolutions and integration areas, the ensemble size depends on location. For selected meteorological parameters ensemble mean and exceedance probabilities of certain thresholds are calculated at each PEPS grid point from the ensemble members using a nearest neighbour approach. All ensemble members are equally weighted and the probability is calculated by counting the corresponding members at each grid point. The results of the SRNWP-PEPS are distributed to the contributing members on an operational basis four times a day. It can be seen from case studies and prob-

abilistic verification for Germany (Trepte et al., 2006) that this ensemble is a valuable tool for severe weather forecasting. Of course its performance is limited by the contributing models. If these are not capable to forecast a given weather pattern, the SRNWP-PEPS will not be capable as well. A major benefit of this multi-model EPS is the possibility to compare the behaviour of all operational European limited area models.

### 3.3 COSMO-DE

Since Spring 2007, the newly developed numerical weather prediction system COSMO-DE (prior name LMK) for very short-range forecasts up to 21 h and with a resolution on the meso- $\gamma$  scale is added to the operational meteorological model chain at the DWD (Doms et al., 2004; Steppeler et al., 2003). The emphasis of this model system lies in the prediction of severe weather events related to deep moist convection and to interactions of the flow with small scale topography. COSMO-DE couples to the models COSMO-EU (meso- $\beta$ ) and GME (global). One of the most important changes from meso- $\beta$  to meso- $\gamma$  resolution is the abandoning of a parameterisation of deep convection. Such a model needs special requirements concerning data assimilation: at this scale highly resolved, rapidly updated observations are needed, which are delivered by precipitation radar data with a horizontal resolution of roughly 1 km. They are assimilated by the latent heat nudging approach (Klink and Stephan, 2005). The assimilation cycle leads to new forecasts every 3 h. There is not much experience with the skill of COSMO-DE in forecasting extreme precipitation events on the convection permitting scale. A hindcast of the 2002 event is presented in Sect. 6.2. Under the assumption that the synoptic scale weather patterns, which force the COSMO-DE at the boundaries, are quite stable with respect to this high frequency of new initialisations, it is reasonable to join the time lagged deterministic runs of COSMO-DE in a lagged average forecast ensemble. Recently, DWD started the development of a physical single model ensemble based on COSMO-DE, the COSMO-DE-EPS, which will be operational in 2011.

## 4 The rainfall-runoff model ArcEGMO

For the Mulde case study presented in this paper, the hydrological model ArcEGMO (Becker et al., 2002) was adapted to the needs of operational flood forecasting. ArcEGMO is a GIS-based modular modelling system containing several sub-models. It can be characterized as a conceptual model, whose parameters have a physical meaning.

The catchment is partitioned into homogenous units (hydrotopes), for which the vertical and horizontal processes are simulated. For a consideration of the spatial variability of land use and soil related properties, areal distribution functions for essential parameters like saturated conductivity,



**Fig. 3.** Topographic map of the Mulde river basin showing the sub-catchments and points of interest mentioned in the text.

field capacity, etc. are applied. Runoff generation is calculated with an empirical formulation for the infiltration capacity (Holtan, 1961) based on the saturated conductivity for the upper most soil layer and the relative soil moisture deficit. Parameterization is therefore directly related to mapped soil characteristics and related attributes. The unsaturated zone in the soil is basically described by two virtual moisture layers which are filled by infiltration and depleted by evapotranspiration. The total thickness of the unsaturated zone is defined as the minimum either of the root depth, the groundwater depth or the soil thickness. Runoff concentration at the surface is modelled with a kinematic wave approach. For the description of the lateral subsurface processes the hydrotopes are reaggregated to different types of hydrological behaviour: e. g. uplands with a deep aquifer supplying mainly stream base flow, hillslopes with dominating lateral interflow and shallow groundwater areas with a more complex groundwater-surface water interaction (e.g. riparian zones, wetlands). Subsurface flow is modelled with a system of linear reservoirs for each of the groups of hydrotopes. A second fast reservoir can be activated when the water level in the main reservoir exceeds a threshold value (Becker et al., 2002). This reservoir system allows the spatially distributed

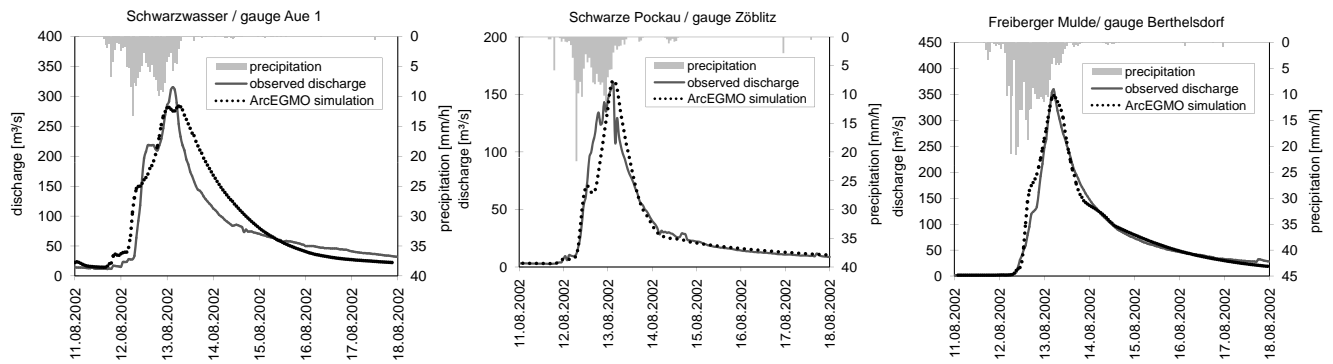
definition of storage constants for slow and fast runoff components, which can be calibrated from measured discharge records via recession curve analysis. Subsurface flow builds up about 90% of the hydrograph, mainly depending on the preconditions of the river basin and the characteristics of the rainfall event. A sensitivity analysis (Wang et al., 2007) revealed that the model parameters for the fast reservoirs are most relevant for the calibration of runoff in the headwaters during the summer season. These parameters can be used as lumped or distributed parameters. Channel routing can be simulated with the Kalinin-Miljukov method or a linear reservoir cascade.

The system states of a continuously operated model with daily time steps can be used as initial conditions for a flood event model operating on sub-daily (commonly hourly) time steps. This event model can be rerun from that state with several parameter sets and different meteorological input (e.g. from ensemble forecasts). So the computational demand is reduced to one single model initialization run and a large number of semi-automated executions of the process computation for a limited number of time-steps. The model can be continuously run in sub-daily resolution within the operational flood forecast system. For the hindcast simulations of historic events, which provide essential knowledge about model parameterization in case of extreme rainfall, only a limited amount of precipitation data is available in sub-daily resolution. Here a switching between different temporal resolutions is advantageous. This strategy allows an efficient integration of ArcEGMO into the computation and optimization of a probabilistic flood forecast chain.

## 5 Study region

The upper Mulde catchment is situated in the Ore Mountains (Germany and Czech Republic, 7400 km<sup>2</sup> total catchment area). It is formed by several parallel sub-basins, draining from South to North. Narrow and steep valleys cause a fast reaction of the watershed and critical superimposition of flood waves. Several cities are located in the flood plain of the lower Mulde river basin (Fig. 3). During west-cyclonic rainfall events, which caused several extreme flood events in the past, the uncertainty of precipitation forecasts in location, time and volume is crucial. Thus the reliability of flood alerts is an issue of concern.

In the last 100 years six local short-time extreme events with catastrophic effects (sub-catchments up to 200 km<sup>2</sup>) and several large scale flood events have been observed in the Mulde river basin. The most recent local extreme event was the Marienberg flood 1999 caused by 144.6 mm rainfall in 90 min over a small sub-catchment (Büttner et al., 2001). The recurrence period of the recorded rainfall intensity is considerably higher than 100 years (63.4 mm for 90-min rainfall at Marienberg according to Bartels et al., 2005). After two extreme floods in 1954 and 1974, the Mulde river basin was



**Fig. 4.** Calibration results for the rainfall-runoff model for sub-catchments located in the western, middle and eastern part of the Ore Mountains (model time step 1 h).

again affected by an extreme flood event in 2002 (Elbe flood). The flood was caused by a so called “Vb” atmospheric circulation pattern, where a low-pressure system moves from the Atlantic southeast to the Mediterranean and then turns northeast across the Alps toward Central Europe. There it causes heavy and intense rainfall, especially in the Central European mountain ranges (Becker and Grünwald, 2003). In Zinnwald-Georgenfeld, close to the upper Freiberger Mulde (Fig. 3), the largest 24-h rainfall ever recorded in Germany was 312 mm on 12 August 2002.

At the occurrence of the 2002 extreme flood event, the ensemble prediction systems used in this case study have not yet been operational. Hindcasts have been computed by rerun of the COSMO-LEPS and the COSMO-DE models. Within the next section we present first results.

## 6 Results

### 6.1 Rainfall-runoff model

For the setup of the Mulde catchment model, a digital elevation model, soil data, land use data and discharge time series were provided by local authorities. Climate time series with daily time steps were used to calibrate and test a continuous model for the period 1950–2006. A single best parameter set has been obtained by calibration with iterations between manual adjustment of parameters and Monte Carlo simulations. We paid specific attention to low flow periods, peak flows and recession curves to minimize errors in initial conditions and to provide a reliable first guess for default parameters. System states are saved by the long-range mode model and loaded as starting values by the short-range mode model. Six summer flood events were calibrated on observed rainfall from 10 to 19 recording climate stations (Fig. 3) in high temporal resolution (hourly time step), namely 2002 (recurrence period  $T$  up to  $>500$  a), 1995 ( $T$  approx. 10a) and four flood events with  $T < 5$  a (1995–1998, Wang et al., 2007). Event specific parameter sets and a common “best” compromise

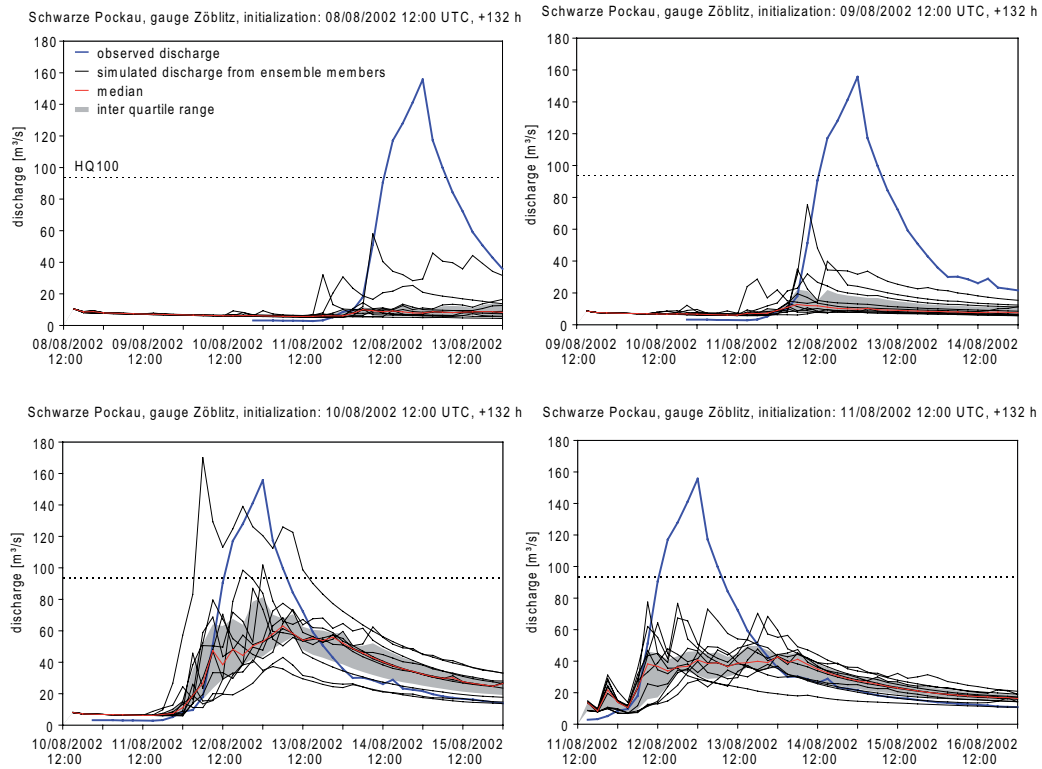
parameter set for all calibrated events have been obtained by iterations between expert and Monte Carlo simulations. A test of the event model has been performed on four flood events in the 1950-ies and 1980-ies.

In this paper we focus on processing of the meteorological ensembles by the hydrological model, but not on hydrological uncertainty. We use the calibration for the 2002 flood event to simulate ensembles of discharge driven by meteorological ensembles. Figure 4 shows the calibration results for three head waters of different hydrological characteristics for the 2002 flood event. Nash-Sutcliffe model efficiency is 0.80 for Aue 1 gauge (Schwarzwasser, 362 km<sup>2</sup>), 0.93 for Zöblitz gauge (Schwarze Pockau, 129 km<sup>2</sup>), and 0.95 for Berthelsdorf gauge (Freiberger Mulde, 243 km<sup>2</sup>). Deviation of peak discharge is 10% for Aue 1 and below 5% for the other two gauges.

The computational efficiency of the adaptive modelling strategy outlined in Sect. 2 could be proven. For the 5442 km<sup>2</sup> catchment area of the Golzern gauge (Fig. 3) and a flood event with 220 hourly time-steps the computation time was only 14 s on a standard personal computer with a 3 GHz Pentium processor.

### 6.2 Stream flow ensemble hindcasts of the 2002 extreme flood in the Mulde catchment

Within the framework of the European Project PREVIEW (<http://www.preview-risk.com>), a rerun of the period between 20/07/ and 31/08/2002 of ECMWF EPS and COSMO-LEPS (10 members) was performed. COSMO-LEPS hindcasts with daily initialization between 07/08/2002 and 12/08/2002 have been used to drive a single ArcEGMO model according to path 1 in Fig. 2 (Sect. 2). The results of the hydrological simulations are ensembles of possible future development of discharge at several gauges within the river basin. Figure 5 shows a sequence of discharge ensembles for the Zöblitz gauge. The hydrographs integrate uncertainty from the meteorological forecast as well as from the hydrological model (cf. Fig. 4). The latter are not explicitly



**Fig. 5.** Sequence of discharge forecasts before the disastrous 2002 flood for the Schwarze Pockau catchment (gauge Zöblitz). COSMO-LEPS hindcasts were initialized at 08/08, 09/08, 10/08 and 11/08/2002 at 12:00 UTC and processed by the hydrological model ArcEGMO (model time step 3 h).

**Table 1.** Temporal evolution of forecasted exceedance probabilities for alert discharge levels at Berthelsdorf gauge (Freiberger Mulde). All COSMO-LEPS (C-L) members are equal weighted.

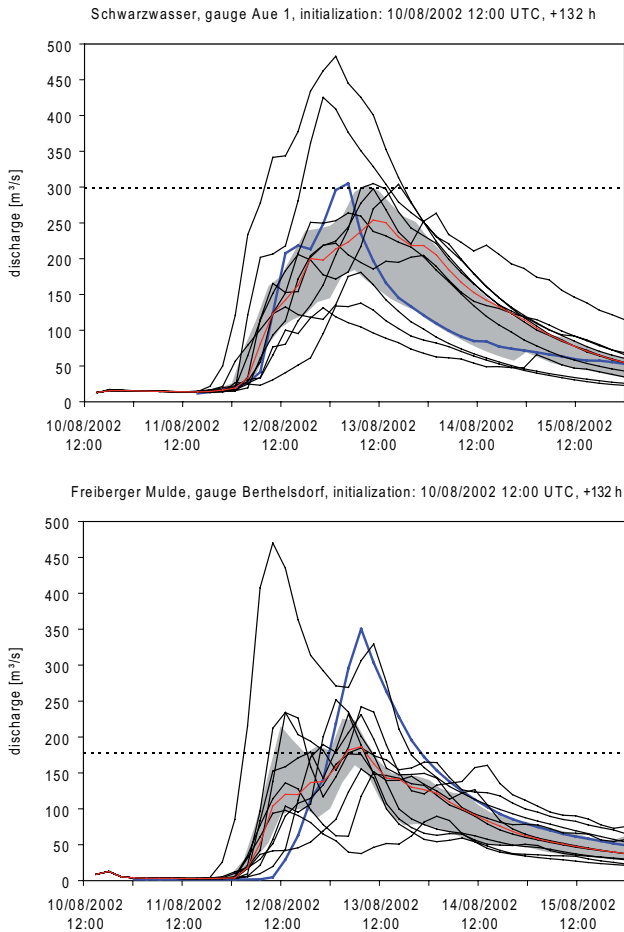
Flood alert levels and activities	C-L 7.8. +5.5d	C-L 8.8. +4.5d	C-L 9.8. +3.5d	C-L 10.8. +2.5d	C-L 11.8. 1.5d	C-L 12.8. 0.5d
1: report	0.5	0.4	0.6	1	1	1
2: control	0.3	0.2	0.5	1	1	1
3: prepare	0.2	0.1	0.4	0.9	1	1
4: defend	0	0	0	0.6	0.2	1

analyzed for this case study, but have been reduced by calibrating the model on observed discharge (which is of course only possible when simulating hindcasts). Between 08/08/ and 10/08/ the median and the spread of the ensemble (shown as interquartile range IQR = 0.75-quantile–0.25-quantile) enlarged. The forecast initialized at 11/08/ indicated that the situation was less tense again. The outlier from the 10/08/ initialization was closest to the discharge peak level observed 2.5 days later. For the Schwarze Pockau catchment (Zöblitz gauge) the event was generally underpredicted (Fig. 5). Sim-

ulations for the Aue 1 gauge (left in Fig. 6) in the Western Ore Mountains show a “good” ensemble embracing the observed runoff, which was close to a 100 year recurrence level. The observed peak level is near the upper limit of the IQR confidence band and relatively close to the median of the ensemble members, but the spread is higher. For the Freiberger Mulde (Berthelsdorf gauge, right in Fig. 6) the flood event was underpredicted. Table 1 lists the probabilities of exceeding flood alert levels at the Berthelsdorf gauge. All COSMO-LEPS members have been equally weighted. The significant increase in probability of exceeding levels 3 and 4 between 09/08/ and 10/08/ is remarkable, as well as the decrease in 11/08/. Even though the forecasts underpredicted peak flow, a probability of 90% for significant inundation (alert level 3) and still 60% for severe damage and danger of life (alert level 4) would have shown a clear evidence of a large flood event two and a half days in advance.

The COSMO-DE model provides deterministic weather forecasts for the very short-range. A rerun for the period around the 2002 flood event has been initialized in a 3 hourly interval with a lead time of 21 h. Figure 7 shows a sequence of discharge simulations driven by different initialization times of COSMO-DE. This combination of forecasts is known as lagged average forecast ensemble (LAF). From



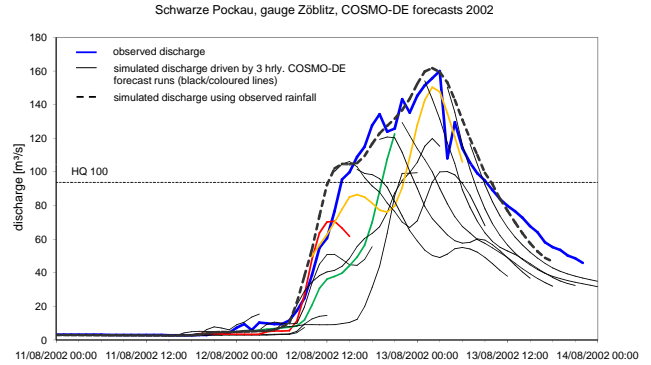


**Fig. 6.** Discharge ensembles for different sub-catchments initialized from COSMO-LEPS hindcasts on 10/08/2002 12:00 UTC with two and a half days lead time to the peak discharge.

the beginning of the event to the initialization of the current forecast, the hydrograph is simulated using combined information from rain gauges and radar observations provided by Haberlandt (2007) for the 2002 case study. When a new forecast is available, the latest system state of the model (driven by observed rainfall) provides hydrological initial conditions for the forecast mode.

The hydrological simulations driven by COSMO-DE forecasts underestimate discharge at the Zöblitz gauge with respect to the amplitude and the duration. In particular, the precipitation forecasts generate discharges of varying quality indicating that the latest COSMO-DE forecast not necessarily gives the best estimation of the observed discharge. This strengthens the need for a LAF approach as shown in Fig. 7.

The COSMO-DE modelling chain predicts two major discharge events. The first peak is strongly underestimated by most runs, while for the second peak at least four runs indicate a strong increase of discharge. Figure 8 shows precipitation patterns from two selected COSMO-DE runs dis-

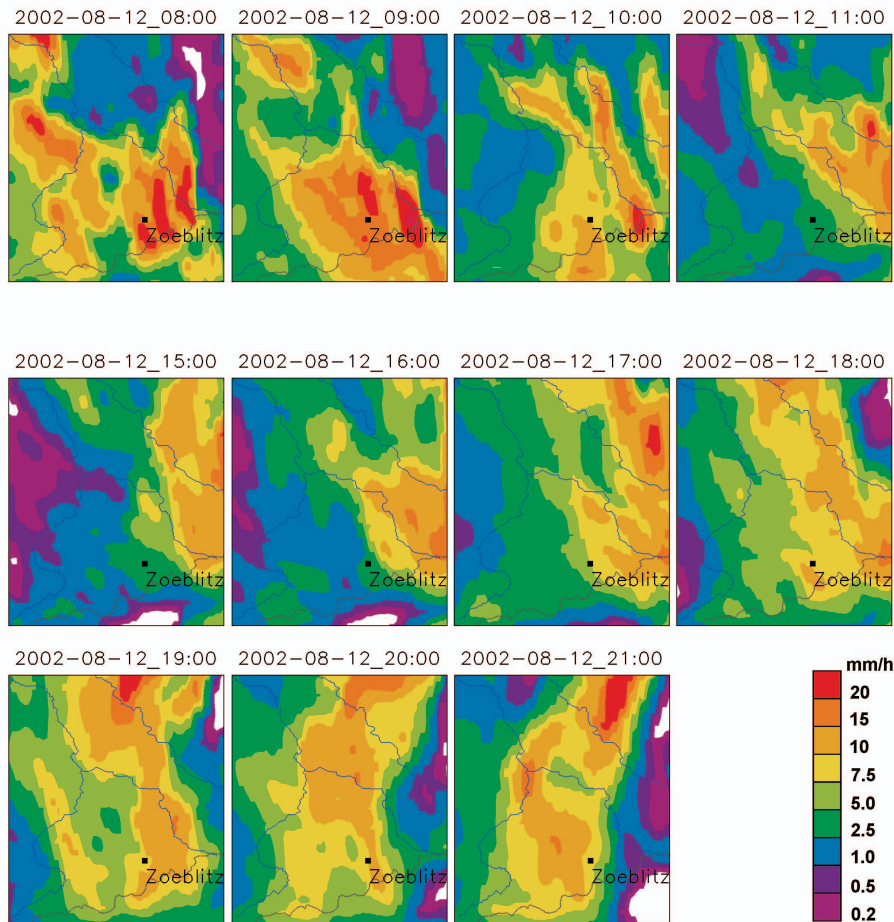


**Fig. 7.** Simulated discharge for the Zöblitz gauge using observed rainfall and COSMO-DE predictions of precipitation (model time step 1 h).

**Table 2.** Exceedance of flood alert levels at Berthelsdorf as forecasted by COSMO-DE. The head row shows the forecast hours from 11/08/2002 20:00. The other rows show the alert levels, which are exceeded by the forecasted discharge for each time step. The two bottom rows show the number of overlapping forecasts, which exceed alert level 3 resp. 4. If the level has been exceeded, the cells are marked as grey.

init. / forecast hr.	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	
11/08/2002 18:00	0	0	1	1	1	0																					
11/08/2002 21:00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12/08/2002 00:00	0	0	0	0	0	0	0	0	0	1	1	1	2														
12/08/2002 03:00	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1										
12/08/2002 06:00	0	0	0	0	0	0	0	0	0	1	1	1	1	2	2	2	2	3	3								
12/08/2002 09:00	0	0	1	1	1	1	2	2	2	3	3	3	3	3	4	4	4	4	4	4	4						
12/08/2002 12:00		1	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4
12/08/2002 15:00			3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
12/08/2002 18:00																											
12/08/2002 21:00																											
13/08/2002 00:00																											
13/08/2002 03:00																											
13/08/2002 06:00																											
13/08/2002 09:00																											
13/08/2002 12:00																											
13/08/2002 15:00																											
13/08/2002 18:00																											
exceed. of level 3	0	0	0	1	2	2	4	4	4	5	5	6	7	7	7	7	7	7	7	7	7	7	7	7	7	7	
exceed. of level 4	0	0	0	0	0	0	1	1	1	2	2	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	

playing the two-peak structure consistently forecasted by all COSMO-DE runs. The first precipitation event lasts approximately four hours showing strong localized precipitation maxima, while the second event incorporates 7 h of heavy precipitation distributed over a much wider area. Due to the more spacious character of the forecasted patterns of the second event, the Zöblitz discharge forecasts are much more reliable. The time-lag between the forecasted two rainfall events was very small (3 h). Thus the hydrographs do not show a noticeable two-peak structure because of the superimposition of the different components of the hydrological system. When comparing COSMO-DE precipitation forecasts with hourly observations from the Haberlandt study (not shown), the observed precipitation patterns show a similar spacious character than the COSMO-DE forecasts of the second peak. In contrast to the COSMO-DE forecasts this spacious character dominates the whole period of more than



**Fig. 8.** COSMO-DE forecasts for the Mulde catchment with forecast initialisations at 11/08/2007 18:00 UTC (first row, red forecast in Fig. 7) and 12/08/2007 00:00 UTC (second and third rows, green forecasts in Fig. 7) showing the first and second precipitation maximum in the Zöblitz catchment, respectively.

20 h of observed precipitation generating the highest ever recorded flood peak at the Zöblitz gauge.

Table 2 shows the persistence of threshold exceedances for flood alert levels at the Berthelsdorf gauge for the deterministic COSMO-DE forecasts. At 12/08/2002 06:00 UTC COSMO-DE gave the first evidence of a potential severe flood event exceeding alert level 3 19 h later (13/08/ 01:00). In the following initializations of the modelling chain the predicted time of exceeding level 3 moved backwards to 12/08/ 15:00 (which is partly due to hydrologic uncertainty, see rising limb of the calibrated hydrograph shown in Fig. 4). The model initializations from 12/08/ 18:00 on gave strong evidence of an extreme flood event. Nevertheless there was always at least one overlapping model run, which did not forecast the exceedance of level 4 ( $210 \text{ m}^2/\text{s}$ ). The observed flood peak was  $360 \text{ m}^2/\text{s}$  at 13/08/ 03:00.

Compared to the results of the forecasts driven by COSMO-LEPS, the performance of the COSMO-DE initializations before 12/08/ 18:00 was surprisingly weak.

The very short-range forecasts did not provide more reliable exceedance probabilities than the medium-range forecasts did. However the very short-range forecasts performed better in estimating the peak discharge during the event (cf. results for the Zöblitz gauge shown above).

## 7 Conclusions

A flood forecast strategy combining meteorological ensembles with different spatial and temporal resolution and a calibrated rainfall-runoff model enable the simulation of probabilistic discharge forecasts for meso-scale catchments. These forecasts can account for different sources of uncertainties, e.g. uncertainties of the initial conditions, model structure and model parameters from both the meteorological and the hydrological models. Data assimilation allows an update of the probabilistic assessment of the ensembles in an operational real-time environment. There is a chance, but no guarantee to reduce uncertainty successively during the event

with an adaptive model control strategy. Compared with deterministic forecasts, probabilistic forecasts of discharge provide additional information to flood managers, who are addressed by the flood management strategy presented in Sect. 2 of this paper.

A part of the proposed flood management strategy has been successfully applied for the simulation of hindcasts for an extreme flood event. The COSMO-LEPS and COSMO-DE hindcasts consistently underpredicted the rainfall rate of the 2002 extreme event. Nevertheless both systems evaluated within this case study produced evidences for a rainfall event which could cause a flood in the dimension of a 100 year recurrence period. If such a probabilistic forecast would already have been operationally available in August 2002, a reliable flood forecast would have been possible with a lead time of 2 to 3 days. Flood management would have been significantly improved compared to the deterministic forecast, which had been issued at that time. Both the COSMO-LEPS and COSMO-DE simulations did not always improve with more recent forecasts.

The derivation of decision rules, e.g. which exceedance probabilities should be used for issuing alerts or initiating flood defence measures, is subject of ongoing research. Within the study area, ensemble hindcasts were only available for one single extreme event. General conclusions about the skill of the models cannot be drawn from this case study. Currently the authors evaluate false alarms to extend their knowledge about the predictive skills of these models. In future work related to the forecast skill of the meteorological models, the study area will be extended in order to evaluate a larger number of extreme rainfall events. Further work of the authors is focused on the hydrological ensemble generation mechanisms. A detailed description and results from evaluation of this method will be published in a subsequent paper. The prototype of an operational system will be extended to the lower Mulde catchment including the adaptive control of flood routing and inundation models at vulnerable hot spots. The probabilistic approach will then enable an explicit consideration of risk and can thus further improve the basis for decision making in operational flood management.

The presented ensemble based flood forecast strategy does not directly address long-term strategic flood management. Nevertheless it is aimed to be an integral part of a sustainable flood management policy. The longer forecasting horizon, the possible reduction of the number of false alerts and a detailed prediction of expected flood damage can improve the preparedness of affected people and thus reduce potential damage.

*Acknowledgements.* 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. B. Pfützner (Büro für Angewandte Hydrologie BAH Berlin) provided source code and support of the ArcEGMO model. 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 National German Weather Service (DWD).

Edited by: S. Vannitsem

Reviewed by: J. Cullmann and another anonymous referee

## 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.
- Ajami, N. K., Gupta, H., Wagener, T., and Sorooshian, S.: Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system, *J. Hydrol.*, 298, 1–4, 112–135, 2004.
- Anderson, J. L.: A method for producing and evaluating probabilistic forecasts from ensemble model integrations, *J. Clim.*, 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/>.
- Bartels, H., Dietzer, B., Malitz, G., Albrecht, F. M., and Guttenberger, J.: Starkniederschlagshöhen für Deutschland, KOSTRA-DWD-2000. Fortschreibungsbericht, Deutscher Wetterdienst – Hydrometeorologie, Offenbach, 2005.
- Becker, A. and Grünewald, U.: Disaster management: flood risk in Central Europe, *Science*, 300, 5622, 1099, 2003.
- 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, 914 pp., 2002.
- Beven, K. J., and Binley, A.: The future of distributed models: model calibration and uncertainty prediction, *Hydrol. Process.*, 6, 279–298, 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.
- Büttner, U., Fügner, D., and Winkler, U.: Das Hochwasser am 5./6. Juli 1999 im Raum Marienberg im Erzgebirge, *Hydrologie und Wasserbewirtschaftung*, 45, 3, 102–111, 2001 (The flood on 5 and 6 July 1999 in the region of Marienberg in the Ore Mountains, Saxonia, in German with English abstract and captions).
- 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.

- De Roo, A., Gouweleeuw, B., Thielen, J., et al.: Development of a European Flood Forecasting System, *I. R. B. M.*, 1, 1, 49–59, 2003.
- 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.
- Doms, G. and Förstner, J.: Development of a kilometre-scale NWP-system: LMK, Doms, G., Schättler, U., and Montani, A. (editors), *COSMO Newsletter* No. 4, 168–176, 2004.
- Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, 99, 10 143–10 162, 1994.
- 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.
- Gouweleeuw, B., Thielen, J., Franchello, G., De Roo, A., and Buizza, R.: Flood forecasting using medium-range probabilistic weather prediction, *Hydrol. Earth Syst. Sci.*, 9, 365–380, 2005, <http://www.hydrol-earth-syst-sci.net/9/365/2005/>.
- Haberlandt, U.: Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event, *J. Hydrol.*, 332, 144–157, 2007.
- Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T.: Bayesian model averaging: A tutorial with discussion. *Statistical Science*, 14, 382–417, 1999, Printing errors corrected in version available at <http://www.stat.washington.edu/www/research/online/hoeting1999.pdf>.
- Hoffman, R. N. and Kalnay, E.: Lagged average forecasting, an alternative to Monte Carlo forecasting, *Tellus*, 35A, 100–118, 1983.
- Holtan, H. N.: A concept for infiltration estimates in watershed engineering, *Agr. Res. Service Pub.*, U.S. Dept. Agr., 41–51, 1961.
- Kalnay, E.: *Atmospheric modelling, data assimilation and predictability*, Cambridge University Press, 512 pp., 2002.
- Klink, S. and Stephan, K.: Latent heat nudging and prognostic precipitation, *COSMO Newsletter*, 5, 124–131, 2005.
- Krzysztofowicz, R.: Bayesian system for probabilistic river stage forecasting, *J. Hydrol.*, 268, 16–40, 2002.
- Lorenz, E. N.: Deterministic Nonperiodic Flow, *J. Atmos. Sci.*, 20, 130–141, 1963.
- Marsigli, C., Boccanera, F., Montani, A., and Paccagnella, T.: The COSMO-LEPS mesoscale ensemble system: validation of the methodology and verification, *Nonlin. Processes Geophys.*, 12, 527–536, 2005, <http://www.nonlin-processes-geophys.net/12/527/2005/>.
- 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.
- Montani, A., Marsigli, C., Nerozzi, F., Paccagnella, T., Tibaldi, S., and Buizza, R.: The Soverato flood in Southern Italy: performance of global and limited-area ensemble forecasts, *Nonlin. Processes Geophys.*, 10, 261–274, 2003, <http://www.nonlin-processes-geophys.net/10/261/2003/>.
- Mullusky, M., Demargne, J., Welles, E., Wu, L., and Schaake, J.: Hydrologic applications of short and medium range ensemble forecasts in the NWS, *Advanced Hydrologic Prediction Services (AHPS), Extended Abstract*, 84, AMS Meeting, Seattle, WA, Jan. 10.–16., 2004.
- 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. Hydrology*, 337, 402–420, 2007.
- Schaake, J. C., Hamill, T. M., Buizza, R., and Clark, M.: HEPEX: The Hydrological Ensemble Prediction Experiment, *Bull. Amer. Meteor. Soc.*, 88, 10, 1541–1547, 2007.
- Sloughter, J. M., Raftery, A. E., and Gneiting, T.: Probabilistic quantitative precipitation forecasting using Bayesian Model Averaging, *Tech. report 496*, Department of Statistics, University of Washington, USA, 2006.
- Smith, M. B., Seo, D.-J., Koren, V. I., Reed, S. M., Zhang, Z., Duan, Q., Moreta, F., and Cong, S.: The distributed model inter-comparison project (DMIP): motivation and experiment design, *J. Hydrol.*, 298 (1–4), 4–26, 2004.
- Steppeler, J., Doms, G., Schättler, U., Bitzer, H. W., Gassmann, A., Damrath, U., and Gregoric, G.: Meso-gamma scale forecasts using the nonhydrostatic model LM, *Meteorol. Atmos. Phys.*, 82, 75–96, 2003.
- 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. Discuss.*, 5, 257–287, 2008, <http://www.hydrol-earth-syst-sci-discuss.net/5/257/2008/>.
- Toth, Z., Talagrand, O., Candille, G., and Zhu, Y.: Probability and ensemble forecasts, in: Joliffe, I. T. and Stephenson, D. B. (eds.): *Forecast verification: a practitioner’s guide in atmospheric science*, John Wiley & Sons, 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.
- Wang, Y., Dietrich, J., Voss, F., and Pahlow, M.: Identifying and reducing model structure uncertainty based on analysis of parameter interaction, *Adv. Geosci.*, 11, 117–122, 2007, <http://www.adv-geosci.net/11/117/2007/>.