Ensemble prediction of floods ? catchment non-linearity and forecast probabilities

J. Komma, C. Reszler, G. Blöschl, T. Haiden

▶ To cite this version:

HAL Id: hal-00299439

https://hal.science/hal-00299439

Submitted on 18 Jun 2008

**Natural Hazards
and Earth
System Sciences**

Ensemble prediction of floods – catchment non-linearity and forecast probabilities

J. Komma[1]**, C. Reszler**[1]**, G. Blöschl**[1]**, and T. Haiden**[2]

[1]Institute of Hydraulic and Water Resources Engineering (Vienna University of Technology), Vienna, Austria
[2]Central Institute for Meteorology and Geodynamics (ZAMG), Vienna, Austria

Abstract. Quantifying the uncertainty of flood forecasts by ensemble methods is becoming increasingly important for operational purposes. The aim of this paper is to examine how the ensemble distribution of precipitation forecasts propagates in the catchment system, and to interpret the flood forecast probabilities relative to the forecast errors. We use the $622\,km^2$ Kamp catchment in Austria as an example where a comprehensive data set, including a 500 yr and a 1000 yr flood, is available. A spatially-distributed continuous rainfall-runoff model is used along with ensemble and deterministic precipitation forecasts that combine rain gauge data, radar data and the forecast fields of the ALADIN and ECMWF numerical weather prediction models. The analyses indicate that, for long lead times, the variability of the precipitation ensemble is amplified as it propagates through the catchment system as a result of non-linear catchment response. In contrast, for lead times shorter than the catchment lag time (e.g. 12 h and less), the variability of the precipitation ensemble is decreased as the forecasts are mainly controlled by observed upstream runoff and observed precipitation. Assuming that all ensemble members are equally likely, the statistical analyses for five flood events at the Kamp showed that the ensemble spread of the flood forecasts is always narrower than the distribution of the forecast errors. This is because the ensemble forecasts focus on the uncertainty in forecast precipitation as the dominant source of uncertainty, and other sources of uncertainty are not accounted for. However, a number of analyses, including Relative Operating Characteristic diagrams, indicate that the ensemble spread is a useful indicator to assess potential forecast errors for lead times larger than 12 h.
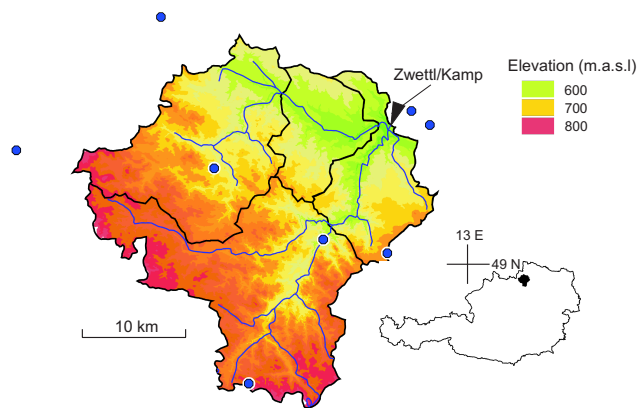
1 Introduction

Quantifying the uncertainty of flood forecasts is becoming increasingly important for operational purposes. This is due to a number of reasons. First, the awareness of the value of uncertainty bounds in flood management has increased. Indeed, it is the uncertainty bounds that will assist flood managers in the trade-off between alternative decisions as they provide information on the likelihood of making less than optimal decisions as a result of forecast errors. Second, flood forecasts are increasingly used for small catchments where the forecast uncertainties tend to be larger than in large catchments. Third, there is a tendency for making forecasts over longer lead times which are associated with larger uncertainties. The most accurate forecasts can be achieved by using observed runoff along with routing models but the forecast lead times are limited to the travel times in the streams. For, say, a $1000\,km^2$ catchment these are on the order of 2 h (Table 1). The values in Table 1 are based on simulation results and hydrograph analyses for various Austrian catchments. Runoff models that use observed precipitation allow to extend the lead times but at the cost of increased uncertainty. Precipitation forecasts allow to further extend the lead times but the uncertainties are still larger.

As the magnitude of the precipitation forecast uncertainty can be large it has been the topic of much recent research. Most of the uncertainty in precipitation forecasts stems from the propagation of small errors in the initial conditions of the atmospheric models (Buizza, 2003). The standard method of estimating this uncertainty is hence to generate an ensemble (or set) of different forecasts of atmospheric processes that differ by their initial conditions (Taylor and Buizza, 2003) in addition to the main (deterministic) forecast. Ensemble forecasts have been operationally issued by the U.S. National Center for Environmental Predictions (NCEP) and the European Centre for Medium Range Weather Forecasts (ECMWF) for more than a decade. Each of the realisations

Table 1. Uncertainties and typical forecast lead times for a 1000 km^2 catchment.

	forecast lead time	forecast uncertainty
River routing using observed runoff	2 h	small
Runoff model using observed precipitation	6 h	medium
Runoff model using precipitation forecasts	48 h	large

Fig. 1. Kamp catchment (622 km^2) with telemetered rain gauges (circles) and stream gauge (triangle) shown. Black lines represent the subcatchments, blue lines the river network. The figure in the lower left corner gives reference about the catchment location in Austria.

(or members) of the ensemble is a possible trajectory of atmospheric processes over the lead time. By examining the distribution of the ensemble one then gets a statistical measure of the forecast uncertainty. The value of making ensemble forecasts lies in the fact that the forecast error changes with time. For some meteorological situations, the likelihood of heavy precipitation will be nil while for others it may be large even though the deterministic forecast does not predict precipitation. More generally speaking, the more the ensemble spread deviates from its climatological mean, the more additional information is provided by the ensemble (Whitaker and Loughe, 1998). Often, the members of the ensemble are assumed to be equally likely and the ensemble spread is assumed to represent the distribution of the forecast errors. However, the statistical interpretation of the ensemble spread is not straightforward. For example, Schaake et al. (2004) analysed the statistical properties of NCEP ensemble precipitation forecasts from 1997–1999 and compared them with measured precipitation. He found that the ensemble forecasts were biased and that the ensemble spread was much smaller than the spread of the error distribution. He proposed methods for bias removal and adjusting the ensemble spread.

Calculating flood runoff from predicted precipitation will modulate the uncertainty of precipitation in two ways. First,

additional sources of uncertainty will come in. These include uncertainties in estimating catchment precipitation and the spatial distribution of precipitation (Siccardi et al., 2005), uncertainties in the soil moisture state of the catchment, as well as uncertainties in the model structure and in the model parameters. Krzystofowicz (2001) presented a formal method of combining the hydrological uncertainties with those of the precipitation forecasts. Second, even when neglecting the hydrological uncertainties, the uncertainty in the flood forecasts will be different from those of precipitation because of the non-linearity of the catchment system. Small inaccuracies can amplify if the system shows strongly non-linear behaviour, for example, if threshold processes are present (Blöschl and Zehe, 2005). However, very little is known on exactly how the uncertainty of precipitation forecasts propagates in the catchment system. Part of the problem is that operational flood management is interested in large floods that tend to exhibit different characteristics from smaller floods, but they are – by definition – rare, so statistical analyses are notoriously limited by small sample sizes.

Given the current issue with ensemble forecasting methods, the aim of this paper is (a) to examine how the ensemble distribution of precipitation forecasts propagates in the catchment system, and (b) to interpret the flood forecast probabilities relative to the forecast errors. We use the Kamp catchment in Austria as an example where an operational flood forecasting system has recently been implemented and a comprehensive data set, including two large floods, is available.

2 Data and methods

2.1 Study area and data

The Kamp catchment is located in northern Austria, approximately 120 km north-west of Vienna. At the Zwettl stream gauge the catchment size is 622 km^2 and elevations range from 500 to 1000 m a.s.l. (Fig. 1). The higher parts of the catchment in the Southwest are hilly with deeply incised channels. Towards the catchment outlet in the Northeast the terrain is flatter and swampy areas exist along the streams. The geology of the catchment is mainly granite and gneiss. Weathering has produced sandy soils with a large storage capacity throughout the catchment. 50% of the catchment

is forested. Mean annual precipitation is about 900 mm of which about 300 mm become runoff (Parajka et al., 2005).

To illustrate the nature of hydrologic response of the Kamp catchment the largest flood events on record and the associated rain events have been analysed using the telemetered rain gauges shown in Fig. 1 and a number of additional rain gauges. Figure 2 shows the event precipitation of these events along with the direct runoff depths. The direct runoff depths were estimated by subtracting baseflow from the event hydrographs that was assumed constant during each event. The events have been ranked according to precipitation. There are two interesting findings. First, for the smallest events only around 10% of rainfall become runoff while the percentage can be much higher for the larger events. During the dry summer months large precipitation depths are necessary to exceed the soil capacity and produce any sizeable runoff as was the case for the extreme event of 8 August 2002 (August 2002a). Clearly, runoff generation is a non-linear process and as the magnitude of the event increases so does the proportion of runoff that is generated. Second, for the same precipitation depth, runoff can vary significantly. As a result of prior snow melt, antecedent soil moisture of the May 1996 event was high which produced a large proportion of runoff. On the other hand, the two July 1997 events had almost the same rainfall as the May 1996 event but much less runoff. There was significant rainfall prior to the 13 August 2002 (August 2002b) event (namely the extreme August 2002a event) which produced more than twice the runoff of the July 1999 event that had similar precipitation but very little antecedent rainfall. It is clear that soil moisture exerts a strong control on runoff response in the Kamp catchment.

In this paper, the analyses of the ensemble forecasts are based on five flood events for which complete data sets of precipitation forecasts were available. These are marked by asterisks in Fig. 2. Details of these events are given in Table 2. The initial moisture state was assessed by examining antecedent rainfall. Both August 2002 events were indeed extraordinary. More details of these events are given in Gutknecht et al. (2002).

2.2 Hydrological model

The model used in this paper is a spatially-distributed continuous rainfall-runoff model (Reszler et al., 2006). The model runs on a 15 min time step and consists of a snow routine, a soil moisture routine and a flow routing routine. The snow routine represents snow accumulation and melt by the degree-day concept. The soil moisture routine represents runoff generation and changes in the soil moisture state of the catchment and involves three parameters: the maximum soil moisture storage FC, a parameter representing the soil moisture state above which evaporation is at its potential rate, termed the limit for potential evaporation LP, and a parameter in the non-linear function relating runoff generation to the soil moisture state, termed the non-linearity parameter
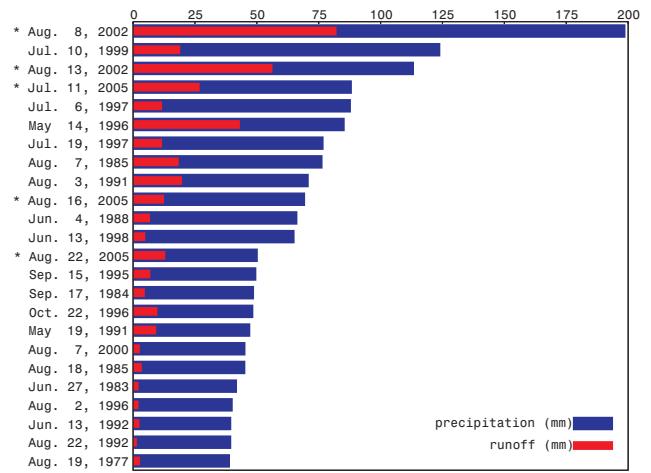

Fig. 2. Event precipitation and direct runoff depths for the largest events on record. Kamp at Zwettl, $622 \, \text{km}^2$. Events marked by asterisks are analysed in this paper.

β. The details of the soil moisture routine are given in Appendix A. Runoff routing on the hillslopes is represented by an upper and two lower soil reservoirs. Excess rainfall enters the upper zone reservoir and leaves this reservoir through three paths, outflow from the reservoir based on a fast storage coefficient K_1; percolation to the lower zone with a percolation rate C_P; and, if a threshold of the storage state LS_{UZ} is exceeded, through an additional outlet based on a very fast storage coefficient K_0. Water leaves the lower zones based on the slow storage coefficients K_2 and K_3. Bypass flow dQ_{by} is accounted for by recharging the lower zone reservoir directly by a fraction of the excess rainfall. K_1 and K_2 as well as C_P have been related to the soil moisture state in a linear way. The outflow from the reservoirs represents the total runoff Q_g on the hillslope scale. These processes are represented on a $1 \, \text{km} \times 1 \, \text{km}$ grid. The model structure and the model parameters have been identified by a five step procedure using field data, comprehensive hydrographic data as well as qualitative evidence during floods (Reszler et al., 2006).

Runoff routing in the stream network is represented by cascades of linear reservoirs with parameters n (number of reservoirs) and k (storage coefficient) that are a function of runoff. Decreasing travel times with increasing flood levels are represented by linearly decreasing k with runoff over a certain range but as the flood water exceeds bank full discharge, k is decreased to represent flood attenuation on the flood plains. The parameters have been found by calibration against observed hydrographs and results of hydro-dynamic simulation models. In the context of this study it is important that the model represents the catchment non-linearities well. Comprehensive tests have shown that this is indeed the case (Blöschl et al., 2006). Of particular value have been the extreme flood events in August 2002 as they allowed to test the

Table 2. Characteristics of the events for which flood forecasts are analysed in this paper. Kamp at Zwettl, 622 km^2.

	Aug 2002a	Aug 2002b	July 2005	Aug 2005a	Aug 2005b
Precipitation (mm)	212	114	88	70	50
Direct runoff depth (mm)	82	56	27	13	13
Runoff coefficient (–)	0.39	0.49	0.30	0.18	0.26
Initial moisture state	Dry	Very wet	Wet	Dry	Wet
Peak discharge (m^3/s)	459	367	95	68	65
Return period (yrs)	~1000	~500	5	3	3
Beginning of event	6 Aug 00:00 h	11 Aug 00:00 h	5 July 00:00 h	14 Aug 00:00 h	20 Aug 00:00 h
End of event	10 Aug 21:00 h	15 Aug 21:00 h	15 July 00:00 h	19 Aug 21:00 h	26 Aug 21:00 h
Beginning of rising limb	6 Aug 12:00 h	11 Aug 12:00 h	10 July 12:00 h	16 Aug 00:00 h	21 Aug 12:00 h
End of rising limb	8 Aug 06:00 h	13 Aug 18:00 h	11 July 06:00 h	17 Aug 21:00 h	22 Aug 12:00 h
Time to peak (h)	36	48	30	18	24

model over a wide range of event magnitudes, from small to extreme, along with the smaller events on record (Fig. 2).

To increase forecast accuracy, two real-time updating procedures are used in the Kamp flood forecasting system. The first procedure assimilates runoff data to update the catchment soil moisture state based on Ensemble Kalman filtering (EnKF) (Evensen, 1994). The strength of the EnKF is that it can accommodate model non-linearity. The model variance represents the errors in the precipitation and evaporation inputs that control the soil moisture state of the catchments and was set to a constant value of 0.005 (mm/15 min)^2 based on sensitivity analyses. The model update is performed for every timestep and the updating is uniform within each gauged catchment. The observation variance represents the discharge measurement errors and is assumed to increase with runoff. The observation variance of runoff was set to $\xi \cdot Q_i^2$ where ξ=0.0025 was obtained from sensitivity analyses and Q_i denotes the observed runoff at timestep i. The soil moisture state of the catchment estimated by the EnKF is used as the initial condition of all forecast runs. The second procedure consists of an additive error model (termed MOS or model output statistics) that updates runoff directly. This error model exploits the autocorrelation of the forecast error and involves an exponential decay of the correction. The autocorrelation lag was found from error analyses of events as 4 h.

2.3 Generating ensemble forecasts

At each time step, precipitation observed at the telemetered rain gauges (Fig. 1) over the past 15 min is interpolated on the 1 km grid using climatologically scaled radar information (Haiden et al., 2007). The climatological scaling is derived from a comparison of monthly totals of the radar and raingauge data at the station locations and varies with location and season. The scaled radar field is linearly combined with the field derived by station interpolation, the weights of this combination depending on the climatological scaling factor.

In regions where this factor is large (i.e., the visibility by the radar network is low), most of the weight is with the station interpolation. Where the factor is close to unity, the scaled radar field dominates the final estimate. The final precipitation analysis reproduces the observed values at the raingauge locations.

Additionally, at each time step, deterministic precipitation forecasts are made at 15 min temporal resolution over a lead time of 48 h. The forecasts consist of two components. The first component is an observation-based extrapolation or nowcast of the interpolated precipitation field using motion vectors. They are determined from consecutive analyses by searching for the spatial shift which gives the best match (lowest root mean square difference) of precipitation patterns (Haiden and Steinheimer, 2007). The second component is a weighted mean of the forecast fields of the ALADIN (Wang et al., 2006) and ECMWF numerical weather prediction (NWP) models. The weighting function to estimate the optimised precipitation forecast P_{opt} can be written as

$$P_{\text{opt}} = w_{\text{ALA}} \cdot P_{\text{ALA}} + w_{\text{ECM}} \cdot P_{\text{ECM}} \tag{1}$$

were w_{ALA} and w_{ECM} are the weights for the ALADIN and the ECMWF precipitation forecasts P_{ALA} and P_{ECM}, respectively. The weights w_{ALA} and w_{ECM} have been derived from several years of comparisons of ALADIN and ECMWF forecasts with observed precipitation by minimising

$$e^* = e_M + 0.5 \cdot e_B \tag{2}$$

where e_M is the mean absolute error and e_B is the absolute value of the bias. This optimisation has been performed separately for moderate (>5 mm/24 h) and heavy (>10 mm/24 h) precipitation events. The sum of the optimized weights w_{ALA} and w_{ECM} can differ from unity to account for biases. The mean error of the combined precipitation forecasts in that period was 20–25% smaller than that of the individual ALADIN and ECMWF forecasts (Haiden et al., 2007) and the biases were negligible.

Another weighting function is used for a smooth transition between the two components (nowcast and NWP forecast) (Golding, 1998). Analyses of the forecast performance indicated that, in most cases, over the first 2–6 h of the forecast the nowcast had smaller errors than the NWP forecast combination. A weighting function was hence chosen that gives full weight to the nowcast during the first 2 h, decreases linearly to zero at 6 h, and remains at zero for larger lead times.

It should be noted that beyond the nowcasting range, the 15 min temporal resolution of the precipitation forecast does not reflect the actual information content of the meteorological models. ALADIN provides output for every hour, and ECMWF provides 6-hourly totals, both of which are linearly interpolated to a uniform 15 min resolution. Similarly, the spatial grid scale of ALADIN (9.6 km) and ECMWF ($\sim$25 km) is much larger than the 1 km grid of the hydrological model. Although the scales of the meteorological models and the hydrological model do not match, sensitivity analyses indicated that the first order effect of precipitation uncertainty on runoff is related to (average) catchment precipitation, while the uncertainty resulting from a lack of knowledge of the spatial detail of precipitation is a second order effect.

In order to quantify the uncertainty of the precipitation forecasts, ensemble forecasts are generated. They are constructed, in a similar way as the deterministic forecasts. The ECMWF model provides, at each run, a set of 50 ensemble forecasts in addition to the main (deterministic) run. The ALADIN model currently does not produce ensemble forecasts operationally, so a set of 25 pseudo-ensembles is generated by spatially shifting the ALADIN forecast in both the x and y directions by a random space lag of up to 40 km. This spatial lag has been introduced to account for some of the small-scale uncertainty in the position of the precipitation forecasts. Each of the ECMWF members is then randomly combined with one of the ALADIN pseudo-ensemble members, and with the nowcast. No uncertainty has been assigned to the nowcasts. This means that, up to 2 h lead time, all ensemble members are identical (zero spread) and the spread increases at longer lead times. We use the pseudo-ensembles of the ALADIN model in the construction of the ensembles because they provide small-scale variance and spread not present in the ECMWF forecasts.

In the case of air temperature, station data are interpolated. The forecasts are based on a combination of the station data with the ALADIN forecasts. No temperature ensembles are computed as their effect on the flood forecasting uncertainty is deemed to be small.

The interpolated precipitation and air temperature fields are used to estimate the state variables of the runoff model such as soil moisture, reservoir storages and snow water equivalent at each time step allowing for EnKF updating. These state variables are used as the initial conditions for the flood forecasts. All members of the ensemble forecasts use the same initial conditions as the deterministic forecast without any perturbation.

The deterministic forecast fields (both precipitation and temperature) are used as an input to the runoff model to compute deterministic flood forecasts. The 50 members of the ensemble forecasts of precipitation along with the deterministic temperature forecasts are used to compute the ensemble flood forecasts, i.e., 50 realisations of runoff over a lead time of 48 h. These are analysed in this paper. Each member of the ensemble forecasts is updated by the additive error model in the same way which means that the error model does not affect the ensemble spread.

3 Results and discussion

3.1 Model performance and deterministic forecasts

To get an appreciation of the performance of the components of the flood forecasting system, the forecast errors e_j were examined for the five flood events of Table 2:

$$e_j = \frac{1}{i_2 - i_1} \sum_{i=i_1}^{i_2} \frac{\left| \hat{Q}_{ij} - Q_i \right|}{Q_i} \qquad (3)$$

where e_j is the mean absolute normalized error in percent for lead time j, $\hat{Q}_{ij}$ is runoff at time step i that is forecast with a lead time of j, Q_i is the observed runoff at time step i, and i_1 and i_2 are the beginning and the end of the analysis interval, respectively. The error analyses were performed separately for the entire flood events (i.e. between the beginning and the end of the flood event as in Table 2) and the rising limbs only (as in Table 2). The rising limb of a flood hydrograph is the period that is of most interest for the users of real-time flood prediction system. For a given lead time j, the errors of the five events were averaged and are shown in Fig. 3. Four cases were considered.

In the first case (blue lines in Fig. 3) we assumed that future precipitation and temperature were known and used their observed interpolated fields as inputs to the runoff model without any updating. Figure 3 indicates that, for this case, the errors do not depend on the forecast lead time. This would be expected as this is the simulation mode. The model errors for the entire flood events are about 15% (Fig. 3, left) while they are about 30% if the rising limb alone is analysed (Fig. 3, right). The better model accuracy for the entire flood events results from including the time periods with no rainfall, i.e. the recession. In these periods the errors are small as no uncertainties about the amount and the spatial distribution of the input rainfall fields are propagated through the rainfall-runoff model. In contrast, the rising limbs are more difficult to simulate.

In the second case we ran the model in a similar way as in case 1, however, allowed for the EnKF updating to estimate the initial conditions of the forecasts (termed updated initial
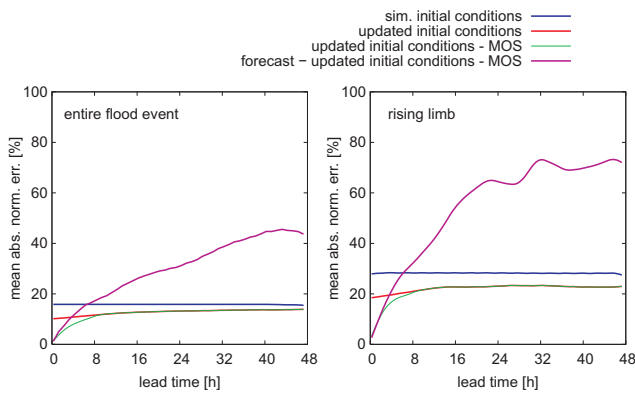
Fig. 3. Average forecast errors e_j of the five flood events of Table 2. Entire flood event (left), rising limb only (right). Kamp at Zwettl, $622\,km^2$.

conditions, red lines in Fig. 3). Updating the initial conditions reduces the errors for both analysis periods. During the rising limbs the updating reduces the errors from about 30% to about 20%. For the entire events the updating reduces the errors from about 15% to about 12%. There is a slight dependence of the error on the lead time with smaller errors for short lead times. This dependence is related to the memory of the hydrological system which is taken advantage of by the updating.

The third case was as case 2 but, in addition, allowed for updating by the additive error model (termed updated initial conditions – MOS, green lines in Fig. 3). The benefit of the additive error model is limited to the first eight hours of the forecasts which is the interval over which the errors are correlated. For larger lead times the additive error model has no effect on the forecasts, so the errors are identical with those of case 2. The fourth case was as case 3 but used forecast precipitation and temperatures rather than the observations (termed forecast – updated initial conditions MOS, purple lines in Fig. 3). The fourth case represents the operational real time configuration, where both updating procedures are used along with the deterministic precipitation and temperature forecasts. In this case, the forecast performance shows a clear dependence on the forecast lead time. In the first 8 h the forecast, errors are less than 30% (rising limb alone) and about 18% (entire flood events). For lead times of 48 h the errors are 75% (rising limb alone) and about 50% (entire flood events). This is much larger than the errors of case 3 where observed precipitation has been used as an input. This clearly demonstrates that the main error source for lead times larger than the travel times is the uncertainty in the precipitation forecasts. The difference between the errors of the two analysis periods is particularly large in case 4, as would be expected, as the precipitation forecasts will be most significant in the rising limb where the rainfall occurs.

3.2 Ensemble forecasting and propagation of non linearity

Depending on the soil moisture state, a change in precipitation input can be amplified (wet conditions) or dampened (dry conditions) by the catchment system. In the Kamp catchment the occurrence of big floods is associated with wet catchment conditions or very large rainfall depths that wet up the catchment during the event. In the case of the flood events examined here, one would hence expect that the precipitation forecast errors will be amplified as they are propagated through the hydrological model. To illustrate the propagation characteristics ensemble forecasts for the flood event of July 2005 are shown in Figs. 4, 5 and 6. The plotted time window ranges from 9 July 2005 00:00 h to 12 July 2005 00:00 h in all three figures. The thick red lines represent observed runoff, the black lines represent the deterministic forecasts and the thin blue lines represent the 50 ensemble members. The light blue shading represents the confidence interval between the 10% and 90% quantiles of the ensemble forecasts.

Figure 4 shows the forecasts on 9 July 2005 00:00 h. In the first 24 h of the forecast lead time only 5 mm of precipitation have been observed. The deterministic precipitation forecast is very accurate but most of the ensemble members predict more than that. In the second 24 h of the forecast lead time about 50 mm of precipitation have been observed. Again, the deterministic forecast is accurate. However, most of the ensemble forecasts underestimate precipitation slightly. The deterministic flood forecast matches the observed hydrograph closely. During the first 12 h of the forecast lead time, the ensemble members are very similar to each other. This is because the forecasts are controlled by observed upstream runoff and observed precipitation through the routing and runoff model components, respectively. In both model components no uncertainty was introduced, i.e. the same data and parameters were used for all members of the ensemble. For lead times of 30 h and more, some of the ensemble members indicate a sudden increase in discharge with a maximum flood peak of $100\,m^3/s$. As compared to precipitation, the ensemble spread for these lead times is much larger. This is where the uncertainty in forecast precipitation becomes important.

The results of the forecast run on 10 July 2005 00:00 h are plotted in Fig. 5. The total observed precipitation during the forecast lead time is about 75 mm while the deterministic forecast predicts about 60 mm. This relatively moderate underestimation of 15% translates into a larger underestimation of runoff with an estimated peak of $57\,m^3/s$ as compared to an observed peak of $95\,m^3/s$, i.e., an underestimation of 40%, and the rising limb is almost completely missed. The main reason is the missing precipitation block at time 36 h. Most ensemble members underestimate precipitation and the ensemble spread is very small up to a lead time of 24 h. Similarly, the spread of the runoff forecasts is small during the first 24 h but in the last 24 h of the forecasts the spread increases significantly. Clearly, this increase is related
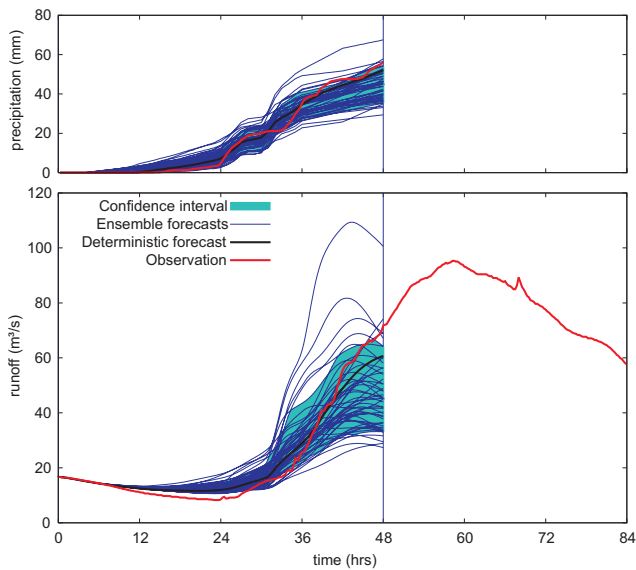
Fig. 4. Ensemble forecasts (top: cumulative catchment precipitation, bottom: runoff) on 9 July 2005 at 00:00 h (time 0 in the figure). Kamp at Zwettl, 622 km^2.
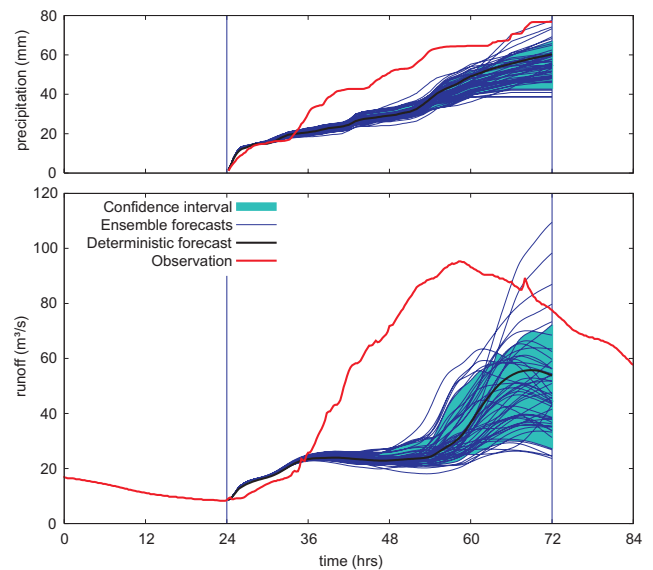

Fig. 5. Ensemble forecasts (top: cumulative catchment precipitation, bottom: runoff) on 10 July 2005 at 00:00 h (time 24 in the figure). Kamp at Zwettl, 622 km^2.

to the non-linearity of the catchment system that translates the somewhat larger spread of precipitation at the end of the forecast period in a much larger spread in runoff. While the observed hydrograph is never within the confidence interval, at the end of the forecast some of the ensemble members do indicate the possibility of a flood on the order of 100 m^3/s. Twelve hours later, the forecasts are much more accurate (Fig. 6). The deterministic precipitation forecast estimates the observed precipitation very well over the entire lead time. The runoff forecasts are very good for the first 12 h but do underestimate runoff for larger lead times. In this case, the underestimation is not a result of precipitation errors but is related to the initial conditions of the hydrologic model that are somewhat too dry.

To illustrate how the spread of the precipitation ensemble, representing the uncertainties in the precipitation forecasts, is propagated through the hydrologic model we analysed the probability distributions of the precipitation ensembles (model input) and the runoff ensemble (model output). In both cases it was assumed that all ensemble members are equally likely. As an example, Fig. 7 shows the probability distributions of the forecast on 10 July 2005 at 00:00 h as in Fig. 5. For a lead time of 24 h the distributions of the precipitation and runoff ensemble are similar. They exhibit a narrow spread and are symmetric. For a lead time of 36 h the two distributions are somewhat different. While the precipitation ensemble spread remains small, the runoff ensemble spread is larger and skewed to the right. This effect is even stronger for a lead time of 48 h and the largest 20% of the runoff ensemble members have increased their spread dramatically. The total range of the precipitation uncertainty
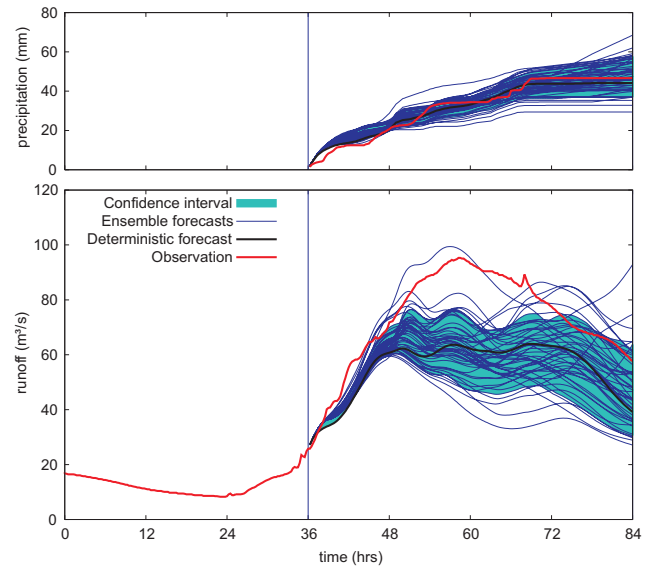

Fig. 6. Ensemble forecasts (top: cumulative catchment precipitation, bottom: runoff) on 10 July 2005 at 12:00 h (time 36 in the figure). Kamp at Zwettl, 622 km^2.

at 48 h is 40 mm or 70% of the median precipitation while the total range of the runoff uncertainty at 48 h is 90 m^3/s or 200% of the median runoff. Small errors in rainfall may translate into larger errors in runoff. The example of Fig. 7 has been extended to all the 232 forecasts examined in this paper. For each of these forecasts, the coefficient of variation of the ensemble members of precipitation and runoff has been calculated. The results for a lead time of 48 h are
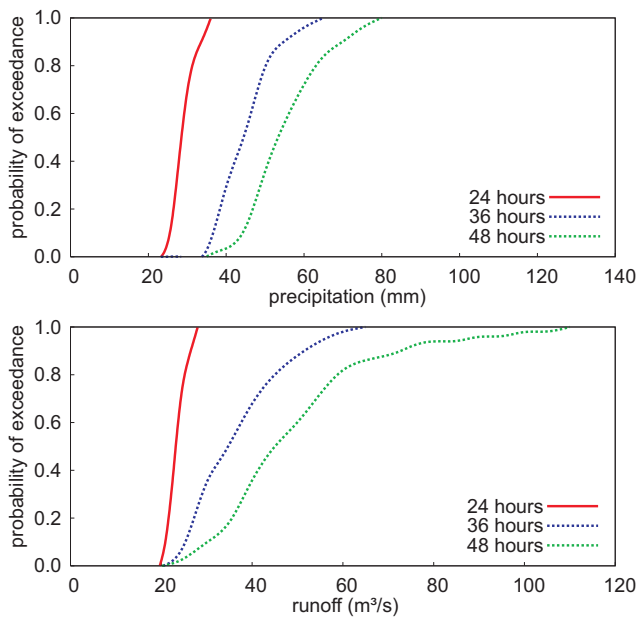
Fig. 7. Mapping of the precipitation uncertainties to runoff uncertainties for forecast lead times of 24, 36 and 48 h. cdf is the cumulative distribution functions assuming all ensemble members are equally likely. Forecast on 10 July 2005 at 00:00 h as in Fig. 5. Kamp at Zwettl, 622 km^2.

Fig. 8. Mapping of the precipitation uncertainties to runoff uncertainties for forecast lead times of 48 h in terms of the coefficient of variation of the ensemble spread of cumulative precipitation CV_P and runoff CV_Q. Five flood events of Table 2, Kamp at Zwettl, 622 km^2. Arrow indicates 48 h forecasts of Fig. 7. Largest dots relate to forecasts with mean cumulative precipitation $\bar{P} > 50$ mm, and the medium and smallest dots relate to forecasts with $50 > \bar{P} > 30$ mm and $30 > \bar{P} > 10$ mm, respectively.

shown in Fig. 8, stratified by the mean cumulative precipitation. The largest dots in Fig. 8 relate to forecasts with mean cumulative precipitation $\bar{P} > 50$ mm, and the medium and smallest dots relate to forecasts with $50 > \bar{P} > 30$ mm and $30 > \bar{P} > 10$ mm, respectively. For small precipitation depths, the uncertainty in precipitation may or may not matter for runoff. Indeed, if precipitation is very small, runoff will be controlled by groundwater response, so any uncertainty in precipitation will not appear in the runoff forecasts. In contrast, for the instances when the forecasted precipitation was large (largest dots in Fig. 8), the coefficients of variations may more than double when moving from precipitation to runoff. In the example of Fig. 7, the coefficient of variation increases from 0.16 to 0.34 when moving from precipitation to runoff (arrow in Fig. 8).

Clearly, this kind of mapping of precipitation uncertainties to runoff uncertainties for large forecast lead times is related to the non-linear nature of catchment response. Non-linearity in runoff response has been observed at all space time scales. Often, the non-linearity is more pronounced in dry climates than in wet ones (Chiew et al., 2006). With 300 mm of annual runoff the Kamp is a rather dry catchment in an Austrian context.

3.3 Ensemble spread and forecast error

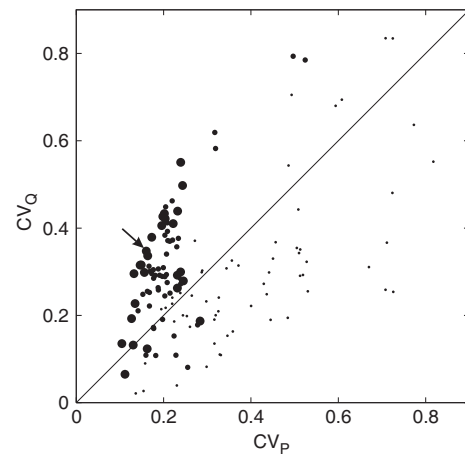Ideally, the ensemble spread should be an estimator of the distribution of the forecast errors. However, in the present study not all sources of uncertainty have been represented in the ensembles. Rain gauge measurement errors, small scale precipitation variability between the raingauges, uncertainty in the routing and runoff models as well as uncertainties in initial soil moisture have not been represented in the ensembles. Also, it is unclear whether the ECMWF ensemble forecasts and the ALADIN pseudo-ensembles are equally probable forecasts in the study region. The obvious method of examining to what degree the ensemble spread actually captures the distribution of the forecast errors is a comparison of the two, based on an analysis of observed flood events. However, the forecast error distribution changes with time. Typically, the forecast errors are large during the rising limbs of floods and small during the recession and low flow periods. Most importantly, one is interested in the forecast errors of large flood events but large events are always rare, so statistical analyses are limited by small sample sizes. As a simplification we assumed here that the forecast errors of all 232 time steps of the five flood events of Table 2 can be combined into a single distribution function for each forecast lead time. It should be noted, however, that not all of the 232 forecasts are completely independent from each other. We calculated the forecast errors as the difference between the deterministic forecast and the observed runoff (positive error for overestimation) from which we derived the distribution function. In a similar vein, we calculated the deviations between the ensemble forecasts and the deterministic forecast (positive deviation if ensemble forecast is larger than the deterministic forecast), and calculated the distribution function for the same time steps as in the case of the errors, assuming that all ensemble members are equally probable.
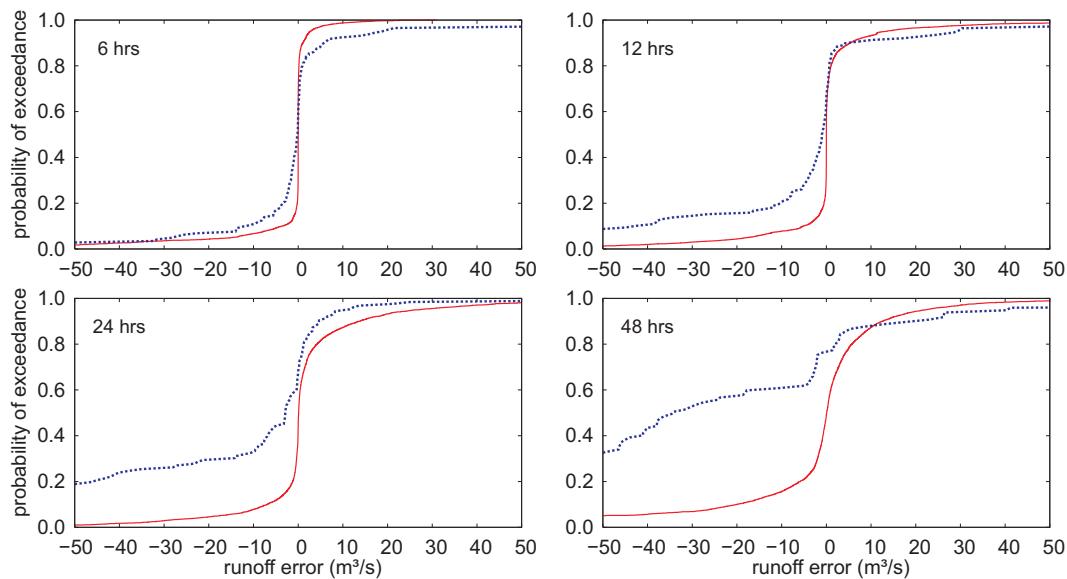
Fig. 9. Comparison of the distribution functions of the average ensemble spread around the deterministic forecast (solid red lines) and forecast errors of the deterministic forecasts (dotted blue lines) for lead times of 6, 12, 24 and 48 h for the five flood events of Table 2. Kamp at Zwettl, 622 km^2.

Figure 9 shows the results of this comparison. The solid red lines represent the average ensemble spread around the deterministic forecasts and the dotted blue lines show the error distributions of the deterministic forecasts. For a forecast lead time of 6 h, about 90% of the ensemble members hardly differ from the deterministic discharge forecast (upper left panel). The remaining members show a slight tendency of being smaller than the deterministic forecast. The forecast errors for this lead time exhibit a slightly wider distribution than that of the ensemble members. For a forecast lead time of 12 h, the spread of the ensemble members increases as would be expected but the forecast errors increase even more. In particular, there are a number of time steps where runoff was significantly underestimated (i.e. negative errors). This tendency continues as one moves to 24 and 48 h lead times. For 48 h, the ensemble spread is larger than that of the other lead times and so are the forecast errors. In about half the time steps, the deterministic forecasts underestimate runoff by more than 30 m^3/s while less than 10% of the ensemble forecasts indicate deviations of less than −30 m^3/s.

The median forecast errors of the 6, 12 and 24 h lead times are close to zero but the large negative errors are more frequent than the large positive errors, i.e., there exists a negative skew. This is even more the case for a lead time of 48 h. This means that the deterministic forecasts underestimate runoff more often than they overestimate runoff in the case of the five flood events. This effect can be potentially related to the tendency of the deterministic precipitation forecast to underestimate extremely high amounts of precipitation during the five flood events. Although a bias correction is used in preparing the forecasts (see Sect. 2.3) it is based on

the analysis of moderate (>5 mm/24 h, >10 mm/24 h) precipitation events. What is of most interest in a flood forecasting context are the very large precipitation events but such extreme events are rare, so sample size is very small. It is likely that the forecast errors and biases of the extreme events will differ from those of the moderate events as one would assume that the error characteristics are heteroscedastic. However, accounting for such biases in practice is very difficult.

The ensemble forecasts are almost symmetric although the 48 h lead times do indicate a slight negative skew. Also, the ensemble spread is always narrower than the distribution of the forecast errors. This would be expected as not all error sources have been represented in the ensembles. However, the ensemble spread increases with lead time in a similar way as the forecast errors. This means that the ensemble spread does provide an indicator to assess potential forecast errors over a range of lead times. Also, one would expect that the most significant changes in the forecast errors as a function of time are captured in the ensembles as they are related to precipitation.

As another possibility of assessing the ability of the ensemble flood forecasts to capture the forecast errors we analysed what we term "range hit rates". A range hit is counted when the observed discharge value is within the range of a certain number of discharge ensemble members. How many of the 50 ensemble members are used to define the upper and lower range is described by the quantile. For the entire forecast ensemble the quantile is 100%. A quantile of 60% means that the highest 20% and the lowest 20% of the ensemble forecast values are not taken into account, i.e., a
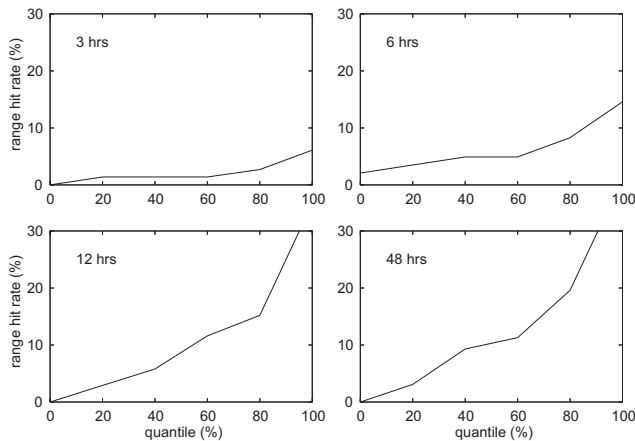
Fig. 10. Range hit rates for lead times of 6, 12, 24 and 48 h for the five flood events of Table 2. Kamp at Zwettl, 622 km². The range hit rate indicates in how many cases, relative to the total number of forecasts, the observed discharge value lies within the range of the ensemble quantiles.

range hit is counted if the observed runoff is within the range covered by the remaining 60% of the ensemble. A quantile of 0% relates here to the deterministic forecast alone, i.e., a range hit is counted if the observed runoff is identical with the deterministic forecast within the numerical accuracy of 2 digits used here. With this definition, the range hit rate indicates in how many cases, relative to the total number of forecasts, the observed discharge value lies within the range of the ensemble quantiles. The range hit rates were calculated for the same forecasts as used for Fig. 9.

Figure 10 shows the results of this analysis for different forecast lead times. For all lead times, the range hit rate increases with the quantiles. Clearly, the wider the uncertainty range the easier it is to capture the observed runoff. For quantiles larger than 60% the range hit rate increases more strongly which is related to the influence of the hydrologic non-linearity combined with the growing deviation from the ensemble mean for the peripheral ensemble members. For a lead time of 3 h, the range hit rates are always very small. This is because most of the forecast error is due to the routing model and the discharge measurements and both error sources have not been included in the ensemble forecasts. However, as the lead time increases, the precipitation forecast error becomes more important and the range hit rates increase. The range hit rates are still much smaller than the quantiles. In fact, if the ensemble forecasts captured all the forecast errors one would expect the range hit rates to lie on the 1:1 line. Clearly, this is not the case as the ensemble forecasts focus on the dominant source of uncertainty, i.e., uncertainty in forecast precipitation. The range hit rates are similar for the 12 h and the 48 h lead times (as well as for 24 and 36 h not shown here) which suggests, again, that the en-

semble spread does provide an indicator to assess potential forecast errors over a range of lead times, provided the lead times are 12 h or larger.

3.4 Relative operating characteristics

Flood management decisions are often based on discharge thresholds, i.e., if a threshold will be exceeded some kind of alarm is triggered. Depending on the context, the warning may result in an alert as is usually the case in early warning, or it may result in flood mitigation action for shorter lead times. When ensemble forecasts are available, any of the ensemble quantiles could be used to trigger an alarm. There is a tradeoff between the ensemble quantile that is used and the usefulness of the alarm. Ensemble members at the upper end (100% quantile) will more likely trigger an alarm but there will also be more false alarms. The opposite is true of the ensemble members at the lower end (0% quantiles). In meteorology, the method of ROC (Relative Operating Characteristic) – diagrams based on threshold analysis are a common method for assessing this tradeoff and hence the performance of probabilistic forecasting systems (Mason and Graham, 1999; Buizza et al., 1999). In analogy, ROC-diagrams are used here to illustrate the alert characteristic of the ensemble flood forecasts for predefined discharge thresholds.

As a first step, hit rates and false alarm rates are defined. The hit rate *HR* is the ratio of the number of correct alarms *H* and the total number of observed events defined by the sum of correct alarms *H* and missed alarms *M*, i.e.

$$HR = \frac{H}{H+M} \tag{4}$$

where a correct alarm is counted if both the observed and forecast hydrograph exceed the threshold within the forecast lead time, and an observed event is counted if the observed hydrograph exceeds the threshold within the forecast lead time. In analogy, the false alarm rate *FAR* is the ratio of the number of false alarms *F* and the total number of no-events defined by the sum of false alarms *F* and correct rejections *C*(neither the observed hydrograph nor the forecast exceeds the threshold), i.e.

$$FAR = \frac{F}{F+C} \tag{5}$$

where a false alarm is counted if the forecast hydrograph exceeds the threshold within the forecast lead time but the observed hydrograph does not, and a no-event is counted if the observed hydrograph does not exceed the threshold within the forecast lead time. $H+M+F+C$ is 232 in this paper as this is the total number of time steps for which forecasts have been analysed. The hit rate and the false alarm rate of the deterministic forecasts can be plotted as a single point on a hit/false alarm rate graph. The same procedure is then repeated for each quantile of the ensemble forecasts separately

which gives a set of points in the hit/false alarm rate graph known as relative operating characteristic (ROC). A perfect forecasting system gives a hit rate of 100% and a false alarm rate of 0%, i.e. the point plots in the top left corner of the ROC diagram. Systems with no skill result in a ROC curve on the 1:1 line.

The relative operation characteristics were calculated for the same forecasts as used for Fig. 9 with a forecast lead time of 48 h and are shown in Fig. 11. Four thresholds were selected. The discharge thresholds of 50 and $100 \, \text{m}^3/\text{s}$ are relevant values for flood warning at the Kamp (see Table 2), the smaller thresholds of 10 and $30 \, \text{m}^3/\text{s}$ were examined for illustrative purposes. The dots represent the ensemble forecasts at intervals of 5%, the crosses represent the deterministic forecasts. A hit rate of 100% is reached for a threshold of $10 \, \text{m}^3/\text{s}$ and the 100% quantile of the ensemble forecast (i.e. the largest of the ensemble members). This perfect hit rate is associated with a false alarm rate of more than 80%. With decreasing ensemble quantiles the false alarm rate decreases to less than 20% while the hit rate is greater than 50% for all ensemble quantiles. For a threshold of $30 \, \text{m}^3/\text{s}$, the hit rate ranges from about 80% (100% quantile) to 15%, while the false alarm rate ranges from 20% to nil. The false alarm rates are even smaller for the 50 and $100 \, \text{m}^3/\text{s}$ thresholds with similar hit rates. For the $100 \, \text{m}^3/\text{s}$ threshold the forecasts of the 100% quantile produce only 10% false alarms with a hit rate of about 80%.

For all discharge thresholds, the hit rates of the 100% quantiles of the ensemble forecasts are larger than those of the deterministic forecasts as would be expected. This is the main reason of using ensemble forecasts in flood management. In general, the ROC curves show that the deterministic and probabilistic forecasts at the Kamp produce only few false alarms and the percentage of false alarms decreases with the magnitude of the discharge threshold. The hit rate is limited to about 80% for the 50 and $100 \, \text{m}^3/\text{s}$ thresholds. The tendency of very small false alarm rates and maximum hit rates below 100% in the ROC curves indicate that the flood forecasts tend to underestimate the observed discharges, particularly for the 50 and $100 \, \text{m}^3/\text{s}$ thresholds. This tendency is consistent with results of meteorological analyses, which have shown that meteorological model forecasts tend to overestimate small precipitation values and underestimate large precipitation values (Buizza, 1999). The analysis in this paper is based on five big flood events with heavy precipitation, so some underestimation of precipitation would be expected. The tendency towards underestimating precipitation during the floods is amplified through the non-linearity of the hydrologic response at the Kamp. Therefore, the flood quantile of choice for flood alarm purposes would be a high ensemble quantile, for example the 90% quantile. There is another argument for using a large quantile which are the relative costs of false alarms and missed alarms. If false alarms are inexpensive it may pay to choose higher flood quantiles as if false alarms were as expensive as missed alarms. However,
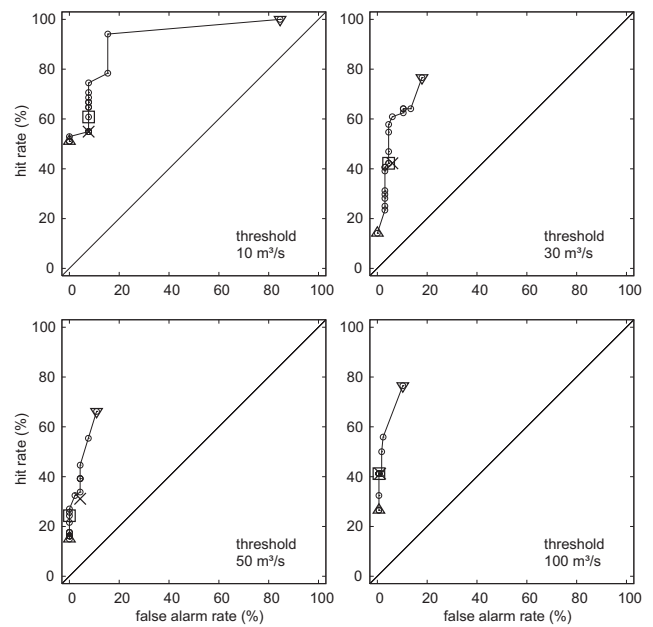

Fig. 11. Relative Operating Characteristic (ROC) curves for the probabilistic (dots) and deterministic (crosses) 48 h flood forecasts for discharge thresholds of 10, 30, 50 and $100 \, \text{m}^3/\text{s}$ for the five flood events of Table 2. Kamp at Zwettl, $622 \, \text{km}^2$. 5% quantile (upward pointing triangle), 50% quantile (square) and 100% quantile (downward pointing triangle).

decisions on alarms are often made based on maximising the credibility of the forecasts rather than cost arguments.

The area under the ROC curve is sometimes used as a measure for the forecast skill (Stanski et al., 1989). The area under the curve decreases from 1 for a perfect prediction system to 0.5 for a prediction system with no skill. Fitting a cubic spline to the ROC curves in Fig. 11 gives areas of 0.90, 0.85, 0.85 and 0.90 for the 10, 30, 50 and $100 \, \text{m}^3/\text{s}$ thresholds, respectively. As compared to precipitation forecasts in the literature, this is a favourable skill. For example, Buizza et al. (1999) found skills in the range of 0.70 and 0.83, depending on precipitation thresholds and for a maximum forecast lead time of 3 days.

4 Conclusions

The real-time flood forecasting system of the Kamp catchment in Austria has been operational since January 2006. It is used in this paper to examine how the ensemble distribution of precipitation forecasts propagates in the catchment system, and to interpret the flood forecast probabilities relative to the forecast errors. The model was tested on five large flood events including a 500 yr flood and a 1000 yr flood.

The analyses indicated that, for long lead times (e.g. 48 h), the variability of the precipitation ensemble is amplified as it propagates through the catchment system. For the example

examined, the total range of the precipitation uncertainty is 70% of the median precipitation while the total range of the runoff uncertainty is 200% of the median runoff. Small errors in rainfall may translate into larger errors in runoff. An analysis of the coefficients of variation of the ensemble members of precipitation and runoff suggests that, for small precipitation depths, the uncertainty in precipitation may or may not matter for runoff. In contrast, for the instances when future precipitation is large, the coefficients of variations may more than double when moving from precipitation to runoff. Also, the ensemble distribution of precipitation is symmetric while that of the flood forecasts is skewed to the right. Clearly, this kind of mapping of precipitation uncertainties to runoff uncertainties for large forecast lead times is related to the non-linear nature of catchment response. In contrast, for short lead times (e.g. 12 h and less), the variability of the precipitation ensemble is decreased as it propagates through the catchment system. This is because the forecasts are mainly controlled by observed upstream runoff and observed precipitation through the routing and runoff model components, as the forecasting system is operated in a real-time mode. The ensemble forecasts focus on the dominant source of uncertainty, i.e., uncertainty in forecast precipitation. For lead times of 12 h and less the ensemble spread is very narrow as other sources uncertainty such as rain gauge measurement errors, small scale precipitation variability between the raingauges, uncertainty in the routing and runoff models as well as uncertainties in the initial soil moisture have not been represented in the ensembles. More generally speaking, it can be expected that the lead time where the uncertainty of the precipitation forecasts starts to amplify will depend on the catchment response characteristics, such as travel times in the river reaches and runoff concentration. In small and flashy catchments this will be a short lead time while for large catchments it will be longer.

The paper also examined the ability of the probabilistic forecasts to capture the distribution of the flood forecast errors. Assuming that all ensemble members are equally likely, the statistical analyses of the ensemble forecasts for five flood events at the Kamp showed that the ensemble spread is always narrower than the distribution of the forecast errors. This would be expected as not all error sources have been represented in the ensembles. Although two updating procedures based on observed runoff have been used to improve the flood forecasts over the simulation mode, there always remains a certain amount of hydrologic uncertainty in the forecasting system. It is also likely, that the precipitation ensembles do not fully represent the precipitation forecast errors Schaake et al. (2004). However, the ensemble spread increases with lead time in a similar way as the forecast errors. This means that the ensemble spread does provide an indicator to assess potential forecast errors over a range of lead times. Also, one would expect that the most significant changes in the forecast errors as a function of time are captured in the ensembles as they are related to precipitation. A

"range hit rate" was defined as the number of cases, relative to the total number of forecasts, in which the observed discharge value lies within the range of the ensemble quantiles. Analyses of the range hit rates indicate that they are small for short lead times but increase with lead time. The range hit rates are similar for lead time of 12 h and more which suggests, again, that the ensemble spread does provide an indicator to assess potential forecast errors over a range of lead times, provided the lead times are 12 h or larger. Finally, the forecast skill of the 48 h ensemble forecasts was tested by ROC (Relative Operating Characteristic) diagrams based on threshold analyses. For all discharge thresholds, the hit rates of the 100% quantiles of the ensemble forecasts are larger than those of the deterministic forecasts. This is the main reason of using ensemble forecasts in flood management. For the largest discharge threshold examined here (100 m^3/s) the forecasts of the 100% quantile produce only 10% false alarms with a hit rate of about 80%. The flood quantile of choice for flood alarm purposes would be a high ensemble quantile, for example the 90% quantile.

Even though the ensemble characteristics do not exactly match the forecast errors, they do provide information about the expected forecast errors. The comparisons indicated that, for lead times larger than 12 h in the case of the 622 km^2 Kamp catchment, the ensemble spread is a useful indicator to the forecast errors. While additional error sources could be included in estimating the flood ensembles it may not be necessary for operational flood forecasting purposes as the uncertainty in forecast precipitation is the dominant source of flood forecast uncertainty for lead times of more than 12 h in catchments such as the Kamp.

Appendix A

Structure of the soil moisture model

A conceptual soil moisture accounting scheme is used at the model grid scale. The sum of rain and melt, $P_r + M$, is split into a component dS that increases soil moisture of a top layer, S_s, and a component Q_p that contributes to runoff. The components are split as a function of S_s:

$$Q_p = \left(\frac{S_s}{L_s} \right)^\beta \cdot (P_r + M) \tag{A1}$$

L_s is the maximum soil moisture storage. β controls the characteristics of runoff generation and is termed the non-linearity parameter. If the top soil layer is saturated, i.e., $S_s = L_s$, all rainfall and snowmelt contributes to runoff and dS is 0. If the top soil layer is not saturated, i.e., $S_s < L_s$, rainfall and snowmelt contribute to runoff as well as to increasing S_s through $dS > 0$:

$$dS = P_r + M - Q_p - Q_{by}$$
$$\text{if } P_r + M - Q_p - Q_{by} > 0 \tag{A2}$$
$$dS = 0 \text{ otherwise}$$

where, additionally, bypass flow Q_{by} is accounted for. Analysis of the runoff data at the Kamp indicated that flow that bypasses the soil matrix and directly contributes to the storage of the lower soil zone is important for intermediate soil moisture states S_s. For $\xi_1 \cdot L_s < S_s < \xi_2 \cdot L_s$ (with $\xi_1 = 0.4$, $\xi_2 = 0.9$) bypass flow was assumed to occur as

$$Q_{by} = \alpha_{by} \cdot (P_r + M) \quad \text{if} \quad \alpha_{by} \cdot (P_r + M) < L_{by}$$
$$Q_{by} = L_{by} \text{ otherwise} \tag{A3}$$

while no by pass flow was assumed to occur for dry and very wet soils. Changes in the soil moisture of the top soil layer S_s from time step $i-1$ to i are accounted for by

$$S_{s,i} = S_{s,i-1} + (dS - E_A) \cdot \Delta t \tag{A4}$$

The only process that decreases S_s is evaporation E_A which is calculated from potential evaporation, E_P, by a piecewise linear function of the soil moisture of the top layer:

$$E_A = E_P \cdot \frac{S_s}{L_P} \quad \text{if} \quad S_s < L_p$$
$$E_A = E_P \quad \quad \text{otherwise} \tag{A5}$$

where L_p is a parameter termed the limit for potential evaporation. Potential evaporation was estimated by the modified Blaney-Criddle method (DVWK, 1996) as a function of air temperature. This representation of potential evaporation was compared to other methods in Parajka et al. (2003) suggesting that it gives plausible results in Austria.

Edited by: L. Ferraris
Reviewed by: G. Boni and two other anonymous referees

References

Blöschl, G. and Zehe, E.: On hydrological predictability. INVITED commentary, Hydrol. Processes, 19(19), 3923–3929, 2005.

Blöschl, G., Reszler, Ch., and Komma, J.: Hochwasservorhersage Kamp – Hydrologie (Kamp flood forecasting system – hydrology), Final Report to the Lower Austria State Government and EVN Energy Supply, February 2006, Institute of Hydraulic and Water Resources Engineering, Vienna University of Technology, 2006.

Buizza, R., Hollingsworth, A., Lalaurette, F., and Ghelli, A.: Probabilistic Predictions of Precipitation Using the ECMWF Ensemble Prediction System, Weather and Forecasting, 14, 168–189, 1999.

Buizza, R.: Weather Prediction: Ensemble Prediction, Encyclopaedia of Atmospheric Sciences, Academic Press, London, pp. 2546–2557, 2003.

Chiew, F. H. S., Peel, M. C., McMahon, T. A., and Siriwardena, L. W.: Precipitation elasticity of streamflow in catchments across the world. Climate Variability and Change-Hydrological Impacts, Proceedings of the Fifth FRIEND World Conference held at Havana, Cuba, November 2006, IAHS Publ. 308, pp. 256–262, 2006.

DVWK: Ermittlung der Verdunstung von Land- und Wasserflächen, DVWK-Merkblätter, Heft 238, Bonn, 1996.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99(C5), 10 143–10 162, 1994.

Golding, B. W.: Nimrod: A system for generating automated very short range forecasts, Meteorol. Appl., 5, 1–16, 1998.

Gutknecht, D., Reszler, Ch., and Blöschl, G.: Das Katastrophenhochwasser vom 7. August 2002 am Kamp – eine erste Einschätzung (The August 7, 2002 – flood of the Kamp – a first assessment), Elektrotechnik und Informationstechnik, 119(12), 411–413, 2002.

Haiden, T., Kann, A., Stadlbacher, K., Steinheimer, M., and Wittmann, C.: Integrated Nowcasting through Comprehensive Analysis (INCA) – System overview. ZAMG report, 49p, http://www.zamg.ac.at/fix/INCA_system.doc, accessed 26 March 2007.

Haiden, T. and Steinheimer, M.: Improved nowcasting of precipitation based on convective analysis fields, Adv. Geosci., 10, 125–131, 2007,
http://www.adv-geosci.net/10/125/2007/.

Kann, A. and Haiden, T.: The August 2002 flood in Austria: sensitivity of precipitation forecast skill to area size and duration, Meteorol. Z., 14, 369–377, 2005.

Krzysztofowicz, R.: Integrator of uncertainties for probabilistic river stage forecasting: precipitation-dependent model, J. Hydrol., 249, 69–85, 2001.

Mason, S. J. and Graham, N. E.: Conditional Probabilities, Relative Operating Characteristics, and Relative Operating Levels, Weather and Forecasting, 14(5), 713–725, 1999.

Parajka, J., Merz, R., and Blöschl, G.: Regionale Wasserbilanzkomponenten für Österreich auf Tagesbasis (Regional water balance components in Austria on a daily basis), Österreichische Wasser- und Abfallwirtschaft, 57(3/4), 43–56, 2005.

Parajka, J., Merz, R., and Blöschl, G.: Estimation of daily potential evapotranspiration for regional water balance modeling in Austria, in: 11th International Poster Day and Institute of Hydrology Open Day "Transport of Water, Chemicals and Energy in the Soil – Crop Canopy – Atmosphere System", 20 November 2003, Bratislava, Slovakia, Published on CD-ROM, Slovak Academy of Sciences, ISBN 80-89139-02-7, pp. 299–306, 2003.

Reszler, Ch., Komma, J., Blöschl, G., and Gutknecht, D.: Ein Ansatz zur Identifikation flächendetaillierter Abflussmodelle für die Hochwasservorhersage (An approach to identifying spatially distributed runoff models for flood forecasting), Hydrologie und Wasserbewirtschaftung, 50(5), 220–232, 2006.

Schaake, J., Perica, S., Mullusky, M., Demargne J., Welles, E., and Wu, L.: Pre-processing of Atmospheric Forcing for Ensemble Streamflow Prediction, Proceedings of the the 84th AMS Annual Meeting held in Seattle, USA, January 2004, 5 p, http://ams.confex.com/ams/pdfpapers/72172.pdf, accessed 18 December 2006, 2004.

Siccardi F., Boni, G., Ferraris, L., and Rudari, R.: A hydrometeorological approach for probabilistic flood forecast, J. Geophys. Res., 110(D5), D05101, doi:10.1029/2004JD005314, 2005.

Taylor, J. W. and Buizza, R.: Using weather ensemble predictions in electricity demand forecasting, Int. J. Forecasting, 19, 57–70, 2003.

Wang, Y., Haiden, T., and Kann, A.: The operational limited area modelling system at ZAMG: ALADIN-AUSTRIA, Österr. Beiträge zu Meteorologie und Geophysik, 37, 33 p, 2006.

Whitaker, J. S. and Loughe, A. F.: The relationship between ensemble spread and ensemble mean skill, Mon. Wea. Rev., 126, 3292–3302, 1998.