Journal of Hydrology 375 (2009) 613–626

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Review

Ensemble flood forecasting: A review

H.L. Cloke [a,*], F. Pappenberger [b]

[a] *Department of Geography, King's College London, Strand, London WC2R 2LS, UK*
[b] *European Centre for Medium Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, UK*

ARTICLE INFO

Article history:
Received 19 August 2008
Received in revised form 2 May 2009
Accepted 4 June 2009

This manuscript was handled by K. Georgakakos, Editor-in-Chief, with the assistance of Kieran M. O'Connor, Associate Editor

Keywords:
Flood prediction
Uncertainty
Ensemble prediction system
Streamflow
Numerical weather prediction

SUMMARY

Operational medium range flood forecasting systems are increasingly moving towards the adoption of ensembles of numerical weather predictions (NWP), known as ensemble prediction systems (EPS), to drive their predictions. We review the scientific drivers of this shift towards such 'ensemble flood forecasting' and discuss several of the questions surrounding best practice in using EPS in flood forecasting systems. We also review the literature evidence of the 'added value' of flood forecasts based on EPS and point to remaining key challenges in using EPS successfully.

Contents

* Corresponding author. Tel.: +44 (0) 20 7848 2604; fax: +44 (0) 20 7848 2287.
E-mail address: hannah.cloke@kcl.ac.uk (H.L. Cloke).

0022-1694/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2009.06.005

Introduction

Flood protection and awareness have continued to rise on the political agenda over the last decade accompanied by a drive to 'improve' flood forecasts (Demeritt et al., 2007; DKKV, 2004; Parker and Fordham, 1996; Pitt, 2007; van Berkom et al., 2007). Operational flood forecasting systems form a key part of 'preparedness' strategies for disastrous flood events by providing early warnings several days ahead (de Roo et al., 2003; Patrick, 2002; Werner, 2005), giving flood forecasting services, civil protection authorities and the public adequate preparation time and thus reducing the impacts of the flooding (Penning-Rowsell et al., 2000). Many flood forecasting systems rely on precipitation inputs, which come initially from observation networks (rain gauges) and radar. However, for medium term forecasts (~2–15 days ahead), numerical weather prediction (NWP) models must be used, especially when upstream river discharge data is not available (Hopson and Webster, in press) or when the equipment or transmission of data fails as is often the case in extreme floods. In general NWPs are essential to establish longer leadtimes than the catchment concentration time, but even for shorter leadtimes NWPs provide added value in that predictions can be made for parts of the catchment near to the outlet/point of reference.

Operational and research flood forecasting systems around the world are increasingly moving towards using ensembles of NWPs, known as ensemble prediction systems (EPS), rather than single deterministic forecasts, to drive their flood forecasting systems. This usually involves using EPS as input to a hydrological and/or hydraulic model to produce river discharge predictions (Fig. 1), often supported by some kind of Decision Support System (Fig. 2).

Several different hydrological and flood forecasting centres now use EPS operationally or semi-operationally (Table 1; note that not all ensemble forecasts are publicly available), and many other centres may be considering the adoption of such an approach (Bürgi, 2006; Rousset Regimbeau et al., 2006; Sene et al., 2007). The move towards ensemble prediction systems (EPS) in flood forecasting represents the state of the art in forecasting science, following on the success of the use of ensembles for weather forecasting (Buizza et al., 2005) and paralleling the move towards ensemble forecasting in other related disciplines such as climate change predictions (Collins and Knight, 2007). For hydrological prediction in general, the Hydrologic Ensemble Prediction Experiment (HEPEX) initiative has been setup to investigate how best to produce, communicate and use hydrologic ensemble forecasts (Schaake, 2006, Schaake et al., 2006, 2005, 2007), which are now often referred to as Ensemble Streamflow predictions (ESP) (note that historically the term ESP was used in a slightly different context, usually for longer term predictions (seasonal to yearly) of total volume or peak flows, and with ensembles created from an analysis of historical observations rather than ensemble weather forecasts (Twedt et al., 1977; Day, 1985) – for more discussion see (Seo et al., 2006; Wood and Schaake, 2008)).

In addition, other international bodies are demonstrating their interest in ensemble predictions for hydrological prediction, for

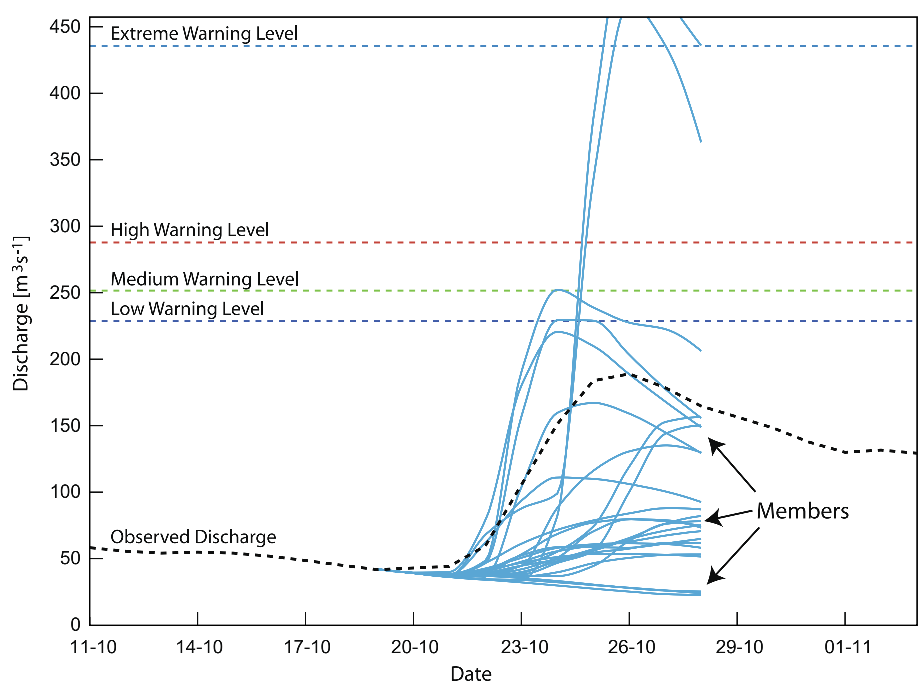

Fig. 1. An example of an ensemble 'spaghetti' hydrograph for a hindcasted flood event. The plot shows the discharge predicted for each ensemble forecast (solid line), the observed discharge (dashed black line) and four flood discharge warning levels (horizontal dashed lines). The example is taken from the October 2007 Flood in Romania on the river Jiu as modelled by the European Flood Alert System (for details of this flooding event see Pappenberger et al., 2008).

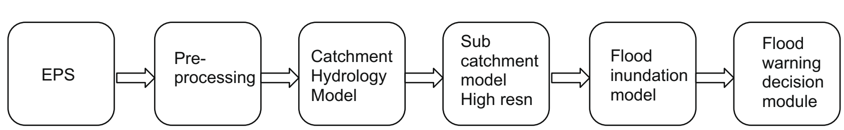

Fig. 2. A possible flood forecasting cascade, showing a cascade of components. Note that not every flood modelling system driven by EPS will have exactly these components; this remains an example, and we have purposefully not included other possible downstream components such as 'warning dissemination' and 'coordination of flood protection measures' as these are beyond the scope of this review.

Table 1
Examples of operational and pre-operational flood forecasting systems routinely using ensemble weather predictions as inputs.

Forecast centre	Ensemble NWP input	Further information
European Flood Alert System (EFAS) of the European Commission Joint Research Centre	European Centre for Medium Range Weather Forecasts (ECMWF) and Consortium for Small-Scale Modelling – Limited-area Ensemble Prediction System (COSMO-LEPS)	Thielen et al. (2009a)
Georgia-Tech/Bangladesh project	ECMWF	Hopson and Webster (2008, in press)
Finnish Hydrological Service	ECMWF	Vehvilainen and Huttunen (2002)
Swedish Hydro-Meteorological Service	ECMWF	Johnell et al. (2007), Olsson and Lindstrom (2008)
Advanced Hydrologic Prediction Services (AHPS) from NOAA	US National Weather Service (NOAA)	http://www.nws.noaa.gov/oh/ahps/; Mcenery et al. (2005)
MAP D-PHASE (Alpine region)/Switzerland	COSMO-LEPS	Rotach et al. (2008)
Vituki (Hungary)	ECMWF	Balint et al. (2006)
Rijkswaterstaat (The Netherlands)	ECMWF, COSMO-LEPS	Kwadijk (2007), Renner and Werner (2007), and Werner (2005)
Royal Meterological Institue of Belgium	ECMWF	Roulin (2007) and Roulin and Vannitsem (2005)
Vlaamse Milieumaatschappij (Belgium)	ECMWF	http://www.overstromings-voorspeller.be (and also Cauwenberghs, 2008)
Météo France	ECMWF and Arpege EPS	Regimbeau et al. (2007) and Rousset-Regimbeau et al. (2008)
Land Oberoestereich, Niederoestereich, Salzburg, Tirol (Austria)	Integration of ECMWF into Aladin	Haiden et al. (2007), Komma et al. (2007) and Reszler et al. (2006)
Bavarian Flood Forecasting Centre	COSMO-LEPS	Hangen-Brodersen et al. (2008)

example, the International Commission for the Hydrology of the Rhine Basin (CHR) and the World Meteorological Organization (WMO)'s Expert Consultation in March 2006 on 'ensemble predictions and uncertainties in flood forecasting', and the International Commission for the Protection of the Danube River's (ICPDR) recent move to adopt the ensemble forecasts of the European Flood Alert System (EFAS) in their flood action plan.

However, there is currently no rigorous critique of the scientific drivers of the move towards the use of EPS in medium range flood forecasting, and in addition there remain many assumptions in the practice of this, for example, the over-reliance on a disjointed set of case studies for evaluation (see later discussion). In this paper we address these issues and outline some of the challenges ahead. First we review the reasons why ensembles of NWPs are so attractive for flood forecasting systems. Second we discuss how uncertainty is represented in, and cascaded through, these systems and some of the assumptions behind these methodologies. Third we discuss the methods used to calculate flood forecasts probabilistically. Fourth, we review the case studies in this field which mostly find that ensemble prediction gives useful information ('added value') for flood early warning and highlight some weaknesses in current practice. Finally we identify the key challenges of using EPS for flood forecasting.

What are EPS and why are they attractive for flood forecasting systems?

The atmosphere is a non-linear and complex system and it is therefore impossible to predict its exact state (Lorenz, 1969). Weather forecasts remain limited by not only the numerical representation of the physical processes, but also the resolution of the simulated atmospheric dynamics and the sensitivity of the solutions to the pattern of initial conditions and sub-grid parameterization (Buizza et al., 1999). Over the last 15 or so years, ensemble forecasting techniques (EPS) have been used to take account of these uncertainties and result in multiple weather predictions for the same location and time (Palmer and Buizza, 2007). This makes EPS forecasts an attractive product for flood forecasting systems with the potential to extend leadtime and better quantify predictability.

The techniques for producing these EPS forecasts are fairly straightforward. Many operational EPS are based on a Monte Carlo framework of NWPs with one realisation starting from a 'central' analysis (the control forecast) and others generated by perturbing the initial conditions (the perturbed forecasts). The number of ensemble members usually varies from 10 to 50 depending on the forecast centre. Initial uncertainty is created by singular vectors (Bourke et al., 2004; Buizza and Palmer, 1995), Ensemble Transform or an Ensemble Transform Kalman Filter approach (Bishop et al., 2001; Bowler et al., 2007; Wei et al., 2006) or empirical orthogonal function based methods (Zhang and Krishnamurti, 1999). Some EPS also additionally incorporate parameter uncertainty in the generation of the ensemble forecasts (Buizza et al., 1999; Houtekamer and Lefaivre, 1997; Shutts, 2005). Moreover, regional EPS exist, which are nested into global EPS to provide EPS forecasts on a smaller spatial scale. An example is COSMO-LEPS, which is a limited-area non-hydrostatic model developed within the framework of the Consortium for Small-Scale Modelling (Germany, Switzerland, Italy, Poland and Greece) and nested on members of the ECMWF global ensemble. The limited-area ensemble forecasts range up to 120 h ahead and ensemble forecasts are produced (Marsigli et al., 2001, 2008). The reader is referred to Park et al. (2008) and Palmer and Buizza (2007) for a summary of different forecast centres issuing EPS meteorological forecasts.

Recent changes in the way that EPS precipitation forecasts are produced means that they are continuing to improve, as seen, for example, in the improvements in 500 hPa geopotential height and precipitation predictions for the new ECMWF ensemble set (Buizza et al., 2005). The increase in forecast skill can be fairly substantial, for example, the ECMWF deterministic model over Europe shows an improvement in precipitation forecast skill (equitable threat score) of roughly 1 day per 7 years for a threshold of 10 mm/24 h (Miller, pers. comm.). However, it is worth noting that in some cases, smaller scale improvements in precipitation forecasts are not evident, which is unfortunate for hydrological applications. For example, Goeber et al. (2004) show that there has been no substantial increase in the skill of precipitation forecasts over the last 8 years for a model by the UKMO for an area over the UK and using a threshold of 4 mm/6 h using the Odds ratio (see also Casati et al., 2008). However, they do find an improvement of bias in which the model no longer produces predominantly large areas of slight precipitation but more realistic, concentrated areas of higher precipitation amounts (however the location of these small scale events remains a problem).

Overall, although precipitation predictions may be improving, these and other predictors from NWPs still require improvement, and the impact of these improvements on hydrological models is uncertain.

It is often thought that a substantial increase in resolution of the models will allow resolution of rainfall cells in predictions and thus remove some of the large errors (Buizza et al., 1999; Undén, 2006). One of the biggest challenges therefore in improving forecasts remains to increase the resolution and identify the adequate physical representations on the respected scale, but this is a resource hungry task. Rather surprisingly perhaps, despite being in the age of supercomputer centres, such as ECMWF's High Performance Computing Facility (HPCF), computing power and storage still limit the production of more advanced ensemble sets and higher resolution forecasts. As a compromise researchers have attempted to cluster EPS for flood predictions in various ways, e.g. by lagging ensembles and deriving representative members through hierarchical clustering over the domain of interest, and thus to produce a reduced ensemble set at higher resolution (Cluckie et al., 2006; Ebert et al., 2007; Marsigli et al., 2001, 2008; Thirel et al., 2008).

An ensemble of weather forecasts can also be constructed from forecasts from many different forecast centres (often known as a 'poor man's ensemble'). For example, Jasper et al. (2002) have used the forecasts provided by five different forecast models in predicting inflow into Lake Maggiore in Italy. Davolio et al. (2008) used six different precipitation forecasts to predict floods in Northern Italy. Such a strategy acknowledges the uncertainty in model structure and variations in meteorological data assimilation. However, strictly speaking forecasts from different models have different error structures and thus cannot be easily combined, although we would argue that using this information with known error structures and restrictions is still more useful than not using it at all.

Another promising approach to the use of EPS in flood forecasting is by using lagged ensembles in which a previous forecast is combined with a more recent one (Dietrich et al., 2008; Hoffman and Kalnay, 1983). This is also effectively the same idea as a persistency forecast (as used in the European Flood Alert System, (Bartholmes et al., in press), in which a warning is only issued if a certain persistency (re-occurrence of the event) is reached).

In order to capture the uncertainties in initial conditions and parameterisations of individual NWP models together with the uncertainties in structure and data assimilation, an excellent strategy is to use a 'grand-ensemble', which means using several EPS together. This is the strategy behind the TIGGE (THORPEX Interactive Grand Global Ensemble) network (Park et al., 2008; Richardson, 2005) which aims to provide a collaboration platform on which to improve development and understanding of ensemble weather predictions from around the world. The TIGGE network now covers large parts of the globe with a detail adequate for flood forecasting (Pappenberger et al., 2008), and will most likely become an intensely used archive.

In order to use EPS in flood forecasting systems some kind of meteorological pre-processing is usually required and thus the meteorological input used by the hydrological model is not equivalent to the original EPS forecasts. First, scale corrections are required, as the time/space scale of the hydrological model will not match the scale of the meteorological model (EPS are not at high enough resolution for this yet, although limited-area prediction such as COSMO-LEPS are moving in the right direction for applications in large catchments). The EPS forecasts are usually therefore downscaled or disaggregated in some way. Second, the ensemble may need to have some kind of correction applied for under-dispersivity (i.e. not enough spread, and thus under-representation of uncertainty) or bias (difference between climatological statistics of ensemble predictions and corresponding statistics of related observations) (Hagedorn et al., 2007). With the latter, Hagedorn et al. (2005) have found that bias-correction does not necessarily lead to an increase in forecast skill, and bias-correction at the input stage may not be the most appropriate method of dealing with bias. However, Fortin et al. (2006) demonstrated that it is possible to significantly improve precipitation and temperature forecasts by pre-processing EPS as input into a hydrological model. Alternatively bias-correction can be dealt with following the propagation of the EPS through the hydrological model (Hashino et al., 2007a; Schaake et al., 2007) (Carpenter and Georgakakos, 2001; Wood et al., 2002), or at the flood warning threshold stage (Reggiani, 2008; Thielen et al., 2009a). The reader is also referred to the work of (Smith et al., 1992; Seo et al., 2006; Wood and Schaake, 2008). Hashino et al. (2007b) argue that all bias-correction methods tend to improve the skill score but that the quantile mapping method tends to lead to the highest sharpness and discrimination.

The HEPEX initiative has identified the need for research into 'hydrological product generators' (post processors of model output designed considering end user needs) (Schaake et al., 2007). Although many correction techniques are well developed and research is progressing fast in this area, a comprehensive comparison of the various techniques for hourly or daily discharge data is outstanding and the individual benefits need to be evaluated.

In summary, EPS forecasts can readily be used as inputs to medium term flood forecasting systems, although it is clear that these precipitation predictions still require significant improvement. Pre-processing of EPS inputs can render them more useful. Grand-ensemble techniques, such as TIGGE, hold great potential for global scale forecasting, which can be essential, for example, for disaster relief preparedness. However, techniques to deal with the combination of models with different error structure have to be developed (see examples and discussions in Berliner and Kim, 2008; Bhowmik and Durai, 2008; Doblas-Reyes et al., 2005; Kantha et al., 2008; Ke et al., 2008; Krishnamurti et al., 1999, 2001; Tippett and Barnston, 2008). In addition, the probabilistic nature of EPS can be particularly attractive when other data for driving flood forecasts is simply not available (Webster et al., submitted for publication) or when alternative anticipatory control measures are required (van Andel et al., 2008).

Capturing and cascading uncertainty

As discussed above, EPS are specifically designed to capture the uncertainty in NWPs, by representing a set of possible future states of the atmosphere. This uncertainty can then be cascaded through flood forecasting systems to produce an uncertain or probabilistic prediction of flooding. Over the last decade or so this potential is beginning to be realised in operational (or pre-operational) forecasting systems. However, there has been little rigorous critique of the main assumptions behind this methodology (cascading the uncertainty; calculating probability etc.). Here we discuss whether rare events such as floods can actually be represented by EPS-based flood forecasting systems, and whether EPS can represent the total uncertainty inherent in the cascaded predictions.

Representing and analysing rare flood events

In order to use EPS for flood forecasting effectively, it is important to establish methodologies to analyse ensemble discharge predictions, and indeed such 'Hydrological Forecast Verification' is one of the highlighted scientific issues of the HEPEX initiative (Schaake et al., 2007). Although evaluation of hydrological models and forecast skills is 'natural' for many hydrologists and land surface modellers and so the verification of the many aspects of flood forecasting systems has been addressed in the general hydrological community (see e.g. Seibert, 1999, 2001), this has not been carried out in an EPS specific context.

The value of hydrological forecasts (discharge, water stage, soil moisture etc) based on ensemble predictions can be evaluated (verified) with scores developed for meteorological applications such as the Brier Score (Jolliffe and Stephenson, 2003), continuous rank probability score (Hersbach, 2000) or the ignorance score (Roulston and Smith, 2002). Regimbeau et al. (2007) uses performance measures, including a rank histogram approach, which allows the quantification of the tendency to over or underpredict. Laio and Tamea (2007) promote evaluation methods based on cost/loss functions (although the exact shape of this cost/loss function may be disputed), which allows the comparison of the value of a deterministic forecast to a probabilistic forecast. However, although a general evaluation of hydrological forecasts based on EPS is possible, it is not straightforward to assess the use of EPS for flood forecasting purposes. Methods to evaluate rare events exist in meteorology (Doswell et al., 1990; Schaefer, 1990; Stephenson et al., 2008), for example the extreme dependency score (Stephenson et al., 2008). However, they are rarely adapted to hydrological circumstances and do not take account of the high autocorrelation which is observed in discharge (e.g. can one really score each discharge prediction in a flood hydrograph? – see discussion by Bogner and Kalas, 2008).

One major difficulty with using EPS for flood forecasting is that the evaluation of meteorological forecasts for hydrological applications (Cloke and Pappenberger, 2008), and thus the evaluation of the flood forecasts themselves, is fundamentally flawed by the low frequency of extreme floods:

i. Flood events are rare, and indeed hydrologists have been faced with the forecasting of rare events well before the advent of EPS. For example, a 1 in 100 year flood, which on most rivers poses potential risk to life and property, has a calculated statistical probability of only ~8% of happening at least 3 times in any period of 100 years. Many major flood events are not adequately measured, and spatial correlation is a major problem with the data that we do have. For example, in the year 2007, 31 major floods occurred in Europe (EM-DAT, 2008). The majority of these events happened at the same time and on the same rivers, but they merely occurred in different countries and so were marked as separate events. These events are too strongly correlated to be independent enough for any meaningful analysis. But even if we ignore spatial correlation and assume measurements are available everywhere, the low statistical probability of these extreme events means there will never be enough flood data to robustly statistically analyse flood predictions.
ii. Even if there was enough data from different flood events at different locations, this does not take into account spatial and temporal non-stationarity. Spatial stationarity cannot be assumed as each catchment is unique (Beven, 2000). Temporal stationarity cannot be assumed as for example, the form of a river bed often changes dramatically after flood events (Li et al., 2004). Changing trends in flood magnitude and frequency at particular locations have been observed in the last century, due to changes in vegetation, human-induced changes (such as dykes, landuse change), climate change and tectonic/isostatic relief change. Thus even consecutive floods cannot robustly be compared.
iii. Evaluation on medium size or mean river flow discharges has nothing to do with the performance of models at flood discharges, due to the non-linear flow processes occurring when a river goes out of bank. In addition, for flood discharge, it is most important to predict the peak of the hydrograph (where flows go overbank) in terms of timing and magnitude. Any use of the average discharge for a forecast period will therefore be unhelpful in this situation.
iv. Floods are seasonal, so we can never average discharge over the same time length as done in 'traditional' meteorology, and in addition the relationship between rainfall and flood discharge is complicated and thus a moving window approach to discharge is required rather than the 24 h total routinely used in meteorology.

The difficulties in assessing flood forecasts because of their rarity can be explained by looking at a contingency table (Table 2). The HIT and MISS fields have a very low frequency and it will be difficult to be statistically robust. The FALSE ALARM field will usually have a higher frequency, although the frequency of false alarms will depend on how realistically the system represents reality, with a very good system having a low frequency of false alarms. The NO EVENT field should have a high frequency.

Bartholmes et al. (2009) explain some of the difficulties in trying to calculate statistics for rare flood events over a 2 year operational period, including the relatively flood-prone year of 2006, for the European Flood Alert System. They note that often there were not enough events to fill in all the fields in the contingency table "which made it impossible to calculate skill scores like odds [...] that need values greater than zero in all fields" (p. 304). Also they found that the number of hits, false alarms and misses were very low, and so the number of positive rejects (no event) was very high, on the order of two magnitudes greater, which strongly affected the outcomes of some of the skill scores used (sets of skill scores are usually used for this and other reasons, see Cloke and Pappenberger, 2008). This is the same issue as in the well known Finley affair for forecasting tornadoes (Murphy, 1996), and the solution of how to best verify a flood forecasting system remains for the present unresolved.

There may be no other option than to analyse the performance of EPS driven flood forecasts on a case by case basis (Pappenberger et al., 2008) (Table 3), although advantage could be taken of specific hindcasting and reforecasting of flood events. Over time a database could be built up containing several hundred flood events on which to base a more thorough flood analysis, however, methodologies and requirements are sure to change over such a time period. Thus forecasters must work on the difficult task of verifying these EPS forecasts and communicating both the uncertainty, and the true value of these forecasts to end users.

Representing the total uncertainty

Many EPS are designed to comprise of equally likely (equiprobable) ensemble members, and to have an adequate number of ensemble members in order to describe the full range of input probabilities. However, EPS in their current format may not represent the full uncertainty of using NWPs to model atmospheric state. As discussed earlier, in many cases only the uncertainty in

Table 2
Contingency table for analysing flood forecasts.

	Observed	Not observed
Forecasted	HIT	FALSE ALARM
Not forecasted	MISS	NO EVENT

Table 3
Key case studies (hindcasts) evaluating ensemble flood forecasting. Note that some case studies concentrate on ESP and not floods specifically. Abbreviations are quoted in this paper as cited in the references.

Case study reference	Catchment/study area	Event/period	Hydrological model	Meteorological input
Balint et al. (2006) and Csík et al. (2007)	Main Danube in Hungary	July/August 2002	NHFS modelling system	EPS ECMWF (with 6 day lead time)
Bartholmes et al. (2007) and Bartholmes et al. (2009)	European Flood events	January 2005 until February 2007	Lisflood-FF (as input to the EFAS)	ECMWF (EPS and deterministic), DWD (global and local)
Bartholmes and Todini (2005)	Po river	October/November 1994	TOPKAPI	ECMWF EPS, HIRLAM EPS
Bogner and Kalas (2008)	Danube	July 2007	Lisflood (FF)	ECMWF (EPS and deterministic up to monthly), DWD (global and local), COSMO-LEPS
Bonta (2006)	Upper Tisza and central Hungary	March 2001 and August 2005	NHFS modelling system	ECMWF EPS
Cluckie et al. (2006)	Brue (in Southwest England)	October 1999, December 1999, April 2000	Simplified grid-based distributed rainfall–runoff model (GBDM)	ECMWF EPS and PSU/NCAR mesoscale model (MM5)
De Roo et al., 2006	Alpine region	August 2005	Lisflood (FF)	ECMWF (EPS and deterministic up to monthly), DWD (global and local)
Davolio et al. (2008)	Reno (in north Italy)	7–9th November 2003, 10–12th April, 2005, 2nd–3rd December 2005	TOPKAPI	Six different forcings (BOLAM, MOLOCH, LM7, LM2.8, WRF7.5, WRF2.5
Dietrich et al. (2008)	Mulde	August 2002	ArcEGMO (note there is also a short-range forecast presented using a large range of different models)	COSMO-LEPS and COMSO-DE
Gabellani et al. (2005)	Reno (in north Italy)	8–10th November 2003	DriFit	COSMO-LEPS
Gouweleeuw et al. (2005)	Meuse, Odra	January 1995 and July 1997	Lisflood-FF (as input to the EFAS)	ECMWF (EPS and deterministic), DWD (global and local)
He et al. (2009)	Upper Severn (UK)	January 2008	Lisflood-FF (here called Lisflood-RR)	TIGGE
Hlavcova et al. (2006)	Upper Hron (tributary to Danube)	August 1997 July 2002	Conceptual semi-distributed rainfall runoff model	ECMWF (EPS and deterministic), HIRLAM, DWD (global and local) and ALADIN
Hopson and Webster (in press)	Ganges and Brahamaputra	Summer 2003, 2004 and 2006	Catchment lumped model (CLM) and Semi distributed model (SDM)	ECMWF EPS
Jasper et al. (2002)	Ticino-Verzasca-Magiia (including smaller sub-catchments smallest 186 km^2)	September 1993 October 1993 October 1994 June 1997 September 1999 October 2000	WaSiM-ETH	Poor man ensemble consisting of Swiss Model, MESO-NH, BOLAM3, MC2, ALADIN
Jaun et al. (2008), Jaun and Ahrens (2009)	Rhine (Swiss part)	August 2005, 2005–2006	Precipitation Runoff Evapotranspiration Hydrotope (PREVAH)	COSMO-LEPS COSMO-LEPS
Johnell et al. (2007), Olsson and Lindstrom (2008)	51 Catchments in Sweden	January 2006–August 2007	HBV	ECMWF EPS
Kalas et al. (2008)	Morava	March–April 2006	Lisflood-FF (as input to the EFAS)	ECMWF (EPS and deterministic), DWD (global and local)
Komma et al. (2007)	Kamp, North Austria	August, 2002 (2 events) July, 2005 August, 2005 (two events)	Model of Reszler et al. (2006)	Combination of ECMWF and ALADIN
Pappenberger et al. (2005)	Meuse (upstream Masseik), Belgium	January 1995	Lisflood-FF, Lisflood-FP	ECMWF EPS
Pappenberger et al. (2008)	Danube, Romania	October 2007	Lisflood (FF)	TIGGE (grand-ensemble)
Regimbeau et al. (2007)	Seine, Herault	March 2001, September 2006	ISBA and MODCOU	ECMWF EPS
Roulin (2007), Roulin and Vannitsem (2005)	Ourthe (Meuse) and Scheldt, Belgium	All events 1997–2006	IRMB (adapted) water balance model	ECMWF EPS
Rousset-Regimbeau et al. (2008), Thirel et al. (2008)	France (on 881 gauges)	March 2005–September 2006	MODCOU	Prevision d'Ensemble ARPege and ECMWF EPS
Siccardi et al. (2005)	NW Italy, Liguria	November 1994	DriFit	LEPS (five clusters)
Thielen et al. (2009b)	Danube, Romania	October 2007	Lisflood (FF)	ECMWF (EPS and deterministic up to monthly), DWD (global and local), COSMO-LEPS
Verbunt et al. (2007)	Upper Rhine (up to Rhinefelsen)	May 1999 November 2002	PREVAH	ECMWF EPS, COSMO-LEPS
Webster et al. (submitted for publication)	Ganges and Brahamaputra	Summer 2007	Catchment lumped model (CLM) and Semi distributed model (SDM)	ECMWF EPS
Younis et al. (2008)	Elbe	March–April 2006	Lisflood-FF (as input to the EFAS)	ECMWF (EPS and deterministic), DWD (global and local)
Zappa et al. (2008)	Linth, Oglio (both in the Alps)	August and November 2007	DIMOSOP	COSMO-LEPS

initial conditions is considered, and only a few EPS incorporate parameter or data assimilation uncertainty. Thus model and observational error is currently ignored. It is possible therefore that the assumptions of equal probability are violated and the total uncertainty is underestimated (Golding, 2000).

It is difficult to establish whether the number of ensemble members used to drive flood forecasts is adequate. Atger (2001) claims that there is a small impact on skill if the number of ECMWF EPS ensembles is reduced from 50 to 21 for a precipitation forecast with a 4 day lead time. Jaun et al. (2008) claims that the benefits of the probabilistic approach for a flood forecasting system may be realized with a comparable small ensemble of only 10 members. However, there are so few case studies addressing this issue that no clear conclusions can be drawn. Experiments in other hydrological modeling exercises with respect to sampling size suggest that a far larger number than 50 is needed (Choi and Beven, 2007; Montanari, 2005; Pappenberger and Beven, 2004).

Meteorological input uncertainty is usually assumed to represent the largest source of uncertainty in the prediction of floods with a time horizon of beyond 2–3 days. However, there are in fact many sources of uncertainties further down in the flood forecasting cascade which could also be significant, for example: the corrections and downscaling mentioned above; spatial and temporal uncertainties as input into the hydrological antecedent conditions of the system (including data assimilation); geometry of the system (including flood defence structures); possibility of infrastructure failure (dykes or backing up of drains); characteristics of the system (in the form of model parameters); and in the limitations of the models available to fully represent processes (for example surface and sub-surface flow processes in the flood generation and routing). These are often termed collectively *model factors*.

The relative importance of the different types of uncertainty will most likely vary with the time (and lead time) of the forecasts, with the magnitude of the event and catchment characteristics. The sensitivity of runoff predictions towards input uncertainty differs for different case studies and thus no general trend is apparent (e.g. see references quoted in Michaud and Sorooshian, 1994; Segond, 2006). Komma et al. (2007) found for a case study in the Alps that for long lead times the uncertainty in the precipitation forecasts will always be amplified through their flood forecasting system. Olsson and Lindstrom (2008) found that the uncertainty is neither dampened nor amplified for catchments in Sweden. Generally it seems that input uncertainties propagating through the forecast system can be amplified or dampened (or neither) depending on the complex interaction of the different system components. For example, the variability of the precipitation inputs may be dampened due to the smoothing effects of the modelled catchments (Obled et al., 1994; Segond, 2006; Smith et al., 2004) which in turn is controlled by the type of dominant surface runoff process operating in the catchment, e.g. infiltration-excess vs. saturation-excess overland flow (Smith et al., 2004) and how each process is modelled (Segond, 2006).

The limits of predictability for a given catchment (identified as a HEPEX scientific issue) are discussed by Komma et al. (2007) and Thirel et al. (2008). These limits depend largely on the interaction between catchment response time, catchment characteristics and resolution of forcing data. For example, Thirel et al. (2008) show that an ensemble with a large resolution (e.g. ECMWF EPS) performs better for large catchments and low flows, whereas a high resolution ensemble (ARPege) is superior for small catchments and high flows.

It is well known that the sensitivity of the flow hydrograph to the uncertainty in rainfall decreases with catchment scale (Rodriguez-Iturbe and Mejia, 1974; Segond, 2006; Sivapalan and Bloschl, 1998; Woods and Sivapalan, 1999). The magnitude of the damping effect non-linearly interacts with the variability of the precipitation pattern. Smith et al. (2004) and Woods and Sivapalan (1999) point out that the distance-averaged rainfall excess needs to be considered in understanding the response of different catchment. Therefore, sensitivity towards precipitation uncertainty can be influenced by the storm movement through the catchment (e.g. Singh, 1997). If the damping effect is large and spatial patterns are of minor importance in the prediction of hydrographs then it is still vital to have accurate information on catchment average precipitation (Andreassian et al., 2004; Naden, 1992; Obled et al., 1994; Smith et al., 2004). The scale of averaging (e.g. size of subcatchments) has to be carefully explored (Dodov and Foufoula-Georgiou, 2005). The importance and sensitivity of the uncertainty of precipitation input is not static and changes spatially as well as temporally (e.g. seasonal due to soil moisture changes). Indeed, this connects the influence of rainfall uncertainty to the uncertainty in the observations. For example a sparse raingauge network has a stronger influence under dry then wet conditions (Shah et al., 1996a,b).

Towards an optimal framework for probabilistic flood predictions

One of the main drivers behind ensemble flood forecasting has been the potential to create and disseminate probabilistic forecasts, which is seen as an attractive 'state of the art' methodology to implement "politically" in operational systems (Sene et al., 2007). Scientifically, probabilistic forecasts are seen as being much more valuable than single forecasts "because they can be used not only to identify the most likely outcome but also to assess the probability of occurrence of extreme and rare events. Probabilistic forecasts issued on consecutive days are also more consistent than corresponding single forecasts" (Buizza, 2008). Thus, probabilistic flood forecasts are potentially very useful for obtaining estimates of flood risk (in its simplest form, probability of flood hazard $\times$ consequence), and methods such as cost–loss functions are geared for understanding this relationship (Laio and Tamea, 2007; Roulin, 2007).

However, for flood forecasting, and especially for severe events, there are still only a very limited number of studies that attempt to quantify the value of a probabilistic approach (see "Ensemble prediction gives useful information at medium term lead times"). In addition, there is currently little guidance on how to derive decisions based on such a complicated set of information (Demeritt et al., 2007). Cost–loss functions, although simple to use in principle (see Laio and Tamea, 2007; Richardson, 2000) do not necessarily lead to optimal decisions for rare events (Atger, 2001) and often fail to incorporate the full spectrum of expert judgments (for example public perception and trust), which is needed in taking action on issuing flood alerts. Doubts about whether ensemble flood forecasts truly represent probabilities and difficulties in easily evaluating their quality do little to reinforce the message that probabilistic flood forecasting is useful. In addition, even though probabilistic forecasts are potentially scientifically useful, they remain a relatively unfamiliar entity for many flood practitioners especially where traditional deterministic forecasts remain dominant in practice (Demeritt et al., 2007; Hlavcova et al., 2006; Zappa et al., 2008).

So which is the best framework to use for producing probabilistic forecasts?

It can be seen that the scales and interactions of those components involved in any flood forecasting system (model factors, inputs etc.) can strongly affect the flood predictions, and thus the nature of these components should of course influence the design of that flood forecasting system (Dietrich et al., 2008; Siccardi et al., 2005; Webster et al., for publication). Importantly, the uncertainty should be tracked using a *full uncertainty analysis* in order to give both the relative importance of various uncertainties in the system, but also the total uncertainty from the combination of each component in the uncertainty in the flood forecast at the end of the cascade (Krzysztofowicz, 2002a; Pappenberger et al., 2005). However, most common operational systems and most research exercises shy away from a full uncertainty analysis due to the intense computational demand that this would require (although a notable exception is Hopson and Webster (2008, in press)). However, even if a full uncertainty analysis cannot be performed some understanding of model sensitivities and uncertainties is a basic requirement in order to intelligently use flood forecast outputs. Moreover, computational burden can be reduced through using clustering techniques for ensemble input or model factors (Ebert et al., 2007; Pappenberger and Beven, 2004; Pappenberger et al., 2005).

Following the discussion in the previous sections, we argue that an optimal framework must be one that concentrates on cascading uncertainties through the flood modelling system. Most operational systems use tested *ad hoc* methods, which (i) fulfil the aim of the particular forecast system, (ii) fit to their historically grown systems and (iii) reflect what is computationally feasible at this particular organization. Depending on the complexity of the forecast system, these methods include routing the ensemble mean through a deterministic hydrological and hydraulic modelling system (Balint et al., 2006) and deterministic routing of all ensemble members through (optimized) hydrological/hydraulic models (Roulin and Vannitsem, 2005; Thielen et al., 2009a). Other operational forecast systems deal with the uncertainty cascade at the decision stage when binary warnings (warning or no warning) are required (Thielen et al., 2009a). Alternatively if hydrograph (exact discharge) predictions are required, an error model can be used to correct the hydrological forecast with observed discharge (Bogner and Kalas, 2008; Olsson and Lindstrom, 2008). Hopson and Webster (2008, in press) present one of the most all encompassing approaches to cascading uncertainty by using multiple corrections (at the input and output stage; 'multi-correction approach') as well as multiple hydrological models (multi-model approach).

Krzysztofowicz and co workers (2001, 2000, 1999, 2002a,b, 2004a,b) proposed a formal Bayesian approach to uncertainty analysis in order to treat the uncertainties in forecasting river stages in the short-range, which involved decomposing the uncertainty into input and hydrological uncertainty (see also Reggiani and Weerts, 2008). Beven et al. (2008) have argued that the formal Bayesian approach might produce misleading results and that the choice of a simple formal likelihood function might be 'incoherent' for real applications subject to input and model structural error (a 'coherent' method produces accurate and well-defined estimates of the parameters and shows convergence of parameter distributions as more data are added). Nevertheless, Krzysztofowicz's Bayesian methodology may be particular suitable for a real-time forecasting environment.

Alternatives to the formal treatment of cascading uncertainties includes a generalized Bayesian approach based on the GLUE methodology (Beven and Binley, 1992) presented by Pappenberger et al. (2005). However, research by Smith et al. (2008) indicates that a GLUE type approach may not be suitable for a real-time forecasting system. Moreover, Beven (2008) argues that as the aim of forecasting is to minimize the bias and variance, data assimilation techniques allow for wrong error assumptions about the error characteristics to be compensated for as a forecast proceeds. A mixture of formal and nonformal approaches is presented by Hopson and Webster (2008), who use a generalized Bayesian approach for most of the modelling chain, however, employ a more statistically rigorous approach in updating the error structure on hydrograph predictions. Reszler et al. (2006) update not only the error of the forecast, but also employ a Kalman Filter approach to update soil moisture.

In summary we argue that any 'optimal framework' will be inevitably a mixture of formal statistical treatments and informal treatment of some parts of the cascade. We suggest that treatments of individual components of the forecasting system will largely compensate for failings in other components. A full treatment of all uncertainties is not only prohibited by theoretical hinderances, but also by the computational burden that such approaches require.

It maybe possible to reduce the burden of treating these uncertainties, for example by simplifying models, (Romanowicz et al., 2008) or assuming fixed error distributions (Krzysztofowicz, 2002a). However, it is questionable whether it is possible to include and quantify all known and unknown uncertainties into such an analysis (Beven, 2008). Moreover multiple other factors will influence the treatment of a forecast system, for example:

- Availability of computer resources to store/retrieve forecasts.
- Spatial scale of predictions (point forecasts vs. spatially distributed).
- Type of prediction (hydrograph line, exceeding warning thresholds or spatially distributed inundation predictions).
- Uncertainty in the observations (for example a bypassed river gauge may lead to huge uncertainties in the observations and thus require a different treatment of the uncertainties).
- Reliability of observations (a failure of measurement equipment may for example render certain real-time updating routines useless).
- Domain (multiple or single catchments).

There is clearly the need both for more theoretical development of flood forecasting systems and a convincing all encompassing strategy for tackling the cascading of uncertainties in an operational framework. Currently, hydrological and hydraulic forecasts based on NWP EPS do not lead to proper probability distributions of any forecast variable. Thus the question remains whether it matters that uncertainties are not treated fully, that assumptions of some of the approaches are violated and that predictions are not true probabilities. These will most likely influence the numerical skill, accuracy and reliability of the system. However, it may well be that this degradation is very small and insignificant in respect to the modelling aim. Therefore it is important to analyse the ways that such a skill is computed and represented. Moreover, the usefulness of such an imperfect system will largely depend on the perception of the end user. This perception maybe partially influenced by the skill, accuracy and reliability of the system, but also by the perceived goodness of the complexity, "trustworthiness" of the issuing institution (influenced for example by disclosing all sources of uncertainty) or previous experiences.

Ensemble prediction gives useful information at medium term lead times

There are several case studies in the published literature that evaluate the use of ensemble prediction for flood forecasting by hindcasting observed flood/high discharge events, in many cases to test the potential or feasibility of a flood forecasting system based on ensemble inputs. We have listed these in Table 3 along with the basic attributes of each study. The majority of these are based on single hindcasted events, although a few do look at longer time series.

Evidence for added value in EPS driven flood forecasting systems

The case studies identified in Table 3 mainly indicate in their conclusions that there may be *added value* in using flood forecasting systems based on ensemble prediction systems, rather than just on single deterministic forecasts, especially in terms of issuing flood alerts or warnings. For example, Balint et al. (2006) wanted to test the feasibility of using ECMWF ensemble predictions to drive the National Hydrological Forecasting Service (NHFS) modelling system of VITUKI (Hungarian hydrological service) for the Danube proper. They analysed a 1–6 day ahead 'forecast' (hindcast) of the period of the flood events of August 2002, especially concentrating on the Devin and Budapest gauging stations. Using graphical analysis (spaghetti hydrographs, quartile boxplots and forecast distributions per lead time) and numerical analysis (Efficiency index and Nash-Sutcliffe criterion) they conclude that:

> "the use of meteorological ensembles to produce sets of hydrological predictions increased the capability to issue flood warnings" p. 67.

Roulin (2007) looked at longer term hindcasts (November 2000 to January 2006) for two Belgian catchments, and analysed a hydrological ensemble prediction system based on the use of ECMWF ensemble predictions driving a hydrological model. The study aimed to analyse the skill and the relative economic value of probability forecasts and used Brier Skill Scores to assess the skill of the probabilistic streamflow forecast and a cost–loss decision model to assess the value of the system. Roulin's conclusions regarding the skill of the ensemble predictions are positive:

> "The hydrological ensemble predictions have greater skills than deterministic ones". (Roulin, 2007, p. 736).

Similar statements can be found in many of the conclusions of the other case studies listed in Table 3 (following similar graphical and numerical analysis of skill etc.). For example,

> "Ensemble forecast provides a clear indication of the possible occurrence of the event" (Roulin and Vannitsem, 2005, p. 1389)
> "Even if the flood peak is first forecast with an error of one or two days [...] and is underestimated [...], the information given by the ensemble forecast can be of use for flood warning or water management agencies" (Regimbeau et al., 2007, p. 26).
> "Despite the overall poor performance for this particular case, it was shown that the ensemble of flow forecasts provides additional information to the deterministic forecast, i.e. the indication of the possibility of an extreme event" (Gouweleeuw et al., 2005, p. 379).

In many cases (including those quotes listed above) the potential of flood forecasting with EPS is described alongside cautious notes regarding variability, uncertainty, communication of ensemble information, need for decision support and problems of using short time series. For example:

> "in some cases both deterministic and ensemble forecasts gave a clear flood signal up to 4 days in advance, but there was a considerable variability in the forecasts, which would have to be reduced in the future. The analysis of longer time series would have been needed in order to adequately address uncertainty and usefulness of the ensembles. Also ways how to meaningfully interpret the ensembles and communicate the information to the users are not yet fully established". (Hlavcova et al., 2006, p. 89).

Bartholmes and Todini (2005) speculate that the added benefit of ensemble forecasts is not in quantitative flood forecasting (e.g. hydrograph predictions) but in the exceedance of warning levels. However, several authors (Hopson and Webster, in press; Olsson and Lindstrom, 2008) conclude to the contrary. The value of EPS seems to be more clearly seen in long term studies (e.g. Roulin (2007)). Jaun and Ahrens (2009) state based on a 2 year study of the upper Rhine that it "works for a wide band of weather conditions and that the ensemble spread nicely represents the additional uncertainty during weather situations with low predictability" (p. 1860). The authors demonstrate with rank probability skill scores (and other measures) that the approach of using ensemble weather predictions is superior to deterministic alternatives or 'statistical' ensembles (statistically-derived ensembles based on for example the ensemble mean and error structure).

Notably, several case studies conclude that the error in the precipitation predictions dominates the analysis:

> "the analysis suffers from under-performing rainfall predictions and therefore the value of the predictions is lessened" (Pappenberger et al., 2005, p. 391).
> "Precise quantitative precipitation forecasts are an absolute prerequisite to successful flood forecasting, [..] especially in alpine watersheds.[...] Precipitation must be predicted accurately in respect to timing, intensity, amount and spatial distribution. [. . .] NWP models do not capture true rainfall distributions."
> (Jasper et al., 2002) p. 50–p. 51.
> "Although the NWP based QPF could generally catch the rainfall pattern, the uncertainties of rainfall [. . .] are always significant." (Xuan et al., 2005, p. 8).
> "The results demonstrate the poor reliability of the quantative precipitation forecasts produced by meteorological models; this is not resolved by using the Ensemble Forecasting technique" (Bartholmes and Todini, 2005, p. 333).

Computational constraints still affect the resolution of the EPS driving the flood forecasts, and this can mean, for example, that the equivalent deterministic meteorological inputs are at a higher resolution and so:

> "the added value of the probabilistic forecast is therefore the combination with the deterministic forecast, rather than its replacement" (Gouweleeuw et al., 2005, p. 379).

Other general findings include the fact that smaller catchments demonstrate a larger uncertainty in the flood forecast, (Balint et al., 2006), as would be expected due to the smoothing effects of modelling a larger catchment. However, even in larger catchments EPS is 'beneficial' (Hlavcova et al., 2006). In addition, Regimbeau et al. (2007) show that in general flood forecasts driven by EPS have a large proportion of under and over predictions at low lead times and exhibit a negative bias at longer lead times (with forecast being higher than observations). This mimics the findings in meteorological forecasts. They also note that large catchments perform on average better at short lead times than smaller catchments, with this feature disappearing at longer lead times. This is to be expected as predictions for larger catchments will be dominated for a longer period of time by observed precipitation.

However, general literature agreement is that EPS flood forecasting is a useful activity and has the potential to inform early flood warning.

Weaknesses of current practice

We have identified several weaknesses in current practice evident in studies using ensemble predictions for flood forecasting. However note that within the case studies listed in Table 3, a few studies do already address these weaknesses, and thus this section should be seen as a general discussion which encourages future case studies to take certain important points into account:

(i) *It is rare for any case study to report a false alarm* ('failure') of their particular flood forecasting system, if the analysis is based on single case studies. A laudable exception is the study presented by Bartholmes and Todini (2005), who report that an ensemble prediction system did not have any additional value and performed poorly. It is of course possible (though unlikely) that other flood forecasting systems and the EPS used to drive them do not give false alarms. However, we feel that it is more likely due to a reluctance to report such false alarms. This may be due to institutional constraints (e.g. not wanting to criticise the meteorological data provider), because the set up of any such system is very labour-intensive and failure is therefore an undesirable outcome, or perhaps because hindcasts were based on known flood events and not continuous time series. We note that long term studies such as those of Bartholmes et al. (Bartholmes et al., 2009), Olsson and Lindstrom (2008) and Hopson and Webster (in press) are more open in their analyses of 'false alarms'. For example, Hopson and Webster find no high probability for "false positives". Another way of including information about false alarms is presented by Roulin (2007), who computed cost–loss which inherently includes the false alarm rate.

(ii) In many cases, *the published case studies report only qualitative statements on the positive impact of NWP EPS*, which often seems like rather a large leap from any quantitative/graphical results presented. For example, many papers remark that EPS-based forecasts are successful as some of (sometimes only one out of fifty for one single EPS forecast) the ensembles match the observations whereas as a single deterministic forecast does not. However, such success is only true if the decision to issue warnings and evaluation of quality is set in a proper framework, which in turn requires a long time series to establish (as for example noted by Roulin, 2007). Thus even a large number of positive case studies are only an early indication of a potentially successful system. We encourage more studies of long term series of cases such as that presented by Bartholmes et al. (2009), Hopson and Webster (2008), Olsson and Lindstrom (2008), Roulin (2007). In meteorology, where ensemble systems have been used in operational activities for more than a decade, this is common practice (Buizza et al., 2005; Park et al., 2008).

(iii) In some cases where a quantitative measure is attempted, *skill (or other measure) is calculated relative to a reference simulation driven by observed precipitation (and not against observed discharge)* (Pappenberger et al., 2008; Thielen et al., 2009a; Thirel et al., 2008). There may be good reasons for this, such as the unreliability or unavailability of the observed discharge time series, or the calibration of flood warning thresholds to model behaviour (and not observed discharge). However, it makes the comparison of these case studies difficult, and also the assumption that the imperfect modeling system behaves like the real hydrological system is questionable.

(iv) *Case studies are not directly comparable to each other.* Case studies are set in different hydrological and meteorological regimes. Moreover, forecasts systems (meteorological and hydrological) change over time. Therefore, interpretation of the results of case studies can change.

(v) *The contribution to forecast error/uncertainty by all of the different components of the system* is not estimated quantitatively or even qualitatively in most cases.

(vi) *The issue of decision support or communication of these forecasts to end users is not adequately considered.* Some case studies report how decisions for a particular flood event were considered (or would be considered in the case of hindcasting), for example through the use of threshold exceedance (Thielen et al., 2009a), and there is general agreement that appropriate decision support rules are needed to utilize the flood forecasts for flood management and warning purposes. (Balint et al., 2006). However, very little detail is provided on how these frameworks do actually work operationally.

Although the potential of flood forecasting driven by EPS is clear, the precise 'added value' and specifically the six points above, need to be addressed in future case studies.

Conclusions: key challenges of using EPS for flood forecasting

The use of ensemble flood forecasting is becoming a widespread activity. The case studies in the published literature give encouraging indications that such activity brings added value to medium-range flood forecasts, particularly in the ability to issue flood alerts earlier and with more confidence. However, the evidence supporting this is still weak, and many more case studies are needed. Reports of future case studies should be more quantitative in nature and in particular detail quantitative evidence for false alarms and contributions to uncertainty. EPS are in no way the magic solution to estimating the uncertainty of future rainfall and many further improvements are required, including with the EPS inputs themselves. Here we identify what we see as the six key challenges for the use of EPS for early flood warning. We also suggest particular scientific directions from which solutions might be forthcoming.

Key challenge 1: improving NWPs

The NWPs which currently form the EPS inputs to ensemble flood forecasting systems are not good enough, for example they need to be at higher resolution, have an increased number of ensemble members, and deal with current problems of bias and underdispersivity (for discussions on the future of NWPs see the special issue by Miller and Smolarkiewicz, 2008) and papers on the THORPEX iniative (Morss et al., 2008; Rabier et al., 2008). Databases such as the TIGGE archive, higher resolution forecasts such as using LEPS and the use of lagged ensemble techniques may be a short term solution. In the longer term, forecast centres should plan to increase the resolution of their forecasts and the number of ensemble members in their EPS to represent covariance of predictors and the extremes of distributions adequately (for a discussion please see Doblas-Reyes et al., 2008; Ferro et al., 2008; Richardson, 2001). In addition, the importance of flash flood forecasting demands greater research effort in relation to EPS. Observed information such as rain gauges and radar are only partially useful for flash flood predictions, whereas a high resolution NWP system has the potential to transform our ability to forecast flash floods.

Key challenge 2: understanding the total uncertainty in the system

We do not understand the full range and interaction of uncertainties in the forecast systems. A full uncertainty analysis has a high computational burden and moreover current EPS-based forecasts do not result in true probabilities of flooding, as uncertainties are not treated fully and the assumptions of some of the approaches are violated. Any 'optimal framework' will be inevitably a mixture of formal statistical treatments and informal treatment of some parts of the modeling cascade. Accounting for sources of uncertainty in the forecast system and the formulation of the system to account for all effects of uncertainty are two of the scientific issues identified by the HEPEX initiative. Many of the research questions that remain are similar to those in a non-EPS setting (see for example the Prediction of Ungauges Basins iniative Sivapalan et al., 2003).

Key challenge 3: data assimilation

Current data assimilation techniques in flood ensemble prediction systems include soil moisture (Komma et al., 2007), snow cover (Thielen et al., 2009a) or discharge, and some impact on hydrological skill can be seen (see also Wood and Lettenmaier, 2008). However, hydrological data assimilation deserves much more attention than it currently attracts, especially when it can be shown that the initial hydrological state has an important impact on the anticipated lead time. This is also highlighed as a HEPEX initiative scientific issue. In particular scientific studies in the estimation of sub-grid scale heterogeneity are underrepresented.

Key challenge 4: having enough case studies

Evaluating probabilistic forecasts is difficult and so we have to rely on case studies of rare flood events. We simply do not have enough at present (and may never do) to conduct a statistical analysis of the value of EPS driven flood forecasts. Thus we have to rely on the information that we have. In addition, a further consideration is what information can be extracted from particular case studies (event, forecasting system, catchment etc.) so that it can be combined with the results from another case study. However, 'the more the merrier': further case studies are essential, and this could be enhanced by reforecasting studies (Hamill et al., 2004). However, we encourage future case studies to take care to address the weaknesses in establishing 'added value' that we have noted in earlier discussion.

Key challenge 5: having enough computer power

The 'old chestnut' of computing resources still remains a millstone for EPS driven flood forecasting. This is especially important for running operational systems. The simple solution is to keep improving our computing resources wherever possible (such as the movements towards stochastic chip technology), or use them more efficiently such as using clusters of inputs or model factors as a compromise to the full EPS cascade.

Key challenge 6: learning how to use EPS in an operational setting

The approach of probabilistic forecasting in hydrology is still novel. Many organizations just recently adopted an ensemble based strategy. A period of several years will be needed in order to build up the know how of the practitioners and also within the hydrological forecasting agencies in order to fully incorporate benefits of these new operational flood forecasting tools (Zappa et al., 2008). The (pre-) operational services listed in Table 1 routinely derive decisions under uncertainty and display them in various formats. Optimized decision support systems are a major part of operational flood forecasting services. Scientifically more could be done to embed this process into current decision set theoretic frameworks such as described by Ben-Haim (2001). For the development of a decision support system and end user evaluations see (Ramos et al., 2007; Thielen and Ramos, 2006; Thielen et al., 2005). Many routines currently developed for the post processing of meteorological products are unsuitable for hydrological prediction systems. In addition, the balance between pre-and post-processing is poorly understood and requires future research, as does a comparison evaluating the cost benefit for hydrological forecasts. These 'Enduser Issues' such as designing Hydrological Product Generators (postprocessors designed for endusers), effective methods to describe and present results within these, and optimising decisions based on probabilistic information are highlighted scientific issues of the HEPEX initiative (Schaake et al., 2007).

Key challenge 7: communicating uncertainty and probabilistic forecasts

Related to, but distinct from key challenge 6, are the difficulties in communicating uncertain probabilistic flood forecasts alongside the assumptions that go into constructing them. Although the community recognises the need for 'enduser training' (Hlavcova et al., 2006), we still know relatively little about how best to go about this. Of course much will depend on who these endusers are. The only solution is to do more research, such as the focus group studies reported by (Demeritt et al., 2007). In addition, one of the scientific issues identified by the HEPEX initiative for hydrological forecast verification is assessing the "added value of human forecaster". It is clear that such social scientific research questions have not been adequately addressed and are absent from the research literature.

Acknowledgements

We gratefully acknowledge funding provided by the UK's ESRC (ES/F022832/1) and the UK's NERC Flood Risk from Extreme Events (FREE) Programme (NE/E002242/1) and the European PREVIEW and SAFER projects (FP6 and 7). We thank Roberto Buizza and Jutta Thielen and two anonymous reviewers for their valuable comments on this manuscript.

References

Andreassian, V. et al., 2004. Impact of spatial aggregation of inputs and parameters on the efficiency of rainfall–runoff models: a theoretical study using chimera watersheds. Water Resources Research 40 (5).

Atger, F., 2001. Verification of intense precipitation forecasts from single models and ensemble prediction systems. Nonlinear Processes in Geophysics 8, 401–417.

Balint, G., Csik, A., Bartha, P., Gauzer, B., Bonta, I., 2006. Application of meterological ensembles for Danube flood forecasting and warning. In: Marsalek, J., Stancalie, G., Balint, G. (Eds.), Transboundary Floods: Reducing Risks through Flood Management. Springer, NATO Science Series, Dordecht, The Netherlands, pp. 57–68.

Bartholmes, J., Thielen, J., Gentilini, S., 2007. Assessing operational forecasting skill of EFAS. In: Thielen, J, Bartholmes, J., Schaake, J., (Eds.), Abstracts of the 3rd HEPEX Workshop, Stresa, Italy, 27–29th June 2007, European Commission, Institute for Environment and Sustainability, EUR22861 EN.

Bartholmes, J., Thielen, J., Ramos, M., Gentilini, S., 2009. The European flood alert system EFAS – Part 2: statistical skill assessment of probabilistic and deterministic operational forecasts. Hydrology and Earth System Sciences 13 (2) 141–153.

Bartholmes, J., Todini, E., 2005. Coupling meteorological and hydrological models for flood forecasting. Hydrology and Earth System Sciences 9, 333–346.

Ben-Haim, Y., 2001. Information-Gap Decision Theory: Decisions Under Severe Uncertainty. Academic Press, San Diego.

Berliner, L.M., Kim, Y., 2008. Bayesian design and analysis for superensemble-based climate forecasting. Journal of Climate 21 (9), 1891–1910.

Beven, K.J., 2000. Uniqueness of place and process representations in hydrological modelling. Hydrology and Earth System Sciences 4, 203–213.

Beven, K.J., 2008. Environmental Modelling: An Uncertain Future? Routledge, London.

Beven, K.J., Binley, A., 1992. The future of distributed models – model calibration and uncertainty prediction. Hydrological Processes 6 (3), 279–298.

Beven, K.J., Smith, P., Freer, J., 2008. So just why would a modeller choose to be incoherent? Journal of hydrology 354, 15–32.

Bhowmik, S.K.R., Durai, V.R., 2008. Multi-model ensemble forecasting of rainfall over Indian monsoon region. Atmosfera 21 (3), 225–239.

Bishop, C.H., Etherton, B.J., Majumdar, S.J., 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: theoretical aspects. Monthly Weather Review 129, 420–436.

Bogner, K., Kalas, M., 2008. Error correction methods and evaluation of an ensemble based hydrological forecasting system for the Upper Danube catchment. Atmospheric Science Letters 9 (2), 95–102.

Bonta, I., 2006. Using ensemble precipitation forecasts for hydrological purposes. Geophysical Research Abstracts 8, 06027.

Bourke, W., Buizza, R., Naughton, M., 2004. Performance of the ECMWF and the BoM ensemble systems in the southern hemisphere. Monthly Weather Review 132, 2338–2357.

Bowler, N.E., Arribas, A., Mylne, K.R., Robertson, K.B., 2007. The MOGREPS short-range ensemble prediction system. Part I: system description, MetOffice, Exeter, EX1 3PB, UK.

Buizza, R., 2008. The value of probabilistic prediction. Atmospheric Science Letters 9, 36–42.
Buizza, R. et al., 2005. A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. Monthly Weather Review 133 (5), 1076–1097.
Buizza, R., Miller, M., Palmer, T.N., 1999. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. Quarterly Journal of the Royal Meteorological Society 125 (560), 2887–2908.
Buizza, R., Palmer, T.N., 1995. The singular vector structure of the atmospheric general circulation. Journal of Atmospheric Sciences 52, 1434–1456.
Bürgi, T., 2006. Swiss Hydrological Forecast System – objectives and problem statement. In: Abstracts of CHR-Workshop Expert Consultation on 'Ensemble Predictions and Uncertainties in Flood Forecasting', Bern, Switzerland, 30 and 31 March 2006, p. 13.
Carpenter, T.M., Georgakakos, K.P., 2001. Assessment of Folsom lake response to historical and potential future climate scenarios: 1. Forecasting. Journal of Hydrology 249 (1–4), 148–175.
Casati, B., Wilson, L.J., Stephenson, D.B., 2008. Forecast verification: current status and future directions. Meteorological Applications 15 (1), 3–18.
Cauwenberghs, K., 2008. Personal conversation.
Choi, H.T., Beven, K.J., 2007. Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework. Journal of Hydrology (NZ) 332 (3–4), 316–336.
Cloke, H., Pappenberger, F., 2008. Evaluating forecasts of extreme events for hydrological applications: an approach for screening unfamiliar performance measures. Meteorological Applications 15, 181–197.
Cluckie, I.D., Xuan, Y., Wang, Y., 2006. Uncertainty analysis of hydrological ensemble forecasts in a distributed model utilising short-range rainfall prediction. Hydrology and Earth System Sciences – Discussion 3, 3211–3237.
Collins, M., Knight, S., 2007. Ensembles and probabilities: a new era in the prediction of climate change. Philosophical Transactions of the Royal Society A, 1471–2962.
Csík, A., Bálint, G., Bartha, P., Gauzer, B., 2007. Application of meteorological ensembles for Danube flood forecasting and warning. In: Thielen, J, Bartholmes, J., Schaake, J., (Eds.), Abstracts of the 3rd HEPEX Workshop, Stresa, Italy, 27–29th June 2007. European Commission, Institute for Environment and Sustainability, EUR22861 EN.
Davolio, S. et al., 2008. A meteo-hydrological prediction system based on a multi-model approach for precipitation forecasting. Natural Hazards Earth System Sciences 8, 143–159.
Day, G.N., 1985. Extended streamflow forecasting using NWSRFS. Journal of Water Resources Planning and Management 111 (2), 157–170.
de Roo, A. et al., 2003. Development of a European flood forecasting system. International Journal of River Basin Management 1, 49–59.
De Roo, A., et al., 2006. The Alpine floods of August 2005. What did EFAS forecast, what was observed, which feedback was received from end-users? EFAS 25, Post-event summary report, European Commission, EUR 22154 EN, p. 94.
Demeritt, D. et al., 2007. Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting. Environmental Hazards 7 (2), 115.
Dietrich, J. et al., 2008. Combination of different types of ensembles for the adaptive simulation of probabilistic flood forecasts: hindcasts for the Mulde 2002 extreme event. Nonlinear Processes in Geophysics 15, 275–286.
DKKV, 2004. Flood Risk Reduction in Germany. Lessons Learned from the 2002 Disaster in the Elbe Region. Summary of the Study. Deutsches Komitee fuer Katastrophen vorsorge (DKKV) (German Committee for Disaster Reduction). Publication 29e. Bonn.
Doblas-Reyes, F.J., Coelho, C.A.S., Stephenson, D.B., 2008. How much does simplification of probability forecasts reduce forecast quality? Meteorological Applications 15 (1), 155–162.
Doblas-Reyes, F.J., Hagedorn, R., Palmer, T.N., 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting – II. Calibration and combination. Tellus Series A – Dynamic Meteorology and Oceanography 57 (3), 234–252.
Dodov, B., Foufoula-Georgiou, E., 2005. Fluvial processes and streamflow variability: interplay in the scale-frequency continuum and implications for scaling. Water Resources Research 41 (5).
Doswell, C.A., Daviesjones, R., Keller, D.L., 1990. On summary measures of skill in rare event forecasting based on contingency-tables. Weather and Forecasting 5 (4), 576–585.
Ebert, C., Bardossy, A., Bliefernicht, J., 2007. Selecting members of an EPS for flood forecasting systems by using atmospheric circulation patterns. Geophysical Research Abstracts, European Geosciences Union, 9: 08177h.
EM-DAT, 2008. The OFDAI CRED International Disaster Database. Universite Catholique de Louvein, Belgium, Brussels.
Ferro, C.A.T., Richardson, D.S., Weigel, A.P., 2008. On the effect of ensemble size on the discrete and continuous ranked probability scores. Meteorological Applications 15 (1), 19–24.
Fortin, V., Favre, A.C., Said, M., 2006. Probabilistic forecasting from ensemble prediction systems: improving upon the best-member method by using a different weight and dressing kernel for each member. Quarterly Journal of the Royal Meteorological Society 132 (617), 1349–1369.
Gabellani, S. et al., 2005. Applicability of a forecasting chain in a different morphological environment in Italy. Advances in Geosciences 2, 131–134.
Goeber, M., Wilson, C.A., Miltona, S.F., Stephenson, D.B., 2004. Fairplay in the verification of operational quantitative precipitation forecasts. Journal of Hydrology 288, 225–263.
Golding, B., 2000. Quantitative precipitation forecasting in the UK. Journal of Hydrology 239 (1–4), 286–305.
Gouweleeuw, B., Thielen, J., Franchello, G., de Roo, A., Buizza, R., 2005. Flood forecasting using medium-range probabilistic weather prediction. Hydrology and Earth System Sciences 9 (4), 365–380.
Hagedorn, R., Doblas-Reyes, F.J., Palmer, T.N., 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concept. Tellus Series A – Dynamic Meteorology and Oceanography 57 (3), 219–233.
Hagedorn, R., Hamill, T.M., Whitaker, J.S., 2007. Probabilistic forecast calibration using ECMWF and GFS ensemble forecasts. Part I: 2-meter temperature. Monthly Weather Review 136, 2608–2619.
Haiden, T., Kann, A., Stadlbacher, K., Steinheimer, M., Wittmann, C., 2007. Integrated nowcasting through comprehensive analysis (INCA) – system overview, ZAMG report. p. 49. <http://www.zamg.ac.at/fix/INCA_system.doc>.
Hamill, T., Whitaker, J., Wei, X., 2004. Ensemble reforecasting: improving medium-range forecast skill using retrospective forecasts. Monthly Weather Review 132, 1434–1447.
Hangen-Brodersen, C., Vogelbacher, A., Holle1, K.-K., 2008. Operational flood forecast in Bavaria, dealing with uncertainty of hydrological forecasts in the Bavarian Danube catchment. In: Proceedings of the 24th Conference of the Danubian Countries, 2–4 June 2008, Bled, Slovenia, <http://ksh.fgg.uni-lj.si/bled2008/cd_2008/>.
Hashino, T., Bradley, A.A., Schwartz, S.S., 2007a. Evaluation of bias-correction methods for ensemble streamflow volume forecasts. Hydrology and Earth System Sciences 11 (2), 939–950.
Hashino, T., Bradley, A.A., Schwartz, S.S., 2007b. Evaluation of bias-correction methods for ensemble streamflow volume forecasts. Hydrology and Earth System Science 11 (2), 939–950.
He, Y., Wetterhall, F., Cloke, H.L., Pappenberger, F., Wilson, M., Freer, J., McGregor, G., 2009. Tracking the uncertainty in flood alert driven by grand ensemble weather predictions. Meterological Applications 16 (1) 91–101.
Hersbach, H., 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather and Forecasting 15 (5), 559–570.
Hlavcova, K., Szolgay, J., Kubes, R., Kohnova, S., Zvolensky, M., 2006. Routing of numerical weather predictions through a rainfall–runoff model. In: Marsalek, J., Stancalie, G., Balint, G. (Eds.), Transboundary Floods: Reducing Risks through Flood Management. Springer, NATO Science Series, Dordecht, The Netherlands, pp. 57–68.
Hoffman, R.N., Kalnay, E., 1983. Lagged average forecasting, an alternative to Monte Carlo forecasting. Tellus A 35, 100–118.
Hopson, T., Webster, P., 2008. Three-Tier flood and precipitation forecasting scheme for south-east asia. <http://cfab2.eas.gatech.edu/> (accessed 08.05.08).
Hopson, T., Webster, P., in press. A 1–10 day ensemble forecasting scheme for the major river basins of Bangladesh: forecasting severe floods of 2003–2007. Journal of Hydrometeorology.
Houtekamer, P.L., Lefaivre, L., 1997. Using ensemble forecasts for model validation. Monthly Weather Review 125 (2416–2426).
Jasper, K., Gurtz, J., Lang, H., 2002. Advanced flood forecasting in Alpine watersheds by coupling meteorological observations and forecasts with a distributed hydrological model. Journal of Hydrology 267 (1–2), 40–52.
Jaun, S., Ahrens, B., 2009. Evaluation of a probabilistic hydrometeorological forecast system. Hydrology and Earth System Sciences 13, 1031–1043.
Jaun, S., Ahrens, B., Walser, A., Ewen, T., Schär, C., 2008. A probabilistic view on the August 2005 floods in the upper Rhine catchment. Natural Hazards Earth System Sciences 8, 281–291.
Johnell, A., Lindstrom, G., Olsson, J., 2007. Deterministic evaluation of ensemble streamflow predictions in Sweden. Nordic Hydrology 38 (4), 441–450.
Jolliffe, I.T., Stephenson, D.B., 2003. Forecast Verification: A Practitioner's Guide in Atmospheric Science. J. Wiley, Chichester, xiii, p. 240.
Kalas, M., Ramos, M.H., Thielen, J., Babiakova, G., 2008. Evaluation of the medium-range European flood forecasts for the March–April 2006 flood in the Morava river. Journal of Hydrology and Hydromechanics 56 (2), 116–132.
Kantha, L., Carniel, S., Sclavo, M., 2008. A note on the multimodel superensemble technique for reducing forecast errors. Nuovo Cimento Della Societa Italiana Di Fisica C-Geophysics and Space Physics 31 (2), 199–214.
Ke, Z.J., Dong, W.J., Mang, P.Q., 2008. Multimodel ensemble forecasts for precipitations in China in 1998. Advances in Atmospheric Sciences 25 (1), 72–82.
Komma, J., Reszler, C., Blöschl, G., Haiden, T., 2007. Ensemble prediction of floods – catchment non-linearity and forecast probabilities. Natural Hazards Earth System Sciences 7, 431–444.
Krishnamurti, T.N. et al., 1999. Improved weather and seasonal climate forecasts from multimodel superensemble. Science 285 (5433), 1548–1550.
Krishnamurti, T.N. et al., 2001. Real-time multianalysis-multimodel superensemble forecasts of precipitation using TRMM and SSM/I products. Monthly Weather Review 129 (12), 2861–2883.
Krzysztofowicz, R., 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. Water Resources Research 35 (9), 2739–2750.
Krzysztofowicz, R., 2001. Integrator of uncertainties for probabilistic river stage forecasting: precipitation-dependent model. Journal of Hydrology 249 (1–4), 69–85.
Krzysztofowicz, R., 2002a. Bayesian system for probabilistic river stage forecasting. Journal of Hydrology 268 (1–4), 16–40.
Krzysztofowicz, R., 2002b. Probabilistic flood forecast: bounds and approximations. Journal of Hydrology 268 (1–4). PII S0022-1694(02)00107-5.

Krzysztofowicz, R., Kelly, K.S., 2000. Hydrologic uncertainty processor for probabilistic river stage forecasting. Water Resources Research 36 (11), 3265–3277.

Krzysztofowicz, R., Maranzano, C.J., 2004a. Hydrologic uncertainty processor for probabilistic stage transition forecasting. Journal of Hydrology 293 (1–4), 57–73.

Krzysztofwicz, R., Maranzano, C.J., 2004b. Bayesian system for probabilistic stage transition forecasting. Journal of Hydrology 299 (1–2), 15–44.

Kwadijk, J.C.J., 2007. Assessing the value of ensemble prediction in flood forecasting – Report Q4136.40, Delft Hydraulics, Delft.

Laio, F., Tamea, S., 2007. Verification tools for probabilistic forecasts of continuous hydrological variables. Hydrology and Earth System Sciences 11 (4), 1267–1277.

Li, L.Q., Lu, X.X., Chen, Z.Y., 2004. River channel change during the last 50 years in the middle Yangtze River, the Jianli reach. Progress in Physical Geography 28 (3), 405–450.

Lorenz, E., 1969. The predictability of a flow which contains many scales of motion. Tellus A 21, 289–307.

Marsigli, C. et al., 2001. A strategy for high–resolution ensemble prediction. Part II: limited-area experiments in four alpine flood events. Quarterly Journal of the Royal Meteorological Society 127, 2095–2115.

Marsigli, C., Montani, A., Paccagnella, T., 2008. A spatial verification method applied to the evaluation of high-resolution ensemble forecasts. Meteorological Applications 15, 125–143.

Mcenery, J. et al., 2005. NOAA's advanced hydrologic prediction service: building pathways for better science in water forecasting. Bulletin of the American Meteorological Society 86, 375–385.

Michaud, J.D., Sorooshian, S., 1994. Effect of rainfall-sampling errors on simulations of desert flash floods. Water Resources Research 30 (10), 2765–2775.

Miller, M.J., Smolarkiewicz, P.K., 2008. Predicting weather, climate and extreme events. Journal of Computational Physics 227 (7), 3429–3430.

Montanari, A., 2005. Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall–runoff simulations. Water Resources Research 41. doi:10.1029/2004WR003826.

Morss, R.E. et al., 2008. Societal and economic research and applications for weather forecasts – Priorities for the North American THORPEX program. Bulletin of the American Meteorological Society 89 (3), 335+.

Murphy, A.H., 1996. The finley affair: a signal event in the history of forecast verification. Weather and Forecasting 11 (1), 3–20.

Naden, P.S., 1992. Spatial variability in flood estimation for large catchments – the exploitation of channel network structure. Hydrological Sciences Journal – Journal Des Sciences Hydrologiques 37 (1), 53–71.

Obled, C., Wendling, J., Beven, K., 1994. The sensitivity of hydrological models to spatial rainfall patterns – an evaluation using observed data. Journal of Hydrology 159 (1–4), 305–333.

Olsson, J., Lindstrom, G., 2008. Evaluation and calibration of operational hydrological ensemble forecasts in Sweden. Journal of Hydrology 350 (1–2), 14–24.

Palmer, T., Buizza, R., 2007. Fifteenth anniversary of EPS. ECMWF Newsletter 114 (Winter 07/08), 14.

Pappenberger, F. et al., 2008. New dimensions in early flood warning across the globe using grand-ensemble weather predictions. Geophysical Research Letters 35. doi:10.1029/2008GL033837.

Pappenberger, F., Beven, K., 2004. Functional classification and evaluation of hydrographs based on multicomponent mapping. International Journal of River Basin Management 2 (1), 1–8.

Pappenberger, F. et al., 2005. 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). Hydrology and Earth System Sciences 9 (4), 381–393.

Park, Y.-Y., Buizza, R., Leutbecher, M., 2008. TIGGE: preliminary results on comparing and combining ensembles, ECMWF, Reading, UK.

Parker, D., Fordham, M., 1996. Evaluation of flood forecasting, warning and response systems in the European Union. Water Resource Management 10 (279–302).

Patrick, S., 2002. The future of flood forecasting. Bulletin of the American Meteorological Society, p. 183.

Penning-Rowsell, E., Tunstall, S., Tapsell, S., Parker, D., 2000. The benefits of flood warnings: real but elusive, and politically significant. Journal of the Chartered Institution of Water and Environmental Management 14, 7–14.

Pitt, M., 2007. Learning lessons from the 2007 floods: an independent review by Sir Michael Pitt: interim report, London, UK.

Rabier, F. et al., 2008. An update on THORPEX-related research in data assimilation and observing strategies. Nonlinear Processes in Geophysics 15 (1), 81–94.

Ramos, M.H., Bartholmes, J., Thielen, J., 2007. Development of decision support products based on ensemble weather forecasts in the European flood alert system. Atmospheric Science Letters 8, 113–119.

Reggiani, P., 2008. HEPEX Workshop: Uncertainty Post-processing for Hydrologic Ensemble Prediction – http://hydis8.eng.uci.edu/hepex/uncertwrksp/uncertwksp.html. Deltares – WL Delft Hydraulics, Delft, The Netherlands, June 23–25, 2008, Delft.

Reggiani, P., Weerts, A., 2008. Probabilistic quantitative precipitation forecasts for flood prediction: an application. Journal of Hydrometeorology 9 (1). doi:10.1175/2007JHM858.1.

Regimbeau, F.R., Habets, F., Martin, E., Noilhan, J., 2007. Ensemble streamflow forecasts over France. ECMWF Newsletter 111, 21–27.

Renner, M., Werner, M., 2007. Verification of ensemble forecasting in the Rhine basin – Report Q4378, Delft Hydraulics, Delft.

Reszler, C., Komma, J., Bloschl, G., Gutknecht, D., 2006. Ein Ansatz zur Identifikation flachendetaillierter Abflussmodelle fur die Hochwasservorhersage (An approach to identifying spatially distributed runoff models for flood forecasting). Hydrologie und Wasserbewirtschaftung 50 (5), 220–232.

Richardson, D., 2005. The THORPEX interactive grand global ensemble (TIGGE). Geophysical Research Abstracts, 7. <http://www.wmo.ch/pages/prog/arep/thorpex/>.

Richardson, D.S., 2000. Skill and relative economic value of the ECMWF ensemble prediction system. Quarterly Journal of the Royal Meteorological Society 126, 649–667.

Richardson, D.S., 2001. Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. Quarterly Journal of the Royal Meteorological Society 127 (577), 2473–2489.

Rodriguez-Iturbe, I., Mejia, J.M., 1974. The design of rainfall network in time and space. Water Resources Research 10, 713–728.

Romanowicz, R., Young, P., Beven, K.J., Pappenberger, F., 2008. Data based mechanistic approaches to flood forecasting. Advances in Water Resources 31 (8), 1048–1056.

Rotach, M.W. et al., 2008. MAP D-PHASE: real-time demonstration of weather forecast quality in the alpine region. Atmospheric Science Letters 2, 80–87.

Roulin, E., 2007. Skill and relative economic value of medium-range hydrological ensemble predictions. Hydrology and Earth System Sciences 11 (2), 725–737.

Roulin, E., Vannitsem, S., 2005. Skill of medium-range hydrological ensemble predictions. Journal of Hydrometeorology 6 (5), 729–744.

Roulston, M.S., Smith, L.A., 2002. Evaluating probabilistic forecasts using information theory. Monthly Weather Review 130 (6), 1653–1660.

Rousset-Regimbeau, F., Noilhan, J., Thirel, G., Martin, E., Habets, F., 2008. Medium-range ensemble streamflow forecast over France. Geophysical Research Abstracts, 11. EGU2008-A-03111.

Rousset Regimbeau, F., Habets, F., Martin, E., 2006. Ensemble streamflow forecast over the entire France. Geophysical Research Abstracts, 8.

Schaake, J., 2006. Hydrologic ensemble prediction: past, present and opportunities for the future, ensemble predictions and uncertainties in flood forecasting. International Commission for the Hydrology of the Rhine Basin. Expert Consulation Workshop, 30 and 31 March 2006, Bern, Switzerland.

Schaake, J., Franz, K., Bradley, A., Buizza, R., 2006. The hydrologic ensemble prediction experiment (HEPEX). Hydrology and Earth System Sciences – Discussions 3, 3321–3332.

Schaake, J. et al., 2005. Hydrological ensemble prediction: challenges and opportunities. In: International Conference on Innovation Advances and Implementation of Flood Forecasting Technology. ACTIF/FloodMan/FloodRelief, Tromsø, Norway.

Schaake, J.C., Hamill, T.H., Buizza, R., Clark, M., 2007. HEPEX – the hydrological ensemble prediction experiment. Bulletin of the American Meteorological Society 88 (10). doi:10.1175/BAMS-88-10-1541.

Schaefer, J.T., 1990. The critical success index as an indicator of warning skill. Weather and Forecasting 5 (4), 570–575.

Segond, M.-L., 2006. Stochastic modelling of space-time rainfall and the significance of spatial data for flood runoff generation. Ph.D. Thesis, Imperial College London, London, 222 pp.

Seibert, J., 1999. On TOPMODEL's ability to simulate groundwater dynamics. Regionalization in Hydrology (254), 211–220.

Seibert, J., 2001. On the need for benchmarks in hydrological modelling. Hydrological Processes 15 (6), 1063–1064.

Sene, K., Huband, M., Chen, Y., Darch, G., 2007. Probabilistic flood forecasting scoping study. Joint Defra/EA Flood and Coastal Erosion Risk Management R&D Programme, R&D Technical Report FD2901/TR.

Seo, D.J., Herr, H.D., Schaake, J.C., 2006. A statistical post-processor for accounting of hydrologic uncertainty in short range ensemble streamflow predictions. Hydrology and Earth System Sciences – Discussions 3, 1987–2035.

Shah, S.M.S., Oconnell, P.E., Hosking, J.R.M., 1996a. Modelling the effects of spatial variability in rainfall on catchment response. 1. Formulation and calibration of a stochastic rainfall field model. Journal of Hydrology 175 (1–4), 67–88.

Shah, S.M.S., Oconnell, P.E., Hosking, J.R.M., 1996b. Modelling the effects of spatial variability in rainfall on catchment response. 2. Experiments with distributed and lumped models. Journal of Hydrology 175 (1–4), 89–111.

Shutts, G., 2005. A kinetic energy backscatter algorithm for use in ensemble prediction systems. Quarterly Journal of the Royal Meteorological Society 131, 3079–3100.

Siccardi, F., Boni, G., Ferraris, L., Rudari, R., 2005. A hydrometeorological approach for probabilistic flood forecast. Journal of Geophysical Research – Atmospheres, 110. D5.

Singh, V.P., 1997. Effect of spatial and temporal variability in rainfall and watershed characteristics on stream flow hydrograph. Hydrological Processes 11 (12), 1649–1669.

Sivapalan, M., Bloschl, G., 1998. Transformation of point rainfall to areal rainfall: intensity–duration frequency curves. Journal of Hydrology 204 (1–4), 150–167.

Sivapalan, M. et al., 2003. IAHS decade on predictions in ungauged basins (PUB), 2003–2012: shaping an exciting future for the hydrological sciences. Hydrological Sciences Journal – Journal Des Sciences Hydrologiques 48 (6), 857–880.

Smith, J.A., Day, G.N., Kane, M.D., 1992. Nonparametric framework for long-range streamflow forecasting. Journal of Water Resources Planning and Management – ASCE 118 (1), 82–92.

Smith, M.B. et al., 2004. Runoff response to spatial variability in precipitation: an analysis of observed data. Journal of Hydrology 298 (1–4), 267–286.

Smith, P., Beven, K.J., Tawn, J.A., 2008. Informal likelihood measures in model assessment: theoretic development and investigation. Advances in Water Resources 31, 1087–1100.

Stephenson, D.B., Casati, B., Ferro, C.A.T., Wilson, C.A., 2008. The extreme dependency score: a non-vanishing measure for forecasts of rare events. Meteorological Applications 15 (1), 41–50.

Thielen, J., Bartholmes, J., Ramos, M.-H., de Roo, A., 2009a. The European flood alert system – Part 1: concept and development. Hydrology and Earth System Science 13 (2), 125–140.

Thielen, J., Bogner, K., Pappenberger, F., Kalas, M., del Medico, M., de Roo, A., 2009b. Monthly-, medium-, and short-range flood warning: testing the limits of predictability. Meteorological Applications 16 (1), 77–90.

Thielen, J., Ramos, M.H., 2006. Enduser perception and feedback of early flood warnings provided by EFAS, The benefit of probabilistic flood forecasting on European scale – results of the European flood alert system for 2005/2006. European Report EUR 22560 EN, European Communities, Italy.

Thielen, J. et al., 2005. Summary report of the 1st EFAS workshop on the use of ensemble prediction system in flood forecasting, 21–22nd November 2005, Ispra. European Report EUR 22118 EN, European Communities, Italy, 23 p.

Thirel, G., Rousset-Regimbeau, F., Martin, E., Habets, F., 2008. On the impacts of short-range meteorological forecasts for ensemble stream flow predictions. Journal of Hydrometerology 9 (6), 1301–1317.

Tippett, M.K., Barnston, A.G., 2008. Skill of multimodel ENSO probability forecasts. Monthly Weather Review 136 (10), 3933–3946.

Twedt, T.M., Schaake, J.C., Peck, E.L., 1977. National weather service extended streamflow prediction. In: Proceedings, Western Snow Conference, Albuquerque, NM, April, p. 52–57

Undén, P., 2006. Global EPS systems – principles, use and limitations, ensemble predictions and uncertainties in flood forecasting. International Commission for the Hydrology of the Rhine Basin. Expert Consulation Workshop, 30 and 31 March 2006, Bern, Switzerland.

van Andel, S.J., Lobbrecht, A.H., Price, R.K., 2008. Rijnland case study: hindcast experiment for anticipatory water-system control. Atmospheric Science Letters 9 (2), 57–60.

van Berkom, F., van de Watering, C., de Gooijer, K., Neher, A., 2007. Inventory of flood information systems in Europe – a study of available systems in Western-, Central- and Eastern Europe, INTERREG IIIC Network Flood Awareness and Prevention Policy in Border Areas' (FLAPP), The Netherlands.

Vehvilainen, B., Huttunen, M., 2002. Hydrological forecasting and real time monitoring in Finland: The watershed simulation and forecasting system (WSFS). <http://www.environment.fi/waterforecast>.

Verbunt, M., Walser, A., Gurtz, J., Montani, A., Schar, C., 2007. Probabilistic flood forecasting with a limited-area ensemble prediction system: selected case studies. Journal of Hydrometeorology 8 (4), 897–909.

Webster, P., Hopson, T., Hoyos, C., Jian, J., Chang, H.-R., submitted for publication. Extended range probabilistic forecasting of the 2007, Bangladesh floods.

Wei, M. et al., 2006. Ensemble transform Kalman filterbased ensemble perturbations in an operational global prediction system at NCEP. Tellus A 58, 28–44.

Werner, M., 2005. FEWS NL Version 1.0 – Report Q3933, Delft Hydraulics, Delft.

Wood, A., Lettenmaier, D., 2008. An ensemble approach for attribution of hydrologic prediction uncertainty. Geophysical Research Letters 35. doi:10.1029/2008GL034648.

Wood, A.W., Maurer, E.P., Kumar, A., Lettenmaier, D.P., 2002. Long-range experimental hydrologic forecasting for the eastern United States. Journal of Geophysical Research – Atmospheres, 107. D20.

Wood, A.W., Schaake, J.C., 2008. Correcting errors in streamflow forecast ensemble mean and spread. Journal of Hydrometeorology 9 (2), 132–148.

Woods, R., Sivapalan, M., 1999. A synthesis of space-time variability in storm response: rainfall, runoff generation, and routing. Water Resources Research 35 (8), 2469–2485.

Xuan, Y., Cluckie, I.D., Han, D., 2005. Uncertainties in application of NWP-based QPF in real-time flood forecasting, Innovation, advances and implementation of flood forecasting technology, Tromso, Norway, pp. 1–9.

Younis, J., Ramos, M., Thielen, J., 2008. EFAS forecasts for the March–April 2006 flood in the Czech part of the Elbe River Basin – a case study. Atmospheric Science Letters 9 (2), 88–94.

Zappa, M. et al., 2008. MAP D-PHASE: real-time demonstration of hydrological ensemble prediction systems. Atmospheric Science Letters 2, 80–87.

Zhang, Z., Krishnamurti, T.N., 1999. A perturbation method for hurricane ensemble predictions. Monthly Weather Review 127, 447–469.