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)

F. Pappenberger[1], K.J. Beven[1], N.M. Hunter[2], P.D. Bates[2], B.T. Gouweleeuw[3], J. Thielen[3] and A.P.J. de Roo[3]

[1]Lancaster University, Institute of Environmental Science, Bailrigg, Lancaster, LA1 4YQ, UK

[2]University of Bristol, School of Geographical Sciences, Bristol, BS8 1SS, UK

[3]NASA/GSFC, Mail Code 614.3, Greenbelt, MD 20770, USA

Email for corresponding author: f.pappenberger@lancaster.ac.uk

Abstract

The political pressure on the scientific community to provide medium to long term flood forecasts has increased in the light of recent flooding events in Europe. Such demands can be met by a system consisting of three different model components (weather forecast, rainfall–runoff forecast and flood inundation forecast) which are all liable to considerable uncertainty in the input, output and model parameters. Thus, an understanding of cascaded uncertainties is a necessary requirement to provide robust predictions. In this paper, 10-day ahead rainfall forecasts, consisting of one deterministic, one control and 50 ensemble forecasts, are fed into a rainfall–runoff model (LisFlood) for which parameter uncertainty is represented by six different parameter sets identified through a Generalised Likelihood Uncertainty Estimation (GLUE) analysis and functional hydrograph classification. The runoff of these 52 * 6 realisations form the input to a flood inundation model (LisFlood-FP) which acknowledges uncertainty by utilising ten different sets of roughness coefficients identified using the same GLUE methodology. Likelihood measures for each parameter set computed on historical data are used to give uncertain predictions of flow hydrographs as well as spatial inundation extent. This analysis demonstrates that a full uncertainty analysis of such an integrated system is limited mainly by computer power as well as by how well the rainfall predictions represent potential future conditions. However, these restrictions may be overcome or lessened in the future and this paper establishes a computationally feasible methodological approach to the uncertainty cascade problem.

Keywords: uncertainty, flood forecasting, model coupling, calibration, GLUE, Lisflood

Introduction

The European Flood Forecasting System (EFFS) project, funded by the European Union, has the eventual aim of making continuous operational real-time forecasts of river discharges and inundation in the major rivers in Europe up to 10 days ahead, based on mesoscale rainfall forecasts provided by the European Medium Range Weather Forecasting Centre (ECMWF) and other national meteorological forecasting services (de Roo *et al*., 2003). Such a project requires a combination of modelling capabilities, including rainfall forecasting, runoff generation predictions based on those rainfall forecasts, and hydraulic flood wave propagation based on the runoff predictions. Taking account of uncertainty in all of these model components is important in risk-based decision making for issuing flood warnings, ordering evacuations, operation of flood control reservoirs and basins, planning for potential future change in flood behaviour, etc.

While uncertainty in predictions may well decrease in future as the different process representations in the models are improved, it will never be eliminated. Weather forecasts, for example, have greatly improved in the last two decades, as increased computer power and more integrated measurement systems have allowed general circulation

models to run at finer resolutions with improved initialisation. Rainfall forecasts, however, are still heavily dependent on sub-grid scale parameterisations of the rainfall forming processes; these cannot take complete account of understanding of the behaviour of rainfall cells in different meteorological conditions (see e.g. Hewitson and Crane, 1996; Joubert and Hewitson, 1997). Convective cells, in particular, cannot be resolved at the current mesoscale grid sizes of 10–40 km used in 5 to 10 day ahead predictions. Thus, rainfall forecasts are still limited by the resolution of the simulated atmospheric dynamics and the sensitivity of the solutions to the pattern of initial conditions and sub-grid parameterisations (Buizza et al., 1999; Downtown and Bell, 1988; Harrison et al., 1999)

There is a somewhat different problem in the forecasting of flood runoff. Hydrological models, whether using measured or forecast rainfalls, are also limited by their representation of the small scale flow dynamics but this is as much a problem of knowing the local characteristics (or parameters) of sub-catchment areas as it is of representing the dynamics themselves. Effective values of the parameters required by hydrological models, as affected by local spatial heterogeneities and nonstationarities, seem to have only loose associations with the type of soil, rock, topographic forms and vegetative cover for the hillslopes that make up a catchment area (see Beven, in press). Thus, it has proven very difficult to make accurate prior estimates of effective parameter values and to get good simulations of streamflow hydrographs without calibration of parameter values by comparison of model predictions with historical data (Gupta et al., 2003). The problem then is that distributed hydrological models with many different grid elements and many different parameters for each grid element do not provide a well-posed calibration problem when there are no discretisation scale observables with which to assess local parameter values (Beven, 1989, 2001). There will then be some inevitable uncertainty in the predictions of runoff generation, compounded by the uncertainty in the rainfall forecasts.

Similar considerations apply to hydraulic models. Grid scales are still large relative to the scales of momentum dissipation, turbulence and secondary currents caused by local heterogeneities in channel and flood plain geometry and cover, and this will inevitably affect the model parameters (Marks and Bates, 2000). Again, local effective values of model parameters may need to be calibrated but only limited observations may be available to achieve such a calibration. Hence, there will be some inevitable uncertainty in the predictions of flood inundation and flood wave propagation, compounded by the uncertainty in the runoff generation predictions (Aronica et al., 1998, 2002;

Pappenberger et al., in press-b; Romanowicz and Beven, 1998, 2003; Romanowicz et al., 1994,1996).

Uncertainty, therefore, cascades from rainfall forecasts, through the runoff generation prediction to the flood wave forecasts. Each stage of the cascade has to deal with non-linear transformations, from atmospheric conditions to rainfall forecasts, from rainfalls to runoff forecasts, and from runoff to flood wave and inundation forecasts. Thus, it is difficult to use traditional linear statistical methods for cascading the uncertainties through the forecasting system. It is, indeed, difficult to make a rigorous assessment of the uncertainties at all, because of sheer computational constraints, particularly in the atmospheric forecasts. Uncertainties in non-linear systems can often be estimated simply using some form of Monte Carlo simulation technique. However, in such a complex modelling system it is still computationally infeasible to perform such an analysis fully and thus first estimates of the magnitude of the uncertainty can be achieved only by some approximate methods.

There have been previous attempts to deal with uncertainties in real-time flood forecasting systems. Each has been associated with some form of approximation e.g. the Bayesian forecasting system (Krzysztofowicz, 2002; Kyriakidis et al., 2001); the US National Weather service ensemble forecasting (Koren et al., 1999); the River Flow Forecasting system (Moore et al., 1990); Transfer Function methods (Young, 2002; Young and Beven, 1994); the Generalized Likelihood Uncertainty Estimation (GLUE) method (Romanowicz and Beven, 1998, 2003; Romanowicz et al., 1994). However, all the references mentioned constrain the uncertainty analysis to only two model components (e.g. from weather prediction to a rainfall–runoff model, or from an upstream discharge to an inundation model) and no study is known to the authors which propagates the uncertainty from a weather forecast model fully down to an inundation model. Note that the GLUE analysis presented here depends not only on prior estimates of effective model parameter values but also on model conditioning based on performance evaluations against past observations.

The objective of this paper is to introduce a framework which cascades uncertainty through such an entire modelling system (Fig. 1). Each of these components is liable to uncertainty. In the following, the problems in each modelling component are introduced, starting with the rainfall forecast, then the rainfall-runoff model and finally the flood inundation model. A full description of each model is beyond the scope of this paper and the reader is referred to Pappenberger et al. (in press-a), Gouweleeuw et al. (2005), Hunter et al. (2005) and de Roo et al. (2000) for more details.
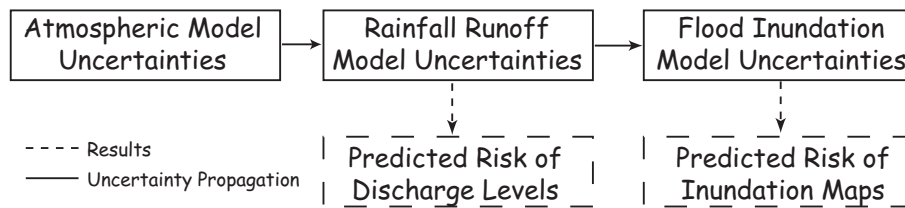
Fig. 1. *Sketch of the uncertainty cascade*

Assessing sources of uncertainty

ASSESSING UNCERTAINTY IN THE RAINFALL
FORECASTS

The rainfall forecasts used in this study are based on the European Centre for Medium Range Weather Forecasting (ECMWF) control and ensemble forecasts (known as the Ensemble Prediction System or EPS). These provide a single deterministic forecast and 50 additional realisations for the next 10 days with a grid scale of 80 km. In addition, a deterministic forecast is provided, which is the control forecast (unperturbed) at higher resolution (40 km) and thus should, in principle, provide better precipitation predictions. A detailed description of the ECMWF Numerical Weather Prediction model can be found in a number of references (Buizza *et al*., 2001; Molteni *et al*., 1996; Persson, 2001) and Gouweleeuw *et al*. (2005).

Rainfall forecasts are available for each grid square (approximately 40 km by 40 km) at one hour time steps over the 10-day period. The forecasts are updated daily with individual realisations based on perturbations of the initial conditions at the start of the period. Furthermore, the EPS includes a simulation of random model error due to parameterised physical processes (so-called stochastic physics, see Buizza *et al*., 1999). Results (Mullen and Buizza, 2001) have shown that both stochastic physics and higher resolution have a measurable positive impact on precipitation prediction. This is, therefore, a limited assessment of the potential uncertainties in future rainfalls. In the near future, with growing computer power, it is expected that the grid resolution will be increased, that the forecasts will be updated more frequently and that more ensembles will be run.

It is not possible to say whether one ensemble forecast is more likely to occur than another. They are therefore taken as equally likely. The possibility of conditioning N day ahead forecasts on the observed rainfalls after 1, 2 or more days, relative to the forecasts of reinitialised forecasts as new data become available, has not been explored in this study since, operationally, the NWP model itself is initialised every 12 hours as new data are received. Moreover, the question of whether the forecast ensembles represent a true sample of rainfall probabilities is not discussed.

ASSESSING UNCERTAINTY IN THE RUNOFF
MODEL

The runoff generation model used here is the LisFlood model (de Roo, 2000; de Roo *et al*., 2000, 2001). This is a distributed model with a water balance version (LisFlood-WB) that is, in this case, run on a daily basis and can then be used to provide initial storage conditions for a flood prediction version (LisFlood-FF) that is run for a 1 km grid for selected basins with a 1h time step. LisFlood-FF can be run using either measured or forecast rainfalls, interpolated to the 1 km grid.

LisFlood-FF uses representations of surface and subsurface flow processes that are lumped at the 1 km grid scale. Thus, although initial estimates of soil hydraulic characteristics and other parameters can be obtained from maps of soil and vegetation type, the process representations must really be considered as sub-grid parameterisations of the complex, heterogeneous and nonstationary nature of the local flow pathways. Thus some calibration of the effective values of the parameters will generally be necessary. Here, this has been achieved by comparing the predictions of the model for a historical period when observed rainfalls are available, prior to the use of the model for forecasting.

The approach to model calibration is based on the General Likelihood Uncertainty Estimation (GLUE) concepts (Beven and Binley, 1992). This approach rejects the idea of an optimal model representation in favour of an equifinality concept that recognises that many different parameter sets might produce acceptable or *behavioural* simulations (Beven, 1993). Predictions made with all the behavioural simulations then allow an evaluation of the output uncertainty arising from parameter and model structure uncertainty. The different behavioural simulations can be associated with a relative likelihood according to how well that parameter set has performed during a calibration or conditioning period. The likelihoods can be updated as new data become available, which may also lead to the rejection of more models as being non-behavioural if they do not

Table 1. Parameters, which have been varied by a multiplier for the Monte Carlo analysis of the LisFlood model (adopted from Pappenberger *et al.*, in press-a)

Parameter	Initial data source	Multiplier (uniform dist)
Saturated hydraulic conductivity (upper soil layer, Ks(S1))	HYPRES database	0.1–10
Saturated hydraulic conductivity (lower soil layer, Ks(S2))	HYPRES database	0.1–10
Porosity (upper soil layer, Theta(S1))	Soil cover / Corine data base	0.1–1.2
Porosity (lower soil layer, Theta(S2))	Soil cover / Corine data base	0.1–1.2
Saturated hydraulic conductivity (upper groundwater layer, Ks(G1))	HYPRES database	0.1–10
Saturated hydraulic conductivity (lower groundwater layer, Ks(G2))	HYPRES database	0.1–10
Manning Roughness Channel (NC)	Channel Order	0.1–5
Manning Roughness Overlandflow (NO)	Land use / Corine data base	0.1–10
Maximum Leave Area Index (ML)	AVHRR satellite images	0.1–10
Maximum percolation from upper to lower response box (P)	Expert Analysis	0.1–10

provide acceptable simulations of the new data (Freer *et al.*, 2003).

Such behavioural simulations are generally identified by the evaluation of different Monte Carlo realisations of parameter sets. A rigorous Monte Carlo study of LisFlood, however, as a distributed model with a large number of grid elements and a large number of parameters required in each grid element, would demand very great computer resources. Therefore, again, some approximations will be required. Here, a selection of the more sensitive parameters was chosen for use in the analysis based on previous experience with the model. In addition, spatially variable parameters (such as hydraulic conductivity of the surface soil layer) were modified only by applying a multiplier for that parameter to the whole field. In this way the dimensionality of the calibration problem was reduced significantly. Table 1 gives a summary of the parameters that were varied in the Monte Carlo sampling. Each model realisation was evaluated against observed discharges measured at the upstream boundary of the nested flood plain inundation model. At this stage, an attempt has been made to condition parameter sets for individual subcatchments for which observed discharges are available. There is, in principle, no problem in extending the analysis in this way, only increased computational requirements.

ASSESSING UNCERTAINTY IN THE HYDRAULIC MODEL

A similar approach was taken to assessing uncertainty in the hydraulic model, LisFlood-FP (Bates and De Roo, 2000). This is a simplified routing model that combines a one-dimensional kinematic representation of the main channel flow with a grid square representation of the flood plain over which two-dimensional flow is simulated using a storage cell concept. Typically, model resolution is of the order of 10–100 m with time steps of the order of 1s. With grids of between 10^4 and 10^5 cells, typical computation times on a desktop PC for full dynamic flood events of ~100 hours duration are ~5 minutes or less. The approach has been shown to be capable of predicting maximum flood extent in areas of complex, low-lying topography and to compare favourably with alternative hydraulic model formulations (Bates and de Roo, 2000; Horritt and Bates, 2003). Boundary condition data required by the model consist solely of the upstream inflow hydrograph at the head of the reach being considered. Other input data are a high resolution, high accuracy Digital Elevation Model of the floodplain topography (typically derived from aerial stereo-photogrammetry or laser altimetry) and basic information on the channel geometry (width, depth and slope). Lastly, the model requires the specification of patterns of effective channel and flood plain roughness parameters which in theory can be specified separately for each computational cell.

Again, some simplification of this high-dimensional calibration problem is required, particularly since it is known that simulations using storage cell codes are relatively insensitive to flood plain roughness, given the approximate representation of the geometry and flood plain infrastructure (Aronica *et al.*, 2002; Romanowicz *et al.*, 1996). As in previous studies, it was here assumed that single effective values of roughness for channel sub-reaches and flood plain elements respectively would be sufficient to obtain behavioural simulations. For the simulations reported in this paper, the roughness parameters were varied by increments across a physically feasible range and the simulations assessed against observed patterns of flood inundation

derived from aerial photography and satellite Synthetic Aperture Radar imagery.

All uncertainties which have been identified previously should be considered within a cascading system. The next section discusses problems with such an approach and suggests a possible methodology.

Cascading uncertainty through the forecasting system — using functional similarity to reduce computational demands

Within the system described above, each time a forecast is made there are 52 rainfall forecasts available (50 ensembles, one high resolution deterministic forecast and one control forecast at the coarser ensemble resolution — see Gouweleeuw *et al*., 2005). Each rainfall forecast should be run through all the behavioural realisations of the runoff generation model to provide upstream discharge inputs to all the behavioural flood wave routing models. Uncertainty in the forecasts should then be assessed over all combinations of rainfall inputs, runoff predictions and flood routing models. It can be readily seen that even for small samples of behavioural simulations, the total number of simulations required would multiply rapidly. Assuming that the rainfall forecasts can be provided on a continuous basis by ECMWF, this may be computationally feasible given, say, a large parallel PC system. However, using the type of distributed models employed in this study, such a task demands very significant computational resources even when the analysis is run off-line (so that time is not critical) rather than in real time.

There are two ways of resolving this computational problem. One would be to simplify the runoff generation and flood wave routing models (for example using the type of transfer function models explored in Lees *et al*., 1994; Romanowicz and Beven, 1998). The other is to reduce the number of runs required. The latter approach has been taken here by using the concept of functional similarity of parameter sets.

The concept of functional similarity arises quite naturally from the idea of equifinality within the GLUE methodology. In any model calibration or evaluation period there will be many parameter sets that provide an acceptable fit to the observations. These will necessarily be somewhat similar in their functioning in the evaluation period, although it is still expected that they will differ in the way in which they achieve acceptability with respect to any particular performance measure and will produce somewhat different simulations when used in forecasting. However, to reduce

the computational burden, the assumption is made that particular parameter sets can be identified within which these differences are smaller than those between regions. Thus, the idea is to classify the behavioural simulations into different functional types, and use only representative parameter sets for each functional type in the forecasting.

To implement this procedure, the behavioural hydrographs predicted by the rainfall-runoff model at a given point on the river network (here the head of the reach for which hydraulic modelling is undertaken) have been classified by cluster analysis, in which hydrographs are compared to each other by a similarity measure. The more dissimilar two model hydrographs being compared are, the more likely it is that they will be assigned to different clusters. It is possible to force the classification of a certain number of clusters to meet computational constraints but this can result in misclassification. The classification method is described in more detail in Pappenberger and Beven (in press). Each class of behavioural hydrographs can then be represented in prediction by a single (or small number of) models, weighted by the sum of likelihood weights of all the behavioural parameter sets in that class.

This method was also used to define functional sets of parameters in the hydraulic model, using as conditioning information gauging station and air photo data from a further large flood event which occurred in 1993. For this event the channel was split into three sub-reaches delimited by internal gauging stations and the friction parameters for each sub-reach were allowed to vary freely. Since the inundation predictions are less sensitive to floodplain friction, the
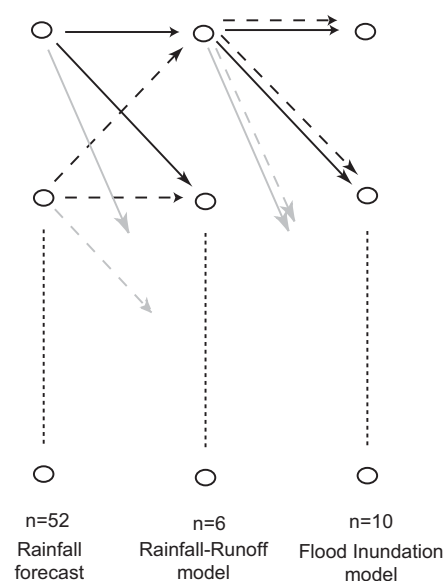

Fig. 2. Sketch illustrating the uncertainty propagation and the implementation

coefficient was fixed at n_{fp}=0.06. A Monte Carlo analysis using 500 simulations was conducted for the 1993 event with Manning's n friction parameters for each sub-reach drawn randomly and uniformly from the range 0.02 to 0.05. The ten best performing parameter sets from this analysis were then used for the simulation of the 1995 event.

The uncertainty cascade is summarised in Fig. 2. All 52 rainfall forecasts are used as inputs into the rainfall-runoff model, which is represented by a certain number of parameter sets (6 in this case). The likelihood weighted outflow hydrographs of the 52*6 realisations builds the upstream boundary condition for the inundation model, which is again represented by a certain number of parameter sets (10 in this case to minimise the computational burden). Thus a total of 3120 different inundation distributions are computed for each forecast. It is possible to construct predictive percentiles from these realisations.

Construction of prediction percentiles

The prediction percentiles for the rainfall-runoff model were constructed by using the likelihood measure associated with each behavioural parameter set since each rainfall-runoff forecast is assumed to be equally likely *'a priori'*. It is therefore not necessary to make strong prior assumptions about the likelihood of different parameter value combinations. In addition, it is also possible to compute prediction percentiles for flows over a certain threshold at any point in the catchment / flood plain system (Gouweleeuw

et al., 2005; Pappenberger *et al.*, in press-a) and maps of inundation probabilities. The information content of the latter, represented as a risk of inundation as the event progresses, is much easier to communicate to end users, including policy makers, emergency services and the public. For the predictions of flood inundation it is necessary to construct a prediction quantile for each model grid cell, which is based on an overall likelihood measure. This measure is computed by multiplying the performance of the rainfall–runoff part, as represented by the cumulative likelihood weights associated with each cluster of predicted discharge hydrographs, with the performance measure from the calibration of the flood inundation model

Application to the Meuse flood of 1995

The methodology for cascading uncertainty through the flood forecasting system was demonstrated by an application to the Meuse catchment, upstream of Maaseik (~21 000 km²) in Belgium. Discharge observations are available at the Borgharen gauging station, at the upstream boundary of the 35-km reach simulated by the LisFlood-FP flood inundation model.

Two different 10-day ahead forecast scenarios were available for the model evaluation, one starting on 21 January 1995 at 12:00 and the other one on 22 January 1995, also at 12:00. The second set has been chosen to evaluate possible updating techniques for the runoff model, which
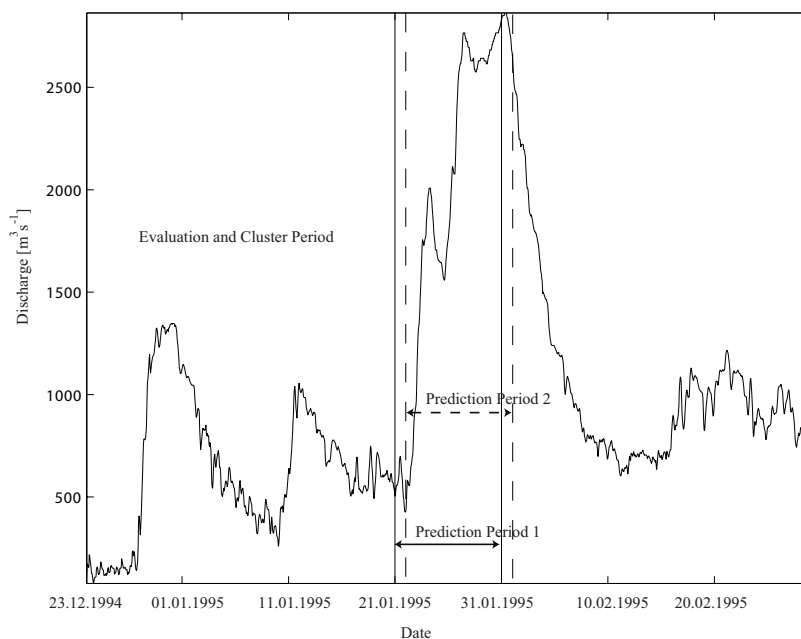

Fig. 3. *Measured outflow hydrograph at Borgharen which shows the period used for the evaluation and cluster process. Furthermore the prediction period of the ECWMF forecast is shown*

could be also extended to the inundation evaluation.

The rainfall–runoff model has been calibrated on hourly flow data, starting from 23 December 1994 right up to the starting time of the forecasts (see Fig. 3). The likelihood measure used is the Multicomponent Mapping method as described in Pappenberger and Beven (in press). This methodology allows the definition of errors around evaluation data as a measure designed in the form of a pyramidal frustum in time and magnitude around observations. The performance of hydrographs is computed according to combinations of the measure over all time steps. The behavioural parameter sets are then clustered into functional classes, which show a similar flow behaviour.

The flood plain model was pre-conditioned on observational data available for the December 1993 flood event; these consisted of a mosaic of air photo images of maximum inundation extent and two stage hydrographs internal to the model domain.

Results and Discussion

FORECASTS FROM 21 JANUARY

Analysis of the functional similarity of behavioural parameter sets (see previous discussion) led to the selection of six parameter sets to represent the parameter uncertainty of the rainfall-runoff model (the performance of these sets is shown in Table 2). The parameter sets have been clustered according to the hydrograph flow at Borgharen they produced in the evaluation period. These six sets resulted in 6 forecasts of the control, 6 forecasts of the high resolution rainfall forecast and 300 forecasts for the ensemble rainfall forecasts. The results of this initial step in the cascading process are shown in Fig. 4. The top sub-figure shows the 5 and 95% percentiles of predicted discharge at Borgharen based on the ensemble runs. The figures show an initialisation problem due to the fact that the ECMWF forecasts need a certain 'spin-up' time to balance internal

Table 2. Likelihood measures of the 6 parameter sets identified in the functional classification

Set	Likelihood Measure until 21January	Likelihood Measure until 22 January
1	0.8324	0.8333
2	0.8516	0.8507
3	0.8836	0.8857
4	0.8649	0.8683
5	0.8481	0.8470
6	0.8713	0.8730

state variables. Therefore, the results would have to be disregarded for a short lead time.

Later, the observations are bracketed by the uncertainty bounds and the model performs reasonably well. However, neither the timing nor the magnitude (especially the width of the uncertainty bounds) are acceptable. The study of Pappenberger *et al.* (in press-a) demonstrated the capability of the model to reproduce and predict within uncertainty bounds when 'true' rainfall is used rather than predicted. This suggests a deficiency in the rainfall forecasts, especially because the performance of the control run and the deterministic forecast drop off after 150 hours. Nevertheless, similar to the ensemble forecasts, the model cascade does show good predictions in the initial phase, which is reasonable bearing in mind higher uncertainties in longer range forecasts.

It is possible to illustrate the capability of this methodology to predict peak over threshold time predictive quantiles (see Fig. 5 for the example of flows over 1500 $m^3\,s^{-1}$). The measured peak over threshold is indicated by the vertical line. Naturally, the three plots show different responses and especially the analysis for the control and high resolution forecasts should be treated with the utmost care due to the small sample size used in calculating these plots. According to this analysis, one would have predicted a possibility of ~18% (for ensembles), ~33% (for control) and ~33% (for high resolution deterministic) to have 0 hours of flow above the threshold chosen. However, the measured data show that the 1500 $m^3\,s^{-1}$ threshold was actually exceeded for 118 h. The cumulative distribution functions give a possibility of ~63% (for ensembles), ~65% (for control) and ~65% (for deterministic) that the flow over 1500 $m^3\,s^{-1}$ would have been exceeded for less than 175 hours. This example illustrates the difference in the approach proposed in this paper in comparison to traditional model exercises. The political decision for further action can be informed by an estimate of the probability of making an error.

The error in the rainfall predictions also has a significant influence on the inundation predictions. The observed discharge at the time of the air photo manages (but only just) to bracket the observed discharge (see Fig. 6) reflecting the findings above. Bankfull discharge in the Borgharen-Maaseik reach is approximately 1500 $m^3\,s^{-1}$ and this would partially explain the limited inundation observed in the 5 and 50% quantile plots. However, given the difficulties of representing complex floodplain geometry (including small-scale dyke structures) within a relatively coarse model, the observed shoreline is reasonably well defined by the 95% quantile.

The spatial distribution of inundation uncertainties for the later, SAR-derived, shoreline (Fig. 7) is remarkably similar
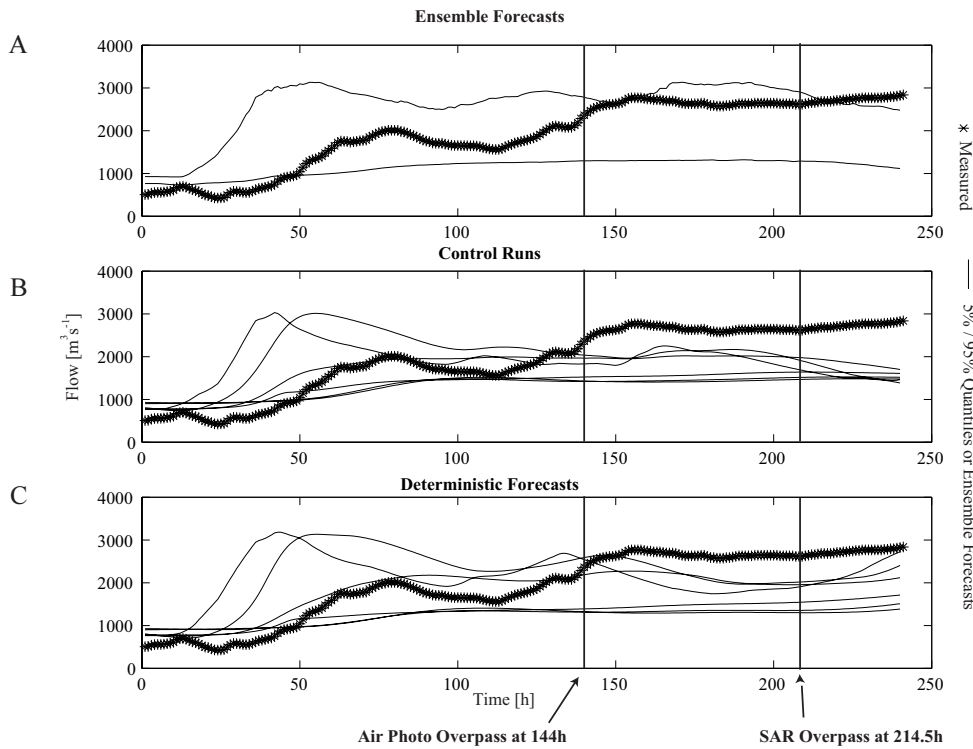
Fig. 4. *A. Prediction quantiles for discharge at Borgharen on the Meuse after the 21st of January based on 10 day ensemble forecasts of rainfalls. B. Predictions using 6 representative hydrological models and the control rainfall prediction. C. Predictions using 6 representative hydrological models and the deterministic rainfall forecast*
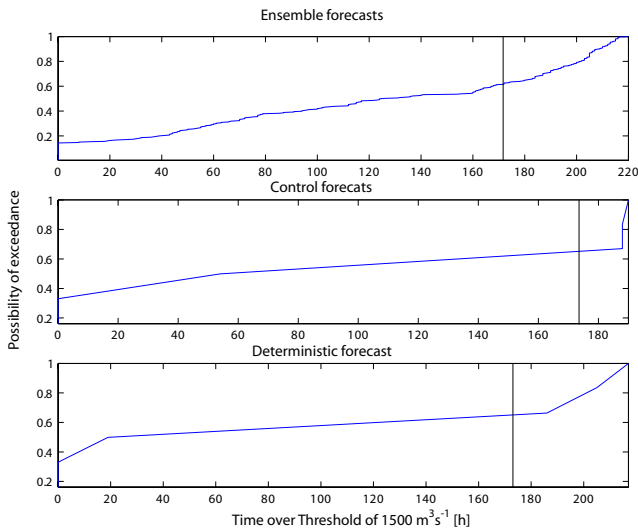

Fig. 5. *Possibilty of time over threshold of the bankfull threshold of 1500 $m^3 s^{-1}$ for the three different forecast scenarios (actual measured time of peak over threshold is 175h -see vertical line)*
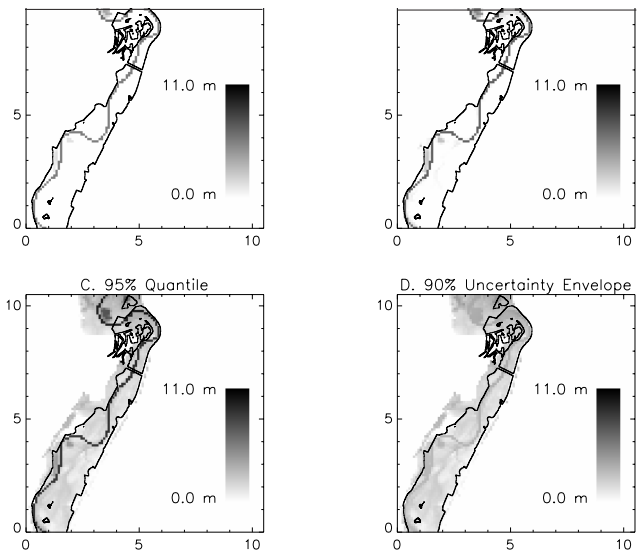
Fig. 6. *Predicted water levels (grey scale) at time of air photo with ensemble forecasts. The 5%, 50% and 95% quantiles are shown as well as the envelope between the 5 and 95% quantiles. The mapped inundation is indicated by a dark black line. (Forecasts from 21*

to those evaluated against the air photo data for the first sub-reach (which is reasonable given that the observations were collected under similar hydraulic conditions). The close proximity of a number of small lakes and ponds to the main

channel accounts for the more obvious inundation observed in the lower two sub-reaches in the 5 and 50% quantile plots. It is interesting (and unsurprising) to note that the uncertainty
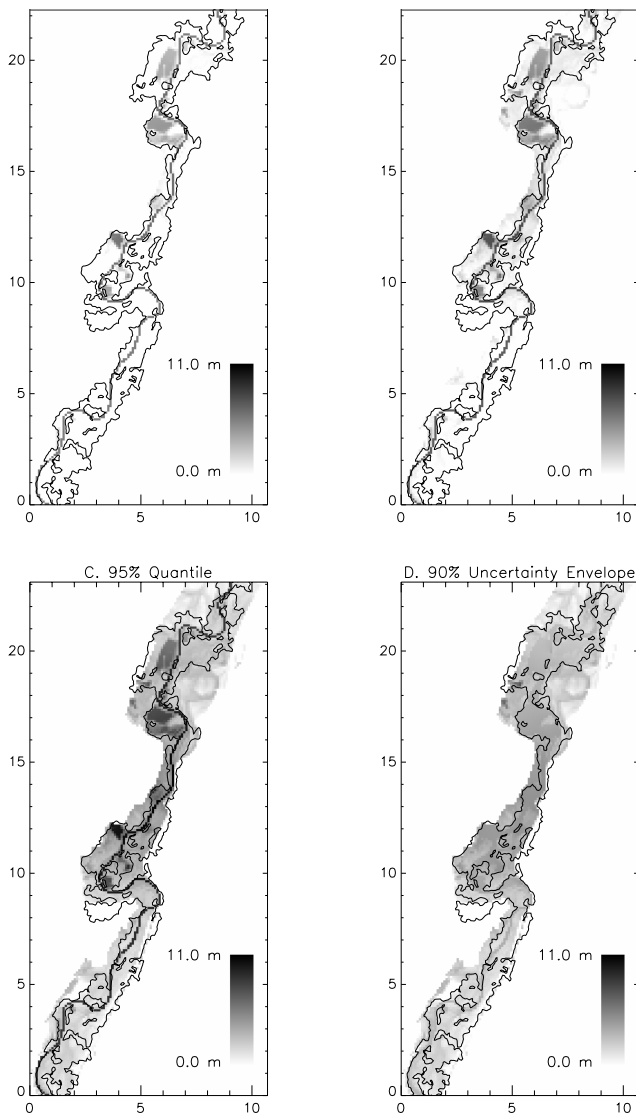
Fig. 8. *Predicted water levels (grey scale) at time of air photo with control forecasts. The 5%, 50% and 95% quantiles are shown as well as the envelope between the 5 and 95% quantiles. The mapped inundation is indicated by a dark black line. (Forecasts from 21 January)*

Fig. 7. *Predicted water levels (grey scale) at time of SAR overpass with ensemble forecasts. The 5%, 50% and 95% quantiles are shown as well as the envelope between the 5 and 95% quantiles. The mapped inundation is indicated by a dark black line. (Forecasts from 21 January)*

bounds of high predicted water levels are larger in regions of constricted flow, particularly in the highly engineered sub-reach 2.

If, as in traditional forecasts, only the control forecast is utilised (Fig. 8, time of air photo) a poor performance with respect to matching the observed shoreline can be observed, which is a direct result of the failure of the rainfall–runoff model to reproduce the observed discharge at Borgharen using the control forecast. Predicted discharges are approximately bankfull or less, which explains why only a very minor inundation was predicted on the floodplain. This lack of water may be compensated for in traditional
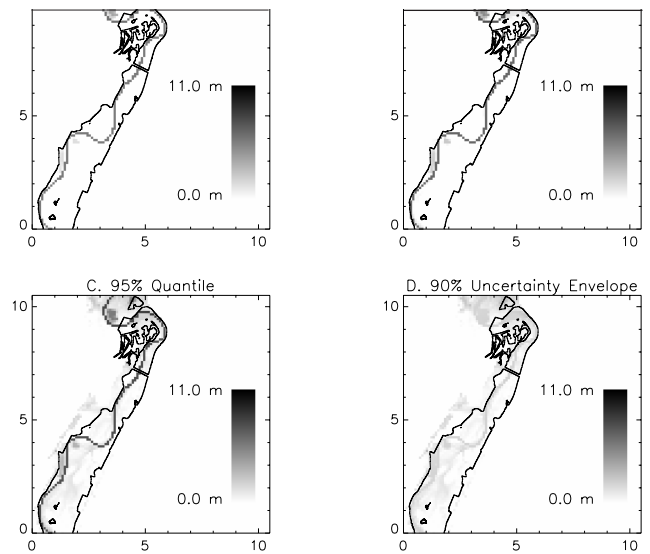
deterministic calibration through the use of large, effective friction coefficients, but not in an integrated system utilising behavioural parameter sets conditioned on a previous event or events. Furthermore, an erroneous result of this considerable underestimation of inundation in all predictions is that the difference between the 95 and 5% quantile plots (i.e. the 90% uncertainty envelope) appears very small in many regions, indicating a more certain but, in this case, incorrect set of predictions. Without an appreciation of all components of the forecasting system, this highlights one stage of the process where considerable opportunity for serious misinterpretation exists.

The high resolution deterministic forecasts (Fig. 9, time of SAR overpass) do not lead to any significant improvement compared to the control or ensemble forecasts and follow the discussion above. This analysis highlights that there is a characteristic division between the six control and deterministic hydrographs in two groups of three. One of these groups predicts the discharge at approximately bankfull or less (for the inundation model) which explains the low inundation of the 5% and 50% percentile. The second group is much closer to the observations and thus the upper limit (95%) leads to more reasonable predictions. This can be explained by two factors. Firstly, as mentioned above, extreme events are much less likely to be predicted correctly. Thus it has to be expected that the upper limit matches the measured disaster shore line. Secondly, the grouping of the hydrographs is based on the much lower peak flows of a previous period; thus it cannot be expected
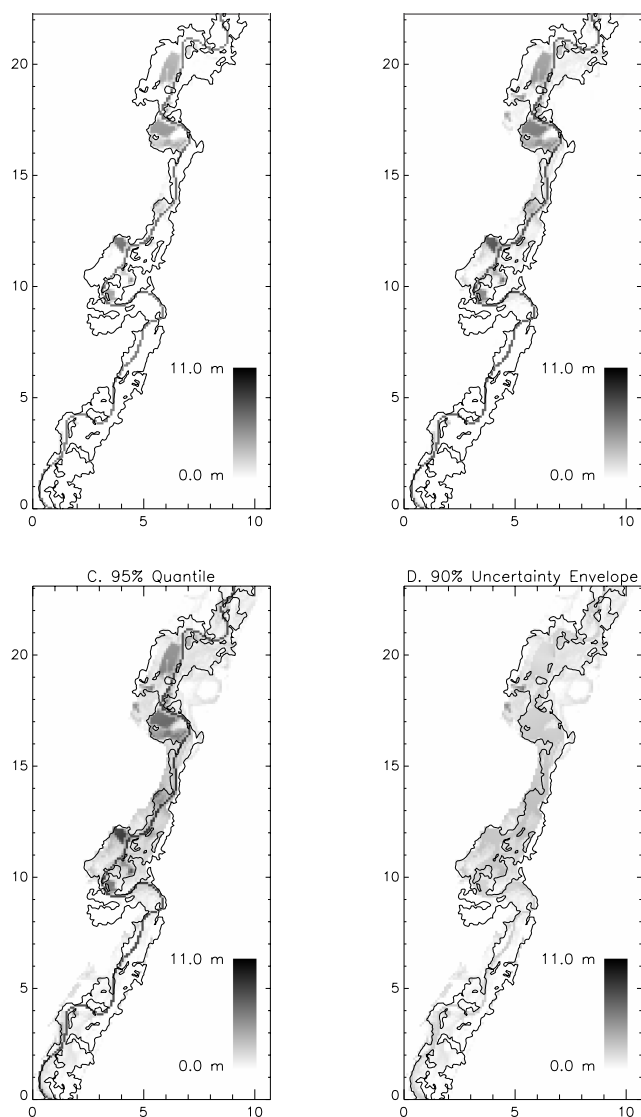
389

Fig. 9. *Predicted water levels (grey scale) at time of SAR overpass with deterministic forecasts. The 5%, 50% and 95% quantiles are shown as well as the envelope between the 5 and 95% quantiles. The mapped inundation is indicated by a dark black line. (Forecasts from 21 January)*

that extrapolated data match with similar performance due to the non-linearity of the response surface. Moreover, effective parameter values might change with event magnitude for good physical reasons (Romanowicz and Beven, 2003).

FORECASTS FROM 22 JANUARY

Figure 10 shows the predictions from 22 January onward; these use recorded rainfall before the updated forecast starts. It is possible to recompute the likelihood weights and clusters of behavioural models at each time that new data

become available, as a form of updating of the forecasting procedure. In this case, the clusters which represent the behavioural parameter sets used for this analysis did not change due to only small changes in the flow within the 24-hour period, although small changes in the likelihood measures can be observed (see Table 2). The predictions suffer from the same difficulties as on the day before. However, the pattern of the flow uncertainty bounds seems to have changed and the model under-performs in the first 75 hours. Nevertheless, the uncertainty bounds of the model still suggest a flood event in the near future.

It is interesting to note that, despite the failure of rainfall-runoff predictions to bracket the observed discharge with the ensemble forecasts, the spatial arrangement for the SAR derived shoreline of inundation uncertainties (Fig. 11) varies very little from the initial forecast period. The general narrowing of the 90% uncertainty envelope, again particularly evident in sub-reach 2, can most likely be attributed to the erroneously lower discharges routed through the rainfall-runoff model.

Moreover, the predictions with the control forecasts at the time of the air photo, mark an improvement in the ability of the rainfall-runoff model to reproduce the observed discharge at Borgharen using the control rainfall-fields for the second forecast period. Similar observations can be made for the deterministic forecast.

Conclusion

In real time flood inundation prediction, it can be useful to include medium range rainfall forecasts in the modelling system to achieve longer lead times for decisions on flood alert status. However, such forecasts, and the prediction of rainfall–runoff and flood inundation models, are inevitably uncertain. Therefore, strategies to cascade uncertainties through such a model framework have to be developed. Here, two 10-day ahead rainfall forecasts in the form of 50 ensemble, 1 control and 1 deterministic forecast (Gouweleeuw *et al*., 2005) illustrate a possible methodology. No prior likelihood can be assigned to these rainfall forecasts so that it is necessary to assume that all these predictions occur with the same probability. These forecasts have been used as an input for the rainfall-runoff model LisFlood for the 1995 flood event on the river Meuse. The parameter uncertainty of the rainfall-runoff model has been determined within a GLUE analysis (Pappenberger *et al*., in press-a). To minimise the computational burden, the responses of the Monte Carlo analysis of the model were classified into functional classes utilising the Multicomponent Mapping method (Pappenberger and Beven, in press). Likelihood measures could be computed for the six parameter sets
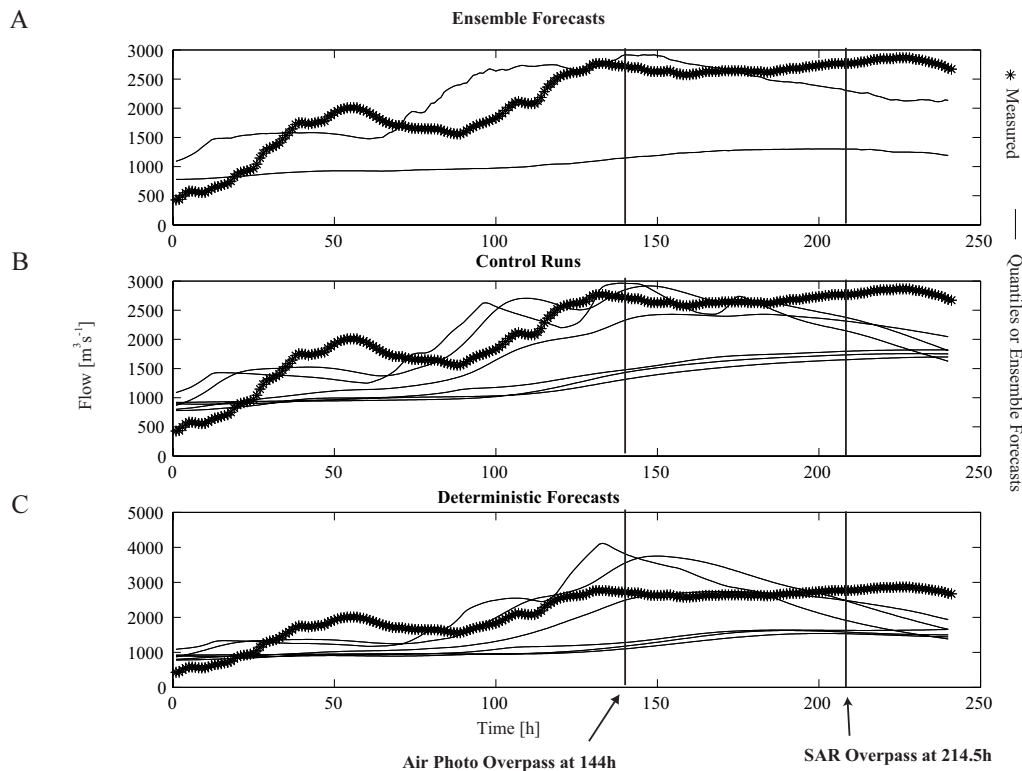
Fig.10. *A. Prediction quantiles for discharge at Borgharen on the Meuse after the 22nd of January based on 10 day ensemble forecasts of rainfalls. B. Predictions using 6 representative hydrological models and the control rainfall prediction. C. Predictions using 6 representative hydrological models and the deterministic rainfall forecast*

stemming from this analysis on the basis of all the behavioural parameters that they represent. The analysis suffers from under-performing rainfall predictions and therefore the value of the predictions is lessened. In this analysis, neither the control nor the deterministic forecasts could improve the model results significantly in comparison to the ensemble forecasts. However, the model predictions are capable of bracketing the real flow over a large range and only initialisation problems occur.

The quality of the input predictions cascades with the inundation model, which has been calibrated on a SAR and an air photo image of flood inundation. It has been demonstrated that the inundation model can give reasonable predictions within the limitations of the input. The predicted inundation percentiles provide further evidence that the forecasting of extreme events is a difficult business, as only the upper boundary is actually in good agreement with the observations.

Admittedly, the significance and dominance of the rainfall prediction error is well known and documented (see e.g. Anderson *et al*., 2002; Koussis *et al*., 2003) and therefore it is not a new finding. Moreover, further research has to be conducted on the influence of the spatial rainfall forecast

variability on the cascading uncertainty system (Arnaud *et al*., 2002).

However, this paper concentrates not on this limitation but rather introduces a method to deal with uncertainties in a cascading model system. Advances in the predictions of rainfall as well as a better understanding of the uncertainties in the rainfall-runoff and inundation components will hopefully narrow the prediction quantiles in the future. However, as demonstrated in this paper, promising results may be expected from such further progress.
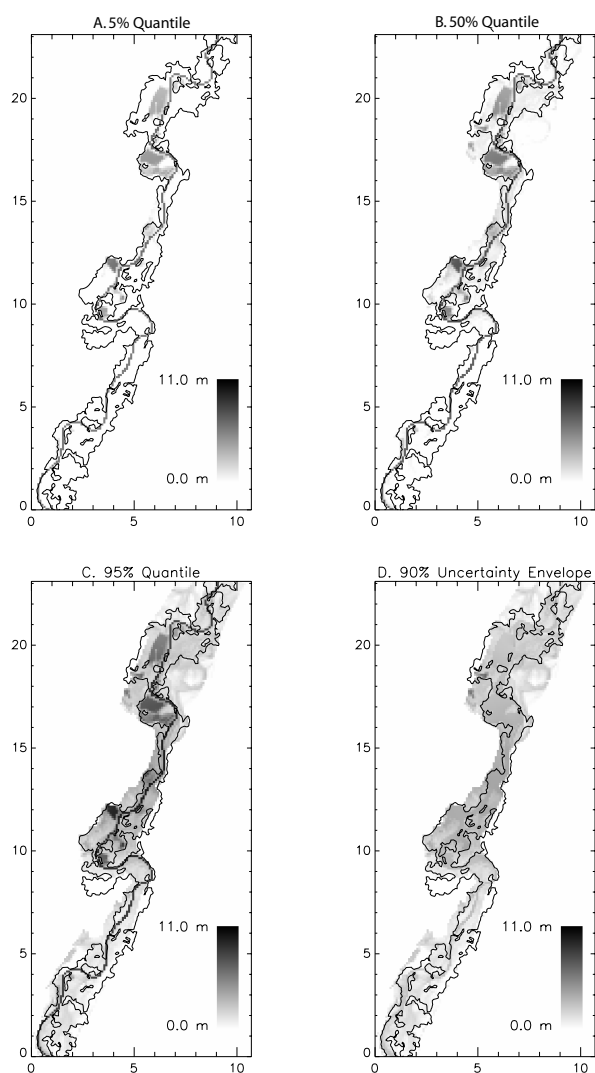
Acknowledgements

Fig. 11. *Predicted water levels (grey scale) at time of SAR overpass with ensemble forecasts. The 5%, 50% and 95% quantiles are shown as well as the envelope between the 5 and 95% quantiles. The mapped inundation is indicated by a dark black line. (Forecasts from 22 January)*

References

Anderson, M.L., Chen, Z.Q., Kavvas, M.L. and Feldman, A., 2002. Coupling HEC-HMS with atmospheric models for prediction of watershed runoff. *J. Hydraul. Eng.*, **7**, 312–318.

Arnaud, P., Bouvier, C., Cisneros, L. and Dominguez, R., 2002. Influence of rainfall spatial variability on flood prediction. *J. Hydrol.*, **260**, 216–230.

Aronica, G., Bates, P.D. and Horritt, M.S., 2002. Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. *Hydrol. Process.*, **16**, 2001–2016.

Aronica, G., Hankin, B. and Beven, K.J., 1998. Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data. *Adv. Water Resour.*, **22**, 349–365.

Bates, P.D. and De Roo, A.P.J., 2000. A simple raster-based model for flood inundation simulation. *J. Hydrol.*, **236**, 54–77.

Beven, K.J., 1989. Changing Ideas in Hydrology: The Case of Physically Based Models. *J. Hydrol.*, **105**, 157–172.

Beven, K.J., 1993. Prophecy, Reality and Uncertainty in Distributed Hydrological Modeling. Adv. *Water Resour.*, **16**, 41–51.

Beven, K.J., 2001. On hypothesis testing in hydrology. *Hydrol. Process.*, **15**, 1655–1657.

Beven, K.J., in press. A Manifesto for the equifinality thesis. *J. Hydrol.*

Beven, K.J. and Binley, A., 1992. The Future of Distributed Models: Model Calibration and Uncertainty Prediction. *Hydrol. Process.*, **6**, 279–298.

Buizza, R., Miller, M. and Palmer, T.N., 1999. Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteorol. Soc.*, **125**, 2887–2908.

Buizza, R., Richardson, D.S. and Palmer, T.N., 2001. The new 80-km high-resolution ECMWF EPS. *ECMWF Newsletter, 90*, 2–9.

de Roo, A., Gouweleeuw, B., Thielen, J., Bartholmes, J., Bongioannini-Cerlini, P., Todini, E., Bates, P., Horritt, M., Hunter, N., Beven, K.J., Pappenberger, F., Heise, E., Rivin, G., Hills, M., Hollingsworth, A., Holst, B., Kwadijk, J., Reggiani, P., van Dijk, M., Sattler, K. and Sprokkereef, E., 2003. Development of a European Flood Forecasting System. *Int. J. River Basin Manage.*, **1**, 49–59.

de Roo, A.P.J., 2000. LisFlood: A rainfall-runoff model for large river basins to assess the influence of land use changes on flood risk.http://natural-hazards.jrc.it/documents/floods/1998-docs/1998-papers/98Giessen.pdf,

de Roo, A.P.J., Odijk, M., Schmuck, G., Koster, E. and Lucieer, A., 2001. Assessing the effects of land use changes on floods in the Meuse and Oder catchment. *Phys. Chem. Earth* (B), **26**, 593–599.

de Roo, A.P.J., Wesseling, C.G. and Van Deursen, W.P.A., 2000. Physically based river basin modelling within a GIS: the LISFLOOD model. *Hydrol. Process.*, **14**, 1981–1992.

Downtown, R. and Bell, R., 1988. The impact of analysis differences on a medium range forecast. *Meteorol. Mag.*, **117**, 279–285.

Freer, J., Beven, K.J. and Peters, N., 2003. Multivariate seasonal period model rejection within the generalised likelihood uncertainty estimation procedure. In: *Calibration of watershed models,* Q.Y. Duan, H. Gupta, S. Sorooshian, A. Rousseau and R. Turcotte (Eds.), American Geophysical Union, Washington, USA. 69–88.

Gouweleeuw, B.T., Thielen, J., de Roo, A.P.J. and Buizza, R., 2005. Flood forecasting using probabilistic weather predictions. *Hydrol Earth Syst Sci.,* **9**, 87–102.

Gupta, H., Sorooshian, S., Hogue, T.S. and Boyle, D.P., 2003. Advances in automatic calibration of watershed models. In: *Calibration of watershed models,* Q.Y. Duan, H. Gupta, S. Sorooshian, A. Rousseau and R. Turcotte (Eds.), American Geophysical Union, Washington, USA.

Harrison, M., Palmer, T.N., Richardson, D.S. and Buizza, R., 1999. Analysis and model dependencies in medium-range ensembles: two transplant case studies. *Quart. J. Roy. Meteorol. Soc.*, **125**, 2487–2515.

Hewitson, B.C. and Crane, R.G., 1996. Climate downscaling: techniques and application. *Climatic Res.*, **7**, 85–95.

Horritt, M.S. and Bates, P.D., 2003. Evaluation of 1D and 2D numerical models for predicting river flood inundation. *J. Hydrol.*, **268**, 87–99.

Hunter, N.M., Bates, P.D., Horritt, M.S., De Roo, A.P.J. and Werner, M.G.F., 2005. Utility of different data types for conditioning flood inundation models within a GLUE framework. *Hydrol Earth Syst Sci.,* **9**, 134–152.

Joubert, A.M. and Hewitson, B.C., 1997. Simulating present and future climates in southern Africa using general circulation models. *Prog. Phys.Geogr.*, **21**, 51–78.

Koren, V.I., Finnerty, B.D., Shaake, J.C., Smith, M.B., Seo, D.-J. and Duan, Q.Y., 1999. Scale dependencies of hydrologic models to spatial variability of precipitation. *J. Hydrol.*, **217**, 285–302.

Koussis, A.D., Lagouvardos, K., Mazi, K., Kotroni, V., Sitzmann, D., Lang, J.G., Zaiss, H., Buzzi, A. and Malguzzi, P., 2003. Flood forecasts for urban basin with integrated hydro-meteorological model. *J. Hydrol. Eng.*, **8**, 1–11.

Krzysztofowicz, R., 2002. Bayesian system for probabilistic river stage forecasting. *J. Hydrol.*, **268**, 16–40.

Kyriakidis, P.C., Miller, N.L. and Kim, J., 2001. Uncertainty propagation of regional climate model precipitation forecasts to hydrologic impact assessment. *J. Hydrometeorol.*, **2**, 140–160.

Lees, M., Young, P.C., Ferguson, S., Beven, K.J. and Burns, J., 1994. An adaptive flood warning scheme for the River Nith at Dumfries. In: *River Flood Hydraulics* W.R. White and J. Watts (Eds.), Wiley, Chichester, UK. 65–75.

Marks, K. and Bates, P., 2000. Integration of high-resolution topographic data with floodplain flow models. *Hydrolog. Process.*, **14**, 2109–2122.

Molteni, F., Buizza, R., Palmer, T.N. and Petroliagis, T., 1996. The ECMWF Ensemble Prediction System: methodology and validation. *Quart. J. Roy. Meteorol. Soc.*, **122**, 73–119.

Moore, R.J., Jones, D.A., Bird, P.B. and Cottingham, M.C., 1990. A basin-wide flow forecasting system for real-time flood warning, river control and water management. In: *River Flood Hydraulics* W.R. White and J. Watts (Eds.), Wiley, Chichester, UK.

Mullen, S.L. and Buizza, R., 2001. The Impact of horizontal resolution and ensemble size on probabilistic forecasts of precipitation by the ECMWF Ensemble Prediction System. *Mon. Weather Rev.*, **129**, 638–663.

Pappenberger, F. and Beven, K., in press. Functional Classification and Evaluation of Hydrographs based on Multicomponent Mapping. *Int. J. River Basin Manage.*

Pappenberger, F., Beven, K.J., de Roo, A., Thielen, J. and Gouweleeuw, B., in press-a. Uncertainty analysis of the rainfall runoff model LisFlood within the Generalized Likelihood Uncertainty Estimation (GLUE). *Int. J. River Basin Manage.*

Pappenberger, F., Beven, K.J., Horritt, M. and Blazkova, S., in press-b. Uncertainty in the calibration of effective roughness parameters in HEC-RAS using inundation and downstream level observations. *J. Hydrol.*

Persson, A., 2001. User Guide to ECMWF forecast products, ECMWF Meteorological Bulletin M3.2. http://www.ecmwf.int/products/forecasts/user_guide.pdf

Romanowicz, R. and Beven, K.J., 1998. Dynamic real-time prediction of flood inundation probabilities. *Hydrolog. Sci. J.*, **43**, 181–196.

Romanowicz, R. and Beven, K.J., 2003. Estimation of flood inundation probabilities as conditioned on event inundation maps. *Water Resour. Res.*, **39**, art. no.-1073.

Romanowicz, R., Beven, K.J. and Tawn, J.A., 1994. Evaluation of predictive uncertainty in nonlinear hydraulic models using a Bayesian Approach. In: *Statistics for the Environment 2, Water Related Issues*, V. Barnett and K.F. Turkman (Eds.), Wiley, New York, USA.

Romanowicz, R., Beven, K.J. and Tawn, J., 1996. Bayesian calibration of flood inundation models. In: *Floodplain Processes*, M.G. Anderson, D.E. Walling and P.D. Bates (Eds.), Wiley, New York, USA.

Young, P., 2002. Advances in real-time flood forecasting. *Phil. Trans.: Math., Phys. Eng. Sci.*, **360**, 1433–1450.

Young, P.C. and Beven, K.J., 1994. Data-Based Mechanistic Modeling and the Rainfall-Flow Nonlinearity. *Environmetrics*, **5**, 335–363.

393