THE UNCERTAINTY CASCADE IN FLOOD FORECASTING

Keith Beven (1), Renata Romanowicz (1), Florian Pappenberger (1), Peter Young (1) and Micha Werner (2)

(1) Environmental Science/Lancaster Environment Centre, Lancaster University, LA1 4YQ, UK
(2) WL|Delft Hydraulics, Hydrology and Flood Management, Rotterdamseweg 185, 2629 HD Delft, The Netherlands

Abstract
A methodology for propagating and constraining the uncertainty inherent in real-time flood forecasting is presented and demonstrated on an application to the River Severn, UK. The flood forecasting system is based on a cascade of rainfall-runoff and flood routing models, developed using stochastic transfer functions with state dependent parameterisations to allow for nonlinearity. The nonlinearities require a Monte Carlo sampling approach to propagation of uncertainty. Model updating and uncertainty constraint as new water level data become available is based on a Kalman filtering approach. The methodology is being implemented into the UK National Flood Forecasting System.

Keywords: flood forecasting, transfer functions, National Flood Forecasting System (NFFS), nonlinearity, uncertainty,

INTRODUCTION

There are many sources of error in making flood forecasts. Such errors mean that all forecasts must be considered uncertain, and that there is a real possibility of getting a forecast wrong, both by not issuing a warning and flood damages being incurred, or by issuing a warning and no flood damages being incurred. It has long been recognised that both types of error will have an effect on the public perception of and reaction to flood warnings (the “crying wolf” problem). Flood forecasting is therefore not just a scientific problem, it is a problem of <u>managing</u> and <u>communicating</u> uncertainty. This involves three critical issues: the representation of different types of uncertainties in the forecasting system; the (preferably optimal) constraint of uncertainty in forecasts by means of real-time data assimilation and updating; and the presentation of forecasts and their associated uncertainties to decision makers and public. This contribution addresses the first two of these problems.

The uncertainties in flood forecasting are manifold. There are spatial and temporal uncertainties in the inputs to the system; in the antecedent conditions of the system; in the geometry of the system (including relevant flood defence infrastructure); in the possibility of infrastructure failure; in the characteristics of the system (in the form of model parameters); and in the limitations of the models available to fully represent the surface and subsurface flow processes in flood generation and routing. The importance of the different types of uncertainties will certainly vary with the time (and lead time) of the forecasts, and with the magnitude of the event.

There are issues of how to represent these different types of uncertainties. We may have input scenarios based on ensemble predictions from rainfall nowcasting or multiple historical analogues (Buizza et al., 1999; Buizza et al., 2001; Molteni et al., 1996); stochastic uncertainties (based on a variety of assumptions) associated with the calibration of models of runoff generation processes (e.g. Romanowicz and Beven, 2005; Romanowicz et al., 2004); and fuzzy representation of uncertainty in flood inundation predictions (Pappenberger and Beven, 2004; Pappenberger et al., 2005a; Pappenberger et al., 2005b; Pappenberger et al., in press).

Water (generally) flows downhill. Thus, flow processes lead to a natural cascade of uncertainty through the system. Rainfall and antecedent condition (including available lying snow) uncertainties combined with model structure and parameter uncertainties will result in uncertainty in runoff generation volume and timing. This provides uncertain inputs to routing algorithms with their own model, parameter, geometry and infrastructure uncertainties. Through such a cascade, uncertainty will grow unless it can be constrained in some way by observations.

Some constraints are natural. The better the estimation of the input of water during an event, the better is the chance for accuracy in forecasts. Mass balance constraints mean that there is an upper limit on how much runoff can be generated, equal to the rainfall inputs plus any snowmelt. Unfortunately, even in extreme events, the percentage runoff may be significantly less than the amount of input, and differences in timing might produce significantly different peak discharges, so this may not be a strong constraint. Far more useful will be observations on discharges or water levels within the system that can be used to see how far forecasts are in error and incorporated into data assimilation algorithms to constrain the growth of uncertainty within the system.

This will be demonstrated in this paper in the context of the National Flood Forecasting System (NFFS) developed for the UK Environment Agency by WL|Delft Hydraulics in an application to the River Severn. The demonstration makes use of techniques of forecasting and data assimilation based on stochastic transfer functions developed by Romanowicz et al. (2005; 2004).

THE NATIONAL FLOOD FORECASTING SYSTEM

DELFT-FEWS/NFFS is a flood forecasting system framework with an architecture which allows different types of models to be linked, and coordinates the model outputs. The NFFS system, as implemented for the UK Environment Agency, consists of a core, which handles data management and display; a set of scripts which configure and call different models, and the models themselves. The script language is XML and all intermediate scripts require a forward and backward conversion of data. The scripts are aligned in a workflow, which determines the sequence of execution. For this paper, several Matlab® functions have been wrapped and compiled so that they can utilize XML scripts as input. These wrapping functions are generic and allow the usage of the uncertainty cascading methodology from any other software package or stand-alone programme, as shown in Figure 1.

Here, the NFFS system has been used to implement the flood forecasting models for the River Severn downstream to Buildwas (3717 km^2) developed by Romanowicz et al. (2004). An online screen from the NFFS system is shown in Figure 2. The models use stochastic transfer functions (STF) for both rainfall-runoff and flood routing components. They include data assimilation and uncertainty estimation routines as described below.

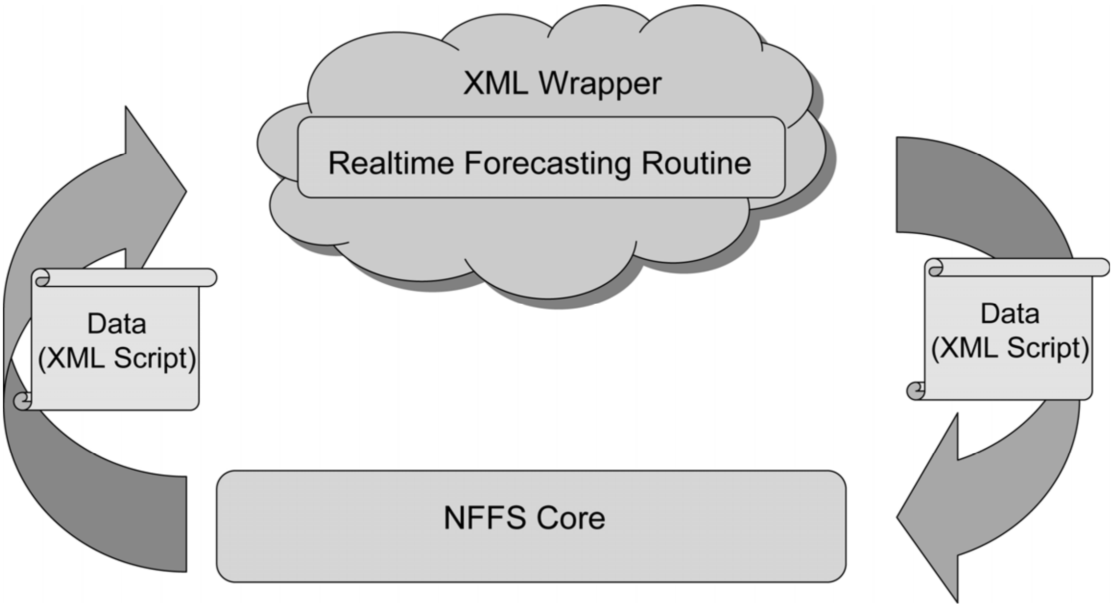

Figure 1 Real time data assimilation in NFFS using XML wrapping

The procedure for the derivation of the full, adaptive rainfall-water level model consists of three stages: (i) the derivation of the linear rainfall- water level TF model structure from the data; (ii) this model structure is used as the initial step in a recursive optimisation routine to find and appropriate nonlinear transformation of the rainfall data and corresponding TF model and finally (iii) an on-line updating procedure of the gain and variance of the TF model predictions uses a Kalman filter data assimilation procedure (Young, 2002). The full flow diagram of the system is shown in Figure 3. The STF models were developed using the autumn 1998 flood events (calibration stage) and autumn 2000 flood events were used for the validation of the models.

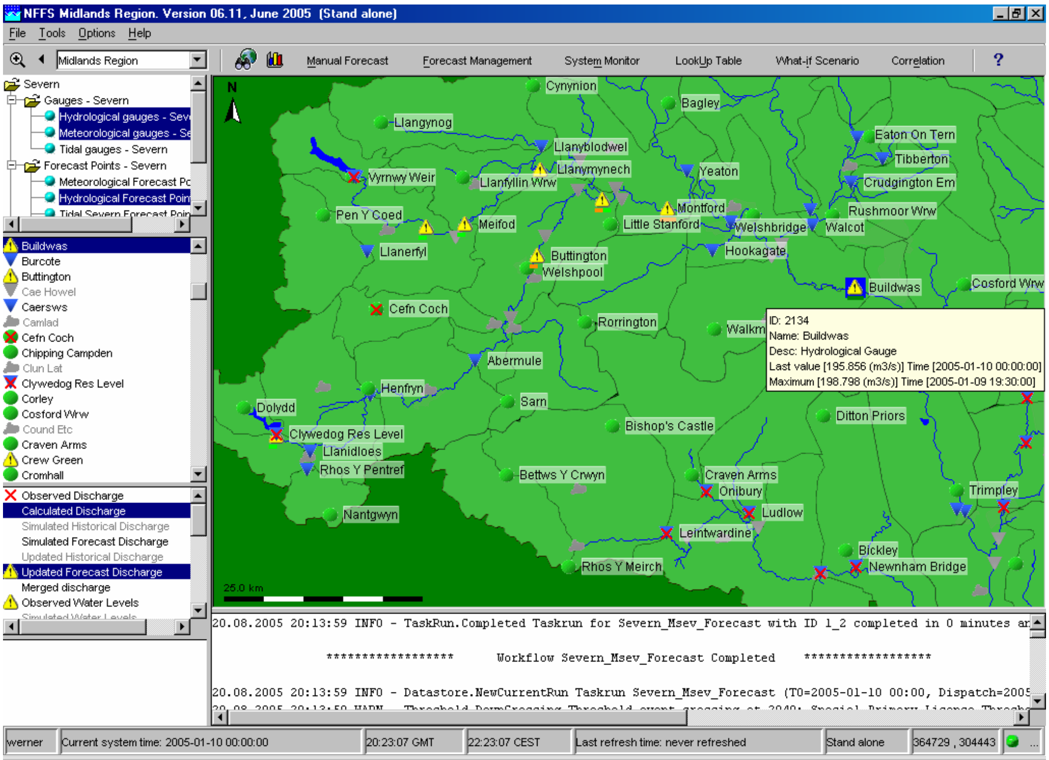

Figure 2 The on-line forecasting system for the River Severn, UK, upstream of Buildwas as presented in NFFS.

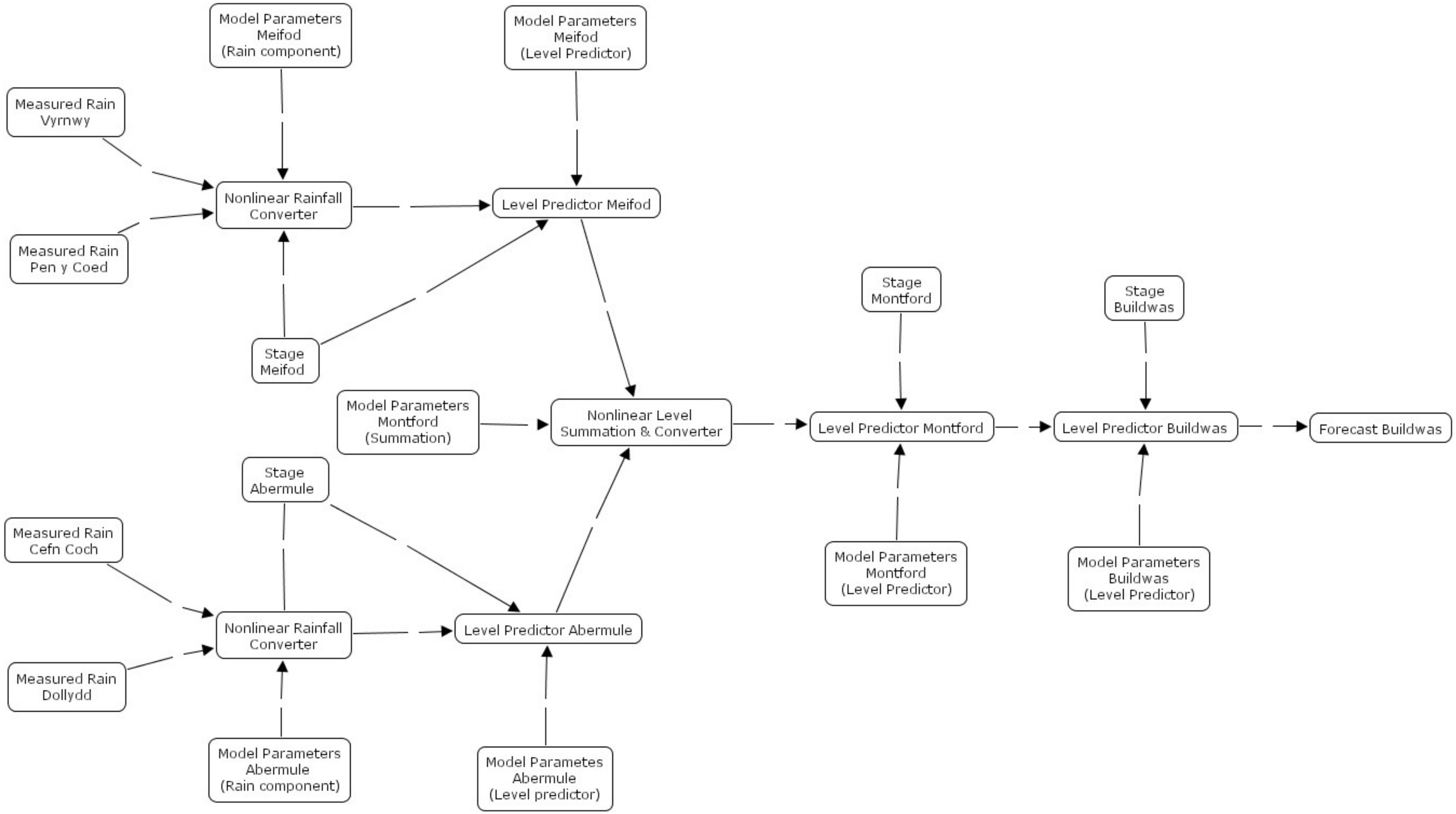

Figure 3 Flow diagram of the Severn forecasting system with data assimilation as implemented in NFFS

THE UNCERTAINTY CASCADE AND DATA ASSIMILATION

Each model component has its own source of uncertainties and therefore uncertainty cascades from rainfall measurements and forecasts as indicators of basin inputs, through the runoff generation prediction to the flood wave forecasts. At each stage of the cascade we are dealing with nonlinear 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. Uncertainties in nonlinear systems can often be estimated simply, however, 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 only achieved by some approximate methods.

There are two ways of resolving this computational problem. One is to reduce the number of runs required. This approach has been taken by Pappenberger et al. (2005b) by using the concept of functional similarity of parameter sets. Another 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, and Romanowicz and Beven, 1998). This approach has been taken in this example, together with numerical estimation of the uncertainty propagating through the series of models in the forecasting system, using MC ensembles of the input variables obtained from the distribution of the preceding model outputs. At each stage the predictions are conditioned on observations of actual water levels, thereby introducing conditioning of the uncertainty in the downstream predictions in an explicit way.

The on-line data assimilation procedure for each step uses a state space formulation of the STF models which is solved on-line by the Kalman filter engine. Additionally, on-line n-step ahead predictions of the water levels are adjusted using a nonlinear gain estimator at each time step on the basis of the new incoming observations. The heteroscedastic variance of the forecast is also updated on-line. Both gain and variance estimators use a recursive least squares random walk algorithm (Young, 1984). The data assimilation methods introduce a number of hyper-parameters to control the Kalman filter and random walk noise to variance ratios.

Owing to the on-line data assimilation procedure, the uncertainty of the predictions is constrained within much smaller bands than is the case without on-line updating. However, when compared with the uncertainty of the forecasts of a single forecasting module, the errors may increase as a result of the uncertainty cascading through the consecutive sequential modules of the forecasting system.

Constraining uncertainty in rainfall-runoff forecasting

The first two sub-models of the on-line forecasting system shown in Figure 3 are rainfall-water level models to estimate water levels at Abermule (580 km^2) and Meifod (675 km^2), Both these models have 5 hours delay, which means that natural lead time for the forecast cannot exceed 5 hours, unless a rainfall forecast is available. The STF model for Abermule has fast and slow TF components in parallel and, with recursively updated nonlinear gain and heteroscedastic variance, yields on average 94 % explanation of the variance of observations in calibration and 93.5% for the 2000 validation event.

The best rainfall-water level model for Meifod, has the same form as for Abermule, and explains on average 94% of the variance of the observations with on-line updating of the gain and variance for the calibration period (October 1998) and 93.8% of the data for the validation period (November 2000). Both rainfall-runoff models were set up in a Monte Carlo simulation framework in order to demonstrate the propagation of uncertainty related to the model structure parameters and the additional hyper–parameters required. One thousand model simulations were performed, with parameter values chosen from the estimated joint distributions, and the resulting uncertainty bands for the forecast median and confidence of variance are illustrated in Figures 4 and 5, for Abermule and Meifod gauging stations, for October, 2000 (validation stage).

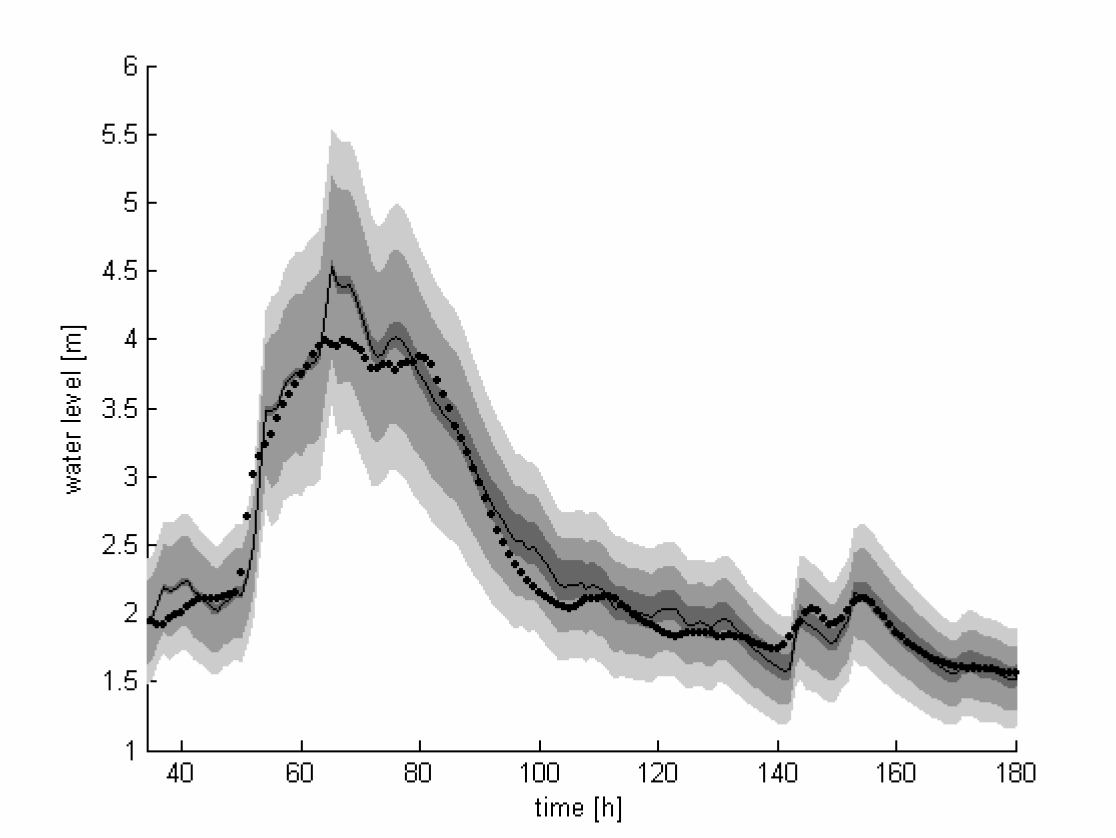

Figure 4 Five hours ahead forecast for Abermule, calibration stage, October 2000, denoted by a continuous line; dark shaded area denote the 0.95 uncertainty bands of the median of the forecast, lighter shaded area describes the 0.95 confidence limits of the forecast and the lightest shaded line shows 0.95 uncertainty on the forecast confidence limits; points denote the observations.

Constraining Uncertainty in Flood Routing

The next sub-model is used to derive the on-line forecasts of water levels at Montford (2025 km^2), above which the River Vyrnwy joins the Severn downstream of the Meiford gauging station. In order to take this tributary into account in our water level routing model, we optimised the weights for combining the stages at Abermule and Meifod during the calibration stage. The results of the optimisation indicated that both stages should be summed up with equal weights. The identified TF model for Montford is first order and has 11 hours delay. Together with the 5 hours ahead forecast from the rainfall-water level models we get 16 hours ahead forecasts at Montford. The last sub-model from Figure 3 provides water level forecasts at Buildwas. The Montford-Buildwas model was identified using October 1998 year floods as a second order TF model with 16 hours delay. Together with the 16 hours ahead forecasts obtained from the previous stages of forecasting, we can get up to a 32 hours ahead forecast at Buildwas.

The propagation of the uncertainty due to the application of forecasted water levels at Meifod and Abermule instead of the observed values in order to prolong the forecast lead time was studied using 1000 Monte Carlo simulations of the entire sequential system with varying model parameters and hyper-parameters according to the calibrated prior distributions. In order to take into account uncertainty of the Montford input, an additional 10 random samples were taken at random from the Abermule and Meifod trajectories varying within the estimated 0.95 confidence bands of their 5 hours ahead forecasts, defined by the outer light shaded areas in Figures 4 and 5. The Buildwas model simulations similarly used 10 randomly sampled trajectories from the Montford forecasts varying within the estimated 0.95 confidence of the 16 hours ahead Montford forecasts. The resulting confidence bands at Buildwas, based on 10,000 simulation runs in each case, for the forecast water levels are shown in Figure 6. The narrow dark grey bounds correspond to uncertainty arising only from the TF model parameter uncertainty. As these are well-defined models, such prediction limits are narrow. However, the forecast uncertainty comprises both observation and input uncertainty, which gives the bounds shown in the total shaded area in these figures. The Montford 16 hours ahead forecasts explained 92% of the variations of the output for 1998 year event and 91.1% of the water level variations for the 2000 year event. The Buildwas model explained on average 83% in calibration (1998 year) and 82.2% of the observation variance for the November 2000 validation period.

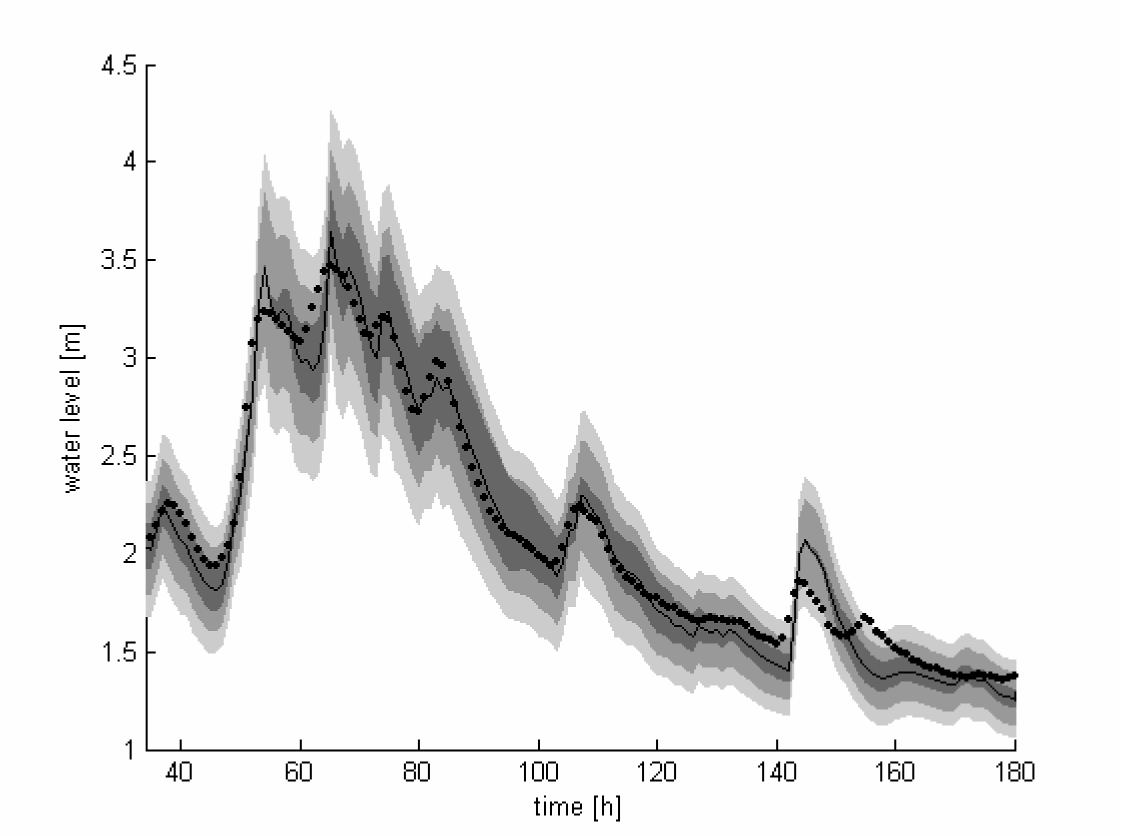

Figure 5 Five hours ahead forecast for Meifod, calibration stage, October 2000; predictions are denoted by a continuous line; dark shaded area denote the 0.95 uncertainty bands of the median of the forecast, lighter shaded area describes the 0.95 confidence limits of the forecast and the lightest shaded line shows the 0.95 uncertainty on the variance itself. Points denote the observations.

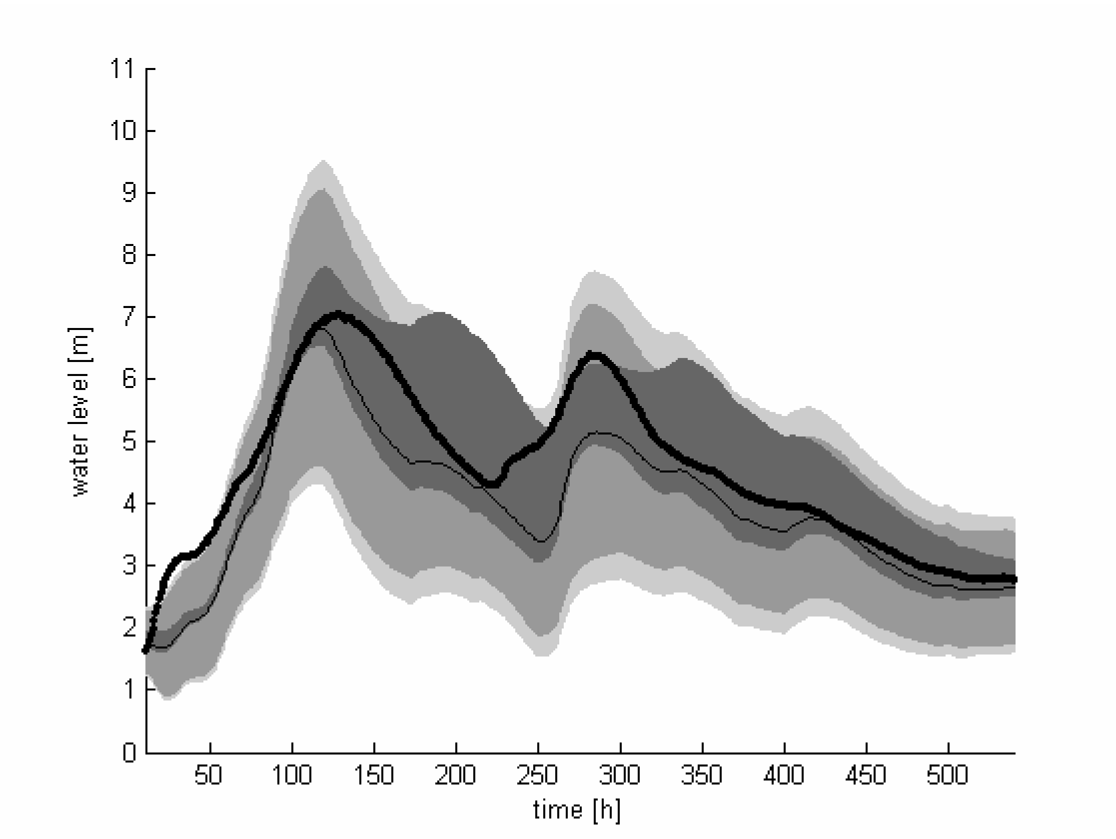

Figure 6 Uncertainty of 32 hours ahead forecast for Buildwas, November 2000, estimated from 1000 Monte Carlo simulations. Median predictions are denoted by a continuous line; dark shaded area denote the 0.95 uncertainty bands of the median of the forecast, lighter shaded area describes the 0.95 confidence limits of the forecast and the lightest shaded line shows the 0.95 uncertainty on the variance itself. Points denote the observations.

CONCLUSIONS

Confidence limits of the forecasts of water levels taking into account the propagation of the uncertainty within the sequential system were derived using a Monte Carlo simulation procedure. We show here, in an application to the River Severn, that the on-line assimilation of water level data constrains the uncertainty propagation within the cascade of model components in the system for flood forecasting. The methodology is being incorporated into the UK National Flood Forecasting System.

ACKNOWLEDGEMENTS

This work was done as a part of the EPSRC project “Predicting the probability of flooding over long reaches (including real time applications)” (GR/R66044/01) and the UK Flood Risk Management Research Consortium (Research Priority Areas 3 and 9, http://www.floodrisk.org.uk/). The Environment Agency is thanked for providing the rainfall and water level data.

REFERENCES

Buizza, R., Miller, M. and Palmer, T.N., (1999). *Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System*. Quarterly Journal of the Royal Meteorological Society, 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.

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: W.R. White and J. Watts (Editors), River Flood Hydraulics. John Wiley and Sons, Chichester, pp. 65-75.

Molteni, F., Buizza, R., Palmer, T.N. and Petroliagis, T., (1996). *The ECMWF Ensemble Prediction System: methodology and validation*. Quarterly Journal of the Royal Meteorological Society, 122: 73-119.

Pappenberger, F. and Beven, K., (2004). *Functional Classification and Evaluation of Hydrographs based on Multicomponent Mapping*. International Journal of River Basin Management, 2(2).

Pappenberger, F., Beven, K.J., Horritt, M. and Blazkova, S., (2005). *Uncertainty in the calibration of effective roughness parameters in HEC-RAS using inundation and downstream level observations.* Journal of Hydrology, 302(1-4): 46-69.

Pappenberger, F., Beven, K.J., Hunter, N., Gouweleeuw, B., Bates, P., de Roo, A. and Thielen, J., (in press) C*ascading 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 Science.

Pappenberger, F., Matgen, P., Beven, K.J., Henry, J.-B., Pfister, L. and de Fraipont, P.,(in press) I*nfluence of uncertain boundary conditions and model structure on flood inundation predictions. Advances in Water Resources.*

Romanowicz, R. and Beven, K., (2005). *Comments on Generalised Likelihood Uncertainty Estimation*. In: K.M. Hanson and F.M. Hemez (Editors), Sensitivity Analysis of Model Output, Los Alamos National Laboratory, http://library.lanl.gov/.

Romanowicz, R. and Beven, K.J., (1998). *Dynamic real-time prediction of flood inundation probabilities.* Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 43(2): 181-196.

Romanowicz, R., Young, P.C. and Beven, K., (2004). *Data assimilation in the identification of flood inundation models: derivation of on-line multi-step ahead predictions.* In: B. Webb et al. (Editors), BHS International Conference: Hydrology, Science and Practice for the 21st century, pp. 348-353.

Young, P., (2002). *Advances in real-time flood forecasting.* Philosophical Transactions: Mathematical, Physical & Engineering Sciences, 360(1796): 1433-1450.

Young, P.C., (1984). *Recursive estimation and time-series analysis: an introduction.* Springer-Verlag, Berlin; New York.