4th International Symposium on Flood Defence:
Managing Flood Risk, Reliability and Vulnerability
Toronto, Ontario, Canada, May 6-8, 2008

Institute for Catastrophic Loss Reduction

USE OF BAYESIAN MODEL AVERAGING TO DETERMINE UNCERTAINTIES IN RIVER DISCHARGE AND WATER LEVEL FORECASTS

J.V.L. Beckers[1], E. Sprokkereef[2] and K.L. Roscoe[1]
1. Deltares, PO Box 177, 2600 MH Delft, The Netherlands
2. Rijkswaterstaat, Centre for Water Management, PO Box 17, 8200 AA Lelystad, The Netherlands

ABSTRACT: Operational water management in the Netherlands depends on reliable water level and discharge forecasts for the Rivers Rhine and Meuse. This need for reliable forecasts initiated the development of FEWS NL, a new water-level and discharge forecasting system for the Dutch rivers. With the introduction of this system the lead time for reliable forecasts has been doubled, contributing substantially to the safety in a large part of The Netherlands. The system uses real-time data provided by a large number of European meteorological and hydrological gauging stations, weather forecasts from three different weather services, and rainfall-runoff and hydraulic models. FEWS NL produces not only a discharge forecast based on deterministic numerical weather predictions from two global and two local models, but also ensemble forecasts from the European Centre for Medium Range Weather Forecasts (ECMWF). The spread of the model realizations within the ensemble gives an indication of the uncertainty of the forecast. However, the translation from spread to uncertainty is not straightforward. To determine the true uncertainty, the distribution of the ensemble members has to be calibrated. To address this, a Bayesian Model Averaging (BMA) method was applied to a set of water-level forecast models used in FEWS NL. It was concluded that BMA is a promising method to produce a calibrated probabilistic forecast from a collection of competing forecast models. Further research will be performed to determine how the probabilistic forecast can be optimized, how it should be presented and how to implement this method in the forecasting and warning procedures.

Key Words: forecasts, floods, DELFT-FEWS, Rhine, Meuse, early warning, uncertainty, ensemble, BMA

1. INTRODUCTION

Reliable water-level forecasts for the Dutch rivers are of great importance for operational water management. About one quarter of the Netherlands lies below sea level and more than 60% of the country is potentially threatened by high water levels at sea and floods from the rivers. The endangered area along the Rhine river is extremely densely populated and is of significant economic and historic value. More than half of the Dutch population lives and works in this part of the country; the harbors of Rotterdam and the national airport Schiphol are located here. The potential economic damage of a flood in this area is roughly estimated at 1,200 billion Euro [Moll et al., 1996].

The accepted risk of an inundation in such an area is low. At the same time extreme events are expected to occur more frequently and in a more severe extent as a result of soil subsidence as well as climate-change induced extreme weather conditions and sea-level rise. One approach to reducing the risks and to limit the consequences of these increasing threats is through the development of improved operational warning systems.

2. WATER LEVEL FORECASTS IN THE NETHERLANDS

Water-level forecasts for the rivers Rhine and Meuse in the Netherlands are the responsibility of the Centre for Water Management (formerly RIZA) of Rijkswaterstaat. Under normal circumstances these forecasts are made every morning on a daily basis (365 days a year), mainly for navigation on the Rhine. During floods the frequency of forecasts is increased to at least twice a day.

Until the late nineties a relatively simple computer model, called LobithW, was used to forecast the water level of the Rhine. This model is based on knowledge and arithmetic principles from the fifties and calculates the water level at the gauging station Lobith near the German-Dutch border on the basis of statistical relations with a number of reference points. The model produces a forecast of the water level at the German-Dutch border with a lead time of four days. Experience shows that only the first two days are reliable [Parnet & Sprokkereef, 1997]. During the last big floods in 1993 and 1995 it was shown that the preparation time for the evacuation of a larger area should be at least 2 ½ to 3 days [Engel et al., 1994]. The existing forecasting system was not able to produce a reliable forecast for this lead time.

In the period after 1998 river-stage forecasting went through a spectacular development, not only regarding computer models, but especially in the field of available data [Sprokkereef, 2001]. Because of the immense increase in available data and the developments in the field of IT (internet, data transmission, faster computers) it became possible to use more advanced physical models in operational mode. The Centre for Water Management and Deltares have developed in the past decade, in cooperation with sister organizations in Switzerland and Germany, a so-called Flood Early Warning System [Werner, 2004]. A version of this system for the Rhine and Meuse Rivers, called FEWS NL, is presently being tested and is expected to become operational in 2008.

FEWS NL is an advanced combination of hydrological and hydraulic models with software for import, validation, interpolation and presentation of data. In comparison to the former statistical model FEWS NL uses significantly more data as input. Every 30 minutes the system receives observed water levels from about 60 gauging stations in the Rhine basin. Every hour meteorological observations are downloaded from servers at the national Dutch (KNMI) and German (DWD) weather services of more than 600 stations in the basin of Rhine and Meuse. The system uses output from four numerical weather models at KNMI, DWD and the European Centre for Medium Range Weather Forecasts (ECMWF).

This extreme increase of available data has great advantages but also creates new problems. In addition to weather forecasts from four deterministic models, the Centre for Water Management also receives ensemble weather predictions from the ECMWF. The ECMWF global ensemble produces 51 ensemble members at 40 km resolution with a lead time up to 14 days [Molteni et al., 1996; Buizza et al., 1999]. The limited area ensemble (COSMO LEPS) is composed of 16 members on a 10 km grid with a lead time of 120 hours [Marsigli et al.]. These ensemble scenarios are all fed into hydrological models, so that in the end the forecaster has more than 70 discharge predictions (see Figures 1 and 2) to choose from. With this amount of information the forecaster gets an impression of the (un)certainty of the forecast, but the translation of the spread in the hydrological output to uncertainty and probability is not straightforward. Another problem is how to communicate uncertainties to those that have to make management decisions, such as whether to evacuate an area or not. A crisis manager requires an unambiguous forecast and is not used to basing his or her decision on probabilities.

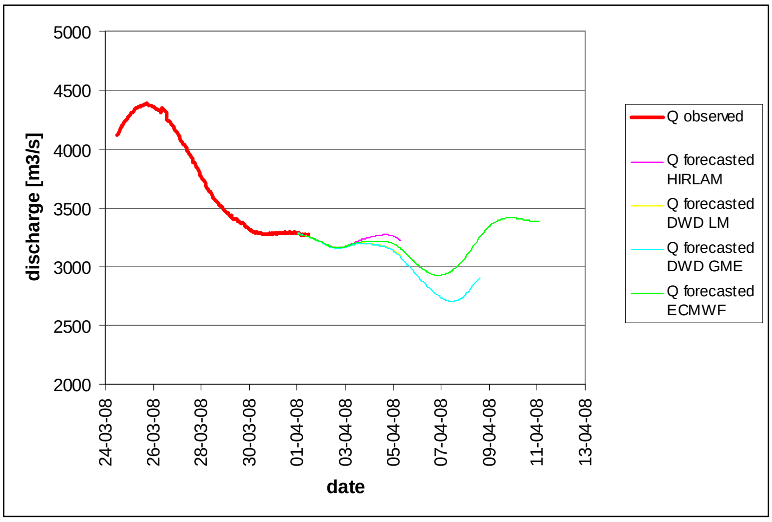

Figure 1: Deterministic discharge forecasts for the Rhine at Lobith with FEWS NL

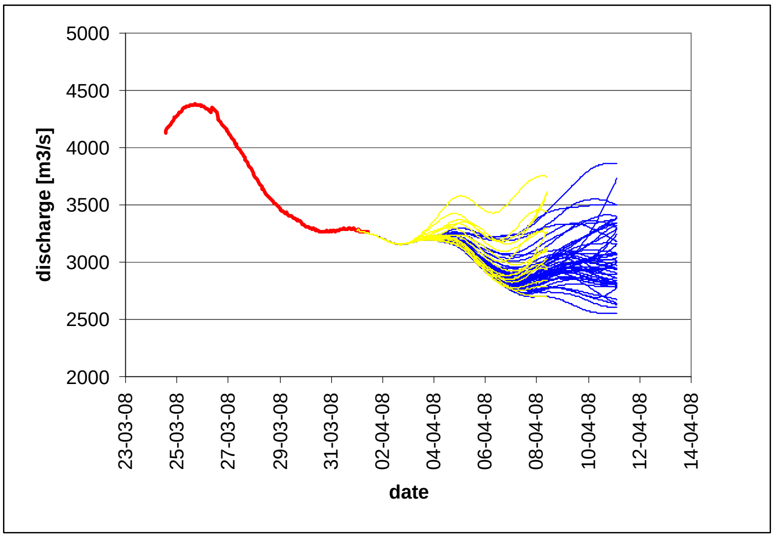

Figure 2: Probabilistic discharge forecasts for the Rhine at Lobith with FEWS NL (red = observed discharge, yellow = forecasts with COSMO LEPS ensembles, blue = forecasts with ECMWF ensembles)

3. METHODS TO CALIBRATE HYDROLOGICAL ENSEMBLE RESULTS

A common method to determine uncertainty in model results is to generate an ensemble of perturbed conditions. The perturbation can be applied to model parameters or to input data. With non-linear models (such as meteorological and hydrological models) these perturbations will grow in time and lead to a spread in the results. It is then tempting to interpret the spread in the ensembles directly as a probability distribution. However, this may lead to false conclusions because the ensemble contains only part of the uncertainty and may additionally be subject to systematic errors (bias). Due to these factors the ensemble probability distribution is not necessarily calibrated, i.e. the probability of a forecasted event is not in accordance with the true probability of the event taking place.

There are several methods to correct these errors in the probability distribution of the ensembles. The spread in the ensemble distribution and a possible systematic deviation from the mean (bias) can be corrected on the basis of historical time series. The Bayesian Forecasting System (BFS, Krzysztofowicz 1999, 2000) uses a Hydrology Uncertainty Processor (HUP) that adds an additional uncertainty to the ensemble forecast, on the basis of hindcasts of similar events. Bayesian Model Averaging (BMA, Raftery et al 2005) uses a training period, prior to the present forecast, to determine a correction for the bias and uncertainty of the ensemble. The Ensemble Kalman Filter (EnKF, Evensen 2004), as already implemented in the hydraulic model of FEWS NL, uses perturbations of model parameters and model input and calculates the effects of these perturbations in the results. The uncertainty can be combined with the spread of the precipitation ensemble. From the alternatives described the BMA and BFS approach are considered the most suitable for FEWS NL. Both methods are presently being investigated in the framework of the FEWS project. In the following section, we discuss the initial results of the application of the BMA method to the FEWS NL system.

4. APPLICATION OF THE BMA METHOD IN FEWS NL

Bayesian Model Averaging (BMA) is a standard statistical approach for post-processing ensemble forecasts from multiple competing models (Laemer, 1978). The method has been widely used in social and health sciences and was first applied to dynamic weather forecasting models by Raftery et al (2005). Details of the method can be found therein.

The basic principle of the BMA method is to generate an overall forecast probability distribution function (PDF) by taking a weighted average of the individual model forecast PDFs. The weights represent the model performance, or more specifically, the probability that a model will produce the correct forecast. In a dynamic model application, the weights are continuously updated by investigating the model performance over the most recent training period. The variance of the overall forecast PDF is the result of two components. The first component is associated with the spread between the model forecasts. The second component is the uncertainty of each individual model forecast. The magnitude of this latter component is also determined over the training period. The construction of the overall forecast PDF is illustrated in Figure 3.

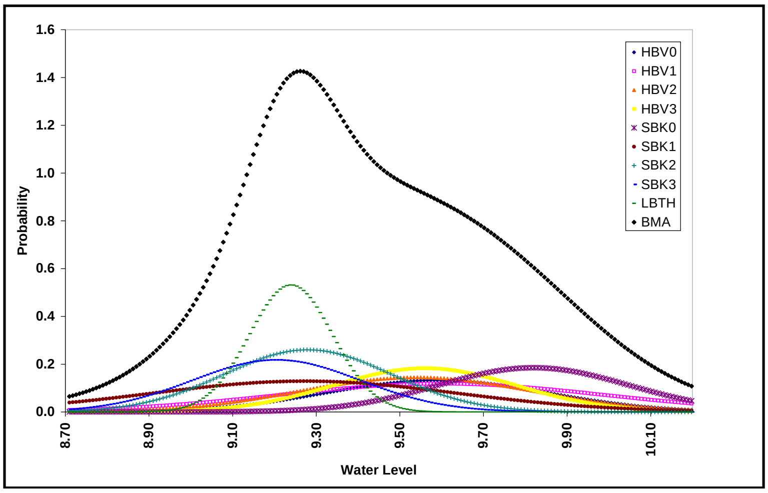

Figure 3: BMA probabilistic forecast for a single time constructed as a weighted sum of individual PDFs.

The BMA method was applied to a set of water-level forecast models run within FEWS NL. Four meteorological models were combined with both a hydrological model (HBV) and a combination hydrological-hydraulic model (HBV/SOBEK) to produce eight competing model forecasts. An additional forecast was provided by the statistical model LobithW. HBV Discharges were translated into water level values using an empirical water-level-discharge relationship at Lobith.

Water-level observations and four-day forecasts at Lobith throughout 2007 were used in the application of the BMA method. Both observations and forecasts were given as hourly values, with the exception of the statistical model, which produces daily values. In order to allow for comparisons between the models, these daily values were interpolated to produce hourly values. The forecast horizon, or lead time, was varied between 1 and 3 days. The length of the training period was varied between 1 and 4 weeks.

4.1 Results

The BMA mean forecast generally results in a lower root mean squared error (RMSE) than that of the individual model forecasts. In the current case study, the BMA RMSE was comparable to the lowest RMSE of the individual models. As shown in Table 1 below, the model with the lowest RMSE varies depending on the lead time, whereas the BMA RMSE is consistently optimal.

The main value of the BMA method lies in the fact that it produces a probabilistic forecast, which can be used to calculate a confidence interval. Figure 4 shows an example of a BMA forecast compared with observations, along with the 10% lower and upper bounds on the confidence interval. Separate training was done for different lead times. Note that the uncertainty increases for longer lead times and that the observations fall within the confidence interval at (almost) all times.

Table 1: RMSE of the individual forecast models and the BMA mean forecast for different lead times, with the lowest RMSE's highlighted in yellow. All calculations used a training period of 28 days.

Forecast	Meteorological input	Hydrological/ hydraulic model	RMSE (24-48 hrs)	RMSE (48-72 hrs)	RMSE (72-96 hrs)
1	HIRLAM	HBV	0.252	0.329	0.428
2	ECMWF	HBV	0.249	0.313	0.379
3	DWD-LM	HBV	0.249	0.302	0.347
4	DWD-GME	HBV	0.249	0.306	0.345
5	HIRLAM	HBV/SOBEK	0.196	0.258	0.381
6	ECMWF	HBV/SOBEK	0.196	0.250	0.340
7	DWD-LM	HBV/SOBEK	0.195	0.238	0.314
8	DWD-GME	HBV/SOBEK	0.195	0.239	0.303
9	LobithW (statistical model)		0.176	0.250	0.366
BMA mean forecast			0.179	0.235	0.307

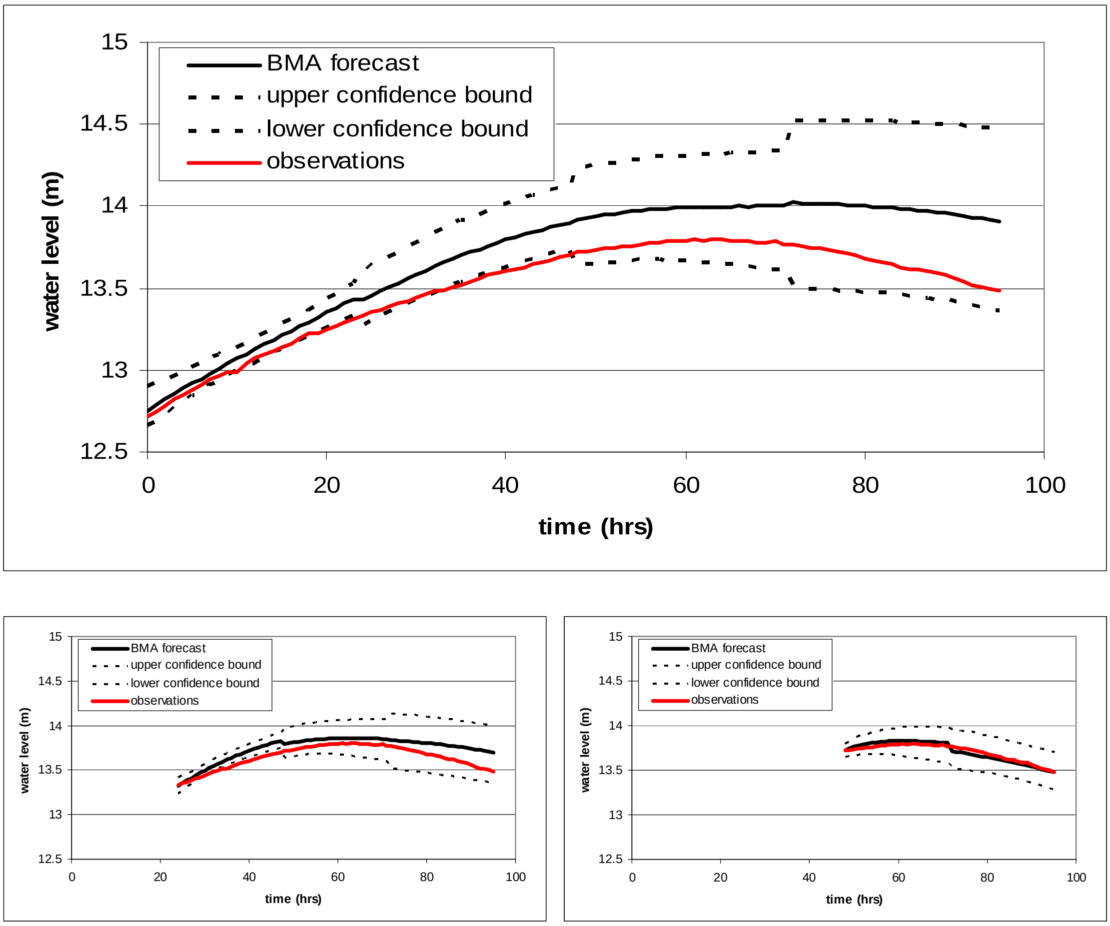

Figure 4: BMA forecast, three days before a water level peak in March 2007,
with upper and lower confidence bounds, compared with water level observations.
The lower two graphs show the forecasts for two days and one day before the maximum water level.

In the current application a confidence level of 80% was specified, which should represent the actual probability of an observation falling within the confidence interval, with 10% of the observations greater than the upper confidence level and 10% less than the lower confidence level. In the 2007 test period, this proved to be approximately the case. Table 2 presents the actual percentage of 2007 forecasts falling outside the confidence intervals, for the four different training periods. Based on these results, it appears that a training period of a minimum of three weeks should be used in the BMA method at Lobith. The agreement between expected and observed water levels outside of the confidence interval demonstrates the usefulness of the BMA method in gauging the uncertainty of a given forecast.

Table 2: Percent of observations falling outside the 80% confidence interval

Training Period (days)	**Average Width of 80% C.I.* (m)**	**> Upper C.I.***	**< Lower C.I.***	**Outside the C.I.***
7	0.52	12.3%	17.8%	30.1%
14	0.61	11.7%	13.0%	24.7%
21	0.62	11.8%	9.4%	21.2%
28	0.61	11.5%	9.7%	21.2%

* C.I. = confidence interval

5. CONCLUSIONS AND OUTLOOK

The multiple deterministic numerical weather predictions and the ensemble forecasts give an indication of the uncertainty of the water level and discharge forecasts. However, a non-calibrated ensemble cannot be translated directly into an uncertainty distribution. There are several methods to generate a calibrated probabilistic forecast. From the alternatives described the BMA and BFS approach are considered the most suitable for FEWS NL. Both methods are presently being investigated in the framework of the FEWS project. The initial results of the BMA approach are described in this article. It is concluded that the BMA is a promising method to produce a calibrated probabilistic forecast from a collection of competing forecast models that predict water levels at Lobith.

Further investigations will be performed to determine how the method can be optimized, such that it provides to the decision-maker a forecast with reliable confidence intervals. Follow-up research will focus on whether the desired forecast reliability can be met and whether decision-makers are able to use the uncertainty information in the decision process for preparation measures in case of an imminent flood.

6. REFERENCES

Beven, K.J. and Binley, A. 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Proc.* 6, 279.

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

Engel, H. et al. 1994. Das Hochwasser 1993/94 im Rheingebiet. Report No. 0833 of the Federal Institute of Hydrology, Coblence, May 1994 (in German).

Evensen, G. 2004. Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics* 54, 539–560.

Hornberger, G.M., Spear, R.C. 1980. Eutrophication in peel Inlet—I. The problem: defining behaviour and a mathematical model for the phosphorus scenario. *Water Res.* 14, 29–42.

International Commission for the Protection of the Rhine 1998.: Action Plan on Flood Defence. ICPR, Coblence, February 1998.

Krzysztofowicz, R. 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resour. Res.* 35, 2739.

Krzysztofowicz, R. Herr, H.D. 2001. Hydrologic Uncertainty Processor for Probabilistic River Stage Forecasting: Precipitation-Dependent Model, *Journal of Hydrology* 249, 46.

Leamer, E.E. 1978. Specification searches. Wiley, NY.

Marsigli, C., Boccanera, F., Montani, A. and Paccagnella, T. 2005. The COSMO-LEPS mesoscale ensemble system: validation of the methodology and verification. *Nonlinear Processes in Geophysics* 12, 527–536.

Moll, J.R., Parmet, B.W.A.H. and Sprong, T.A. 1996.: Floods and droughts in the Netherlands. *Euraqua report. Technical meeting – Rome*, October 1996.

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

Parmet, B.W.A.H. and Sprokkereef, E. 1997. Hercalibratie Model Lobith. Onderzoek naar mogelijke verbeteringen van de voorspellingen met het Meervoudig Lineaire Regressie Model Lobith na de hoogwaters van 1993 en 1995. *RIZA report 97.061*. ISBN 9036951070. Lelystad, December 1997 (in Dutch).

Raftery, A.E., Gneiting, T., Balabdaoui, F. and Polakowski, M. 2005. Using Bayesian Model Averaging to Calibrate Forecast Ensembles. *Mon. Weather Rev.* 133, 1155.

Sprokkereef, E. 2001. Extension of the Flood Forecasting Model FloRIJN. *Final Report of the IRMA-SPONGE project* No. 12. *NCR Publication* 12-2001. ISSN No. 1568-234X.

Werner, M.G.F., Van Dijk, M. and Schellekens, J., 2004, DELFT-FEWS: An open shell flood forecasting system, In 6th international conference on Hydroinformatics, Liong, Phoon and Babovic (Eds.), World Scientific Publishing Company, Singapore, 1205-1212.