IMTC 2005 – Instrumentation and Measurement
Technology Conference
Ottawa, Canada, 17-19 May 2005

Automatic Error Correction of Rainfall-Runoff models in Flood Forecasting Systems

Piet M.T. Broersen and Albrecht H. Weerts[#]
Department of Multi-Scale Physics, Delft University of Technology
P.O.Box 5046, 2600 GA Delft, The Netherlands, phone +31 15 2786419, email broersen@tnw.tudelft.nl
[#] WL | Delft Hydraulics, The Netherlands

Abstract *–Physical modelling of the dynamics of a catchment area produces simulation models with a limited forecasting accuracy for the discharge of rivers. The discrepancies between the simulation model and the actually observed past discharges can be used as additional information for error correction. With a time series model of the recent past error signal, an improved discharge forecast can be made for the next few days. The best type and order of the time series model can be selected automatically. Adaptive modelling in data assimilation calculates updates of the time series model estimated from the error data of only the last few weeks. The use of variable updated models has advantages in periods with the largest discharges, which are most important in flood forecasting.*

Keywords *– ARMA model, data assimilation, hydrological forecasting, order selection, time series model.*

I. INTRODUCTION

The importance of an accurate early warning system for river flooding is obvious. Improving the forecast of the river discharge enhances the safety of riverside residents and can prevent material damage. Moreover, an improved forecast of the river discharge or the water level has economic advantages because the maximum loading of commercial vessels in inland navigation depends on the actual water level in rivers and ports. Therefore, much attention has been given to obtain better forecasts of river discharges.

In principle, it is possible to follow the history of a single drop of rainfall somewhere in the catchment area until it evaporates or disappears in the river. This would lead to very complex models. The development of mathematical models relating the precipitation incident upon a catchment to the emanating stream flow has been a major subject for surface water hydrology for decades [1]. One unexpected conclusion was that the inclusion of spatial data did not lead to improved forecasts. Very often, one quick flow component and one slow flow component are all that can be identified and calibrated for a given catchment [1]. Many other investigations support the conclusion that physical modeling does not produce the desired accuracy in flood forecasting. It is argued that deterministic models are inappropriate because of the inherent uncertainty that characterizes river catchment dynamics [2]. The parameters of over-parameterized complex models cannot be identified with sufficient accuracy from the available data. Only relatively simple simulation models can be identified from past data [2]. Automatic calibration strategies to determine the parameters for the rainfall-runoff simulation model have been compared for a number of specific catchments [3].

The discharge data of the simulation model can be improved by data assimilation. Observations of the actually measured discharge are used to find better forecasts, which will fit closer to future measurements. Different error correction procedures have been considered, including global autoregressive (AR) models, local AR models, neural networks and genetic programming [4]. Test data consisted of hourly data of the average catchment rainfall, discharge at the catchment outlet and daily temperature data to estimate evapotranspiration. After the calibration of a simulation model, the different updating procedures have been tested. The global AR(3) model was the best for small lead times. Artificial neural networks gave the least efficient forecasts [4]. A further study of the transfer function in neural networks gives a preference for the logistic function [5]. A comparison of neural networks and autoregressive moving average (ARMA) models for the error correction concluded that similar results could be obtained with both methods [6]. The amount of user interaction and the specifically required skills of the user are decisive which method to prefer [6].

Recently, a completely automatic algorithm for time series analysis with ARMA models has been developed and made available [7,8]. Many AR, MA and ARMA candidate models are computed for a given random data set. Order selection criteria select a single model. Akaike's AIC criterion [9] is prone to select overfit models whereas consistent order selection criteria [10] have problems with underfit. The best order selection criterion is a compromise between overfit and underfit, and it is also adapted to finite sample properties [11]. The increased computer power enables the routine computation of many candidate models and the improved order selection criteria give a single good time series model, for small or large data sets alike. That model is used to determine spectrum, autocorrelation and prediction.

This paper studies the possibilities of the automatic time series program ARMASA [8] for error correction in rainfall-runoff models. Without user interaction, forecasts are made. The length of the past error signal is chosen to give the best predictions in times of sudden strong discharge increases, which are most important for timely flood warnings.

II. TIME SERIE MODELS

Three different linear types of time series models can be distinguished: autoregressive or AR, moving average or MA and combined ARMA models. An ARMA(p,q) model can be written as [12]

$$x_n + a_1 x_{n-1} + \cdots + a_p x_{n-p} = \varepsilon_n + b_1 \varepsilon_{n-1} + \cdots + b_q \varepsilon_{n-q} \ , \qquad (1)$$

where ε_n is a purely random process of independent identically distributed stochastic variables with zero mean and variance σ_ε^2. It is purely AR for $q = 0$ and purely MA for $p = 0$. The power spectrum $h(\omega)$ of the ARMA(p,q) process is completely determined by the parameters in (1) together with the variance σ_ε^2 and is given by:

$$h(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{\left| B_q\left(e^{j\omega i}\right) \right|^2}{\left| A_p\left(e^{j\omega i}\right) \right|^2} = \frac{\sigma_\varepsilon^2}{2\pi} \frac{\left| 1 + \sum_{i=1}^{q} b_i \, e^{-j\omega i} \right|^2}{\left| 1 + \sum_{i=1}^{p} a_i \, e^{-j\omega i} \right|^2}. \qquad (2)$$

The inverse integral Fourier transformation of (2) gives the autocorrelation function, which may have an infinite length. Also direct transformations from the time series parameters into the autocorrelation function exist [12]. The transform of the true or the estimated parameters of an AR(p) model into the first p lags of a positive semi-definite autocorrelation function is made with the Yule-Walker equations [12]

$$\sum_{i=0}^{p} a_i \rho(m-i) = 0, \qquad m \geq 1 \qquad (3)$$
$$with \ \ \rho(0) = 1, \ \ \rho(-k) = \rho(k), \ k > 0$$

An ARMA(p,q) process is equivalent with an AR(p) process with white noise as input, followed by a MA(q) process. The parameters of the estimated and selected time series model give the best estimators for spectrum and autocorrelation. The parameters of that time series model are also particularly suited for prediction. The optimal prediction with an AR(p) model requires only the last p observations, also for multi-step ahead prediction [12]. The linear prediction with MA or ARMA models uses a recursive method that is also used for the computation the exact likelihood for an arbitrary normally distributed signal with mean zero [13]. The state space representation of Jones is particularly useful [14]. All programs for the computation of the spectral density, the autocorrelation function and the prediction of future observations are part of the ARMASA toolbox [8]. That requires only the data and the desired prediction horizon to compute the predictions.

Generally, the mean value of the data is treated separately in predictions. The mean is subtracted before the time series model is estimated with ARMASA. Also predictions are computed for this signal with subtracted mean. Afterwards, the previously subtracted mean is added to the predicted value to complete the prediction procedure.

III. CASE STUDY FEWS-RHINE

The Flood Early Warning System FEWS-Rhine is a prototype flood forecasting system developed in close cooperation by the Federal Institute of Hydrology in Germany (BfG), the Institute for Inland Water Management and Waste Water Treatment in the Netherlands (RIZA) and Delft Hydraulics. This prototype is now being replaced by a newly developed open architecture flood forecasting system that has been developed by Delft Hydraulics [15] and is used by the Environment Agency (UK) for England and Wales. This new system will probably be operational in 2006.

The Rhine basin is extremely well instrumented. Table 1 shows a list of stations and tributaries currently incorporated in the FEWS-Rhine system. Two types of stations are available. The runoffs observed at stations on major tributaries and at Maxau are used as boundary conditions to the SOBEK routing model [16]. For forecasting, the runoff for these stations is determined using the hydrological rainfall runoff HBV-96 model [17]. The most suitable option for data assimilation is automatic error correction.

Table 1 List of stations on the Rhine and its major tributaries

Station	Tributary	Chainage (km)	Trib-area (km2)	Area (km2)
Maxau		362.30		50,343
Speyer		400.60		53,235
Mannheim		424.90		54,136
Rockenau	Neckar	428.20	14,000	68,486
Worms		443.40		68,936
Oppenheim-Nierstein		480.60		70,462
Frankfurt	Main	496.60	27,200	98,488
Mainz		498.30		98,488
Bingen		528.40		99,277
Grolsheim	Nahe	530.00	4,100	103,407
Kaub		546.20		103,729
Boppard		570.45		103,981
Kalkofen	Lahn	588.00	5,900	109,994
Koblenz		591.50		110,131
Cochem	Mosel	592.00	28,100	138,231
Nettegut	Nette&Wied	610.00	1,100	139,586
Andernach		613.80		139,795
Altenahr	Ahr	629.00	850	140,837
Bonn		654.70		141,162
Menden	Sieg	660.00	2,900	144,217
Koln		688.00		144,612
Opladen	Wupper	702.00	800	145,618
Neubrück	Erft	738.00	1,800	147,948
Dusseldorf		744.20		148,040
Hattingen	Ruhr	779.00	4,500	153,143
Ruhrort		780.80		153,176
Wesel		814.00		154,528
Schermbeck	Lippe	815.00	4,900	159,428
Rees		837.40		159,683
Emmerich		851.90		159,784
Lobith		862.22		160,800

IV. HYDROLOGICAL MODEL OF RHINE BASIN

A hydrological model of the German part of the Rhine basin and the Mosel river, shown in Fig. 1, has been developed by the Federal Institute of Hydrology in Germany [18] for the Institute for Inland Water Management and Waste Water Treatment in the Netherlands (RIZA). The HBV-96 model has been used of the Swedish Meteorological and Hydrological Institute. For modelling purposes, twelve basins of major tributaries have been defined (Neckar, Main, Nahe, Lahn, Upper Mosel, Saar, Sauer, Lower Mosel, Sieg, Erft, Ruhr, Lippe) alongside with four basins of Rhine-stretches (high, upper, middle and lower Rhine).

Fig 1. The Rhine basin.

Subsequently, the basins have been further divided into a number of areas in order to model spatial heterogeneity. These areas can be considered the basic units of the hydrological model of the Rhine. The number of units into which a sub-basin is divided depends on the number of available discharge gauging stations, since discharge data is used to calibrate and validate the model for each unit.

A. HBV-96 hourly model

The HBV-96 hourly model [17] is a conceptual precipitation-runoff model, which simulates snow accumulation, snow melt, actual evapo-transpiration, soil moisture storage, groundwater depth and runoff. The model input consists of precipitation, temperature and potential evaporation. The model consists of three major components: a snow routine, a soil routine and a runoff response routine consisting of a quick and slow reservoir.

B. Nahe river, HBV-96 model application

The Nahe river basin is made up of three HBV sub-basins each having an gauging station at the outlets of Martinstein, Boos and Grolsheim, respectively. The measured or modelled discharge at Grolsheim is used as input into the hydraulic model in the flood forecasting system FEWS-Rhine. Each sub basin is divided into several height zones and each height zone has two forms of vegetation resulting in a quite a number of states.

C. Mosel river, HBV-96 model application

The Mosel river basin is made up of 26 sub-basins. Again, each sub-basin is divided into several height zones and each height zone has two forms of vegetation. The measured or the modelled discharge at Cochem is used in the flood forecasting system as input to the hydraulic model used in the flood forecasting system FEWS-Rhine.

V. APPLICATION OF ERROR CORRECTION

Some results from the hydraulic HBV-96 model, together with the actual measured data are given in Fig.2 for the Mosel river. The high peak around day 70 represents an extremely high discharge, in November 1998. The rainfall-runoff model result is obtained with the HBV-96 hourly hydrological model [17]. The purpose in this paper is to use both this runoff model and the observations measured until a certain time *t* to make the best prediction for the discharge at future times *t+1* until *t+L*. The accuracy of the HBV-96 model is no subject in this paper, but only the improvement that can be obtained by using the error between the actual measurements and the model in the recent past, up to the present moment, for a prediction of the future error of the HBV-96 model.

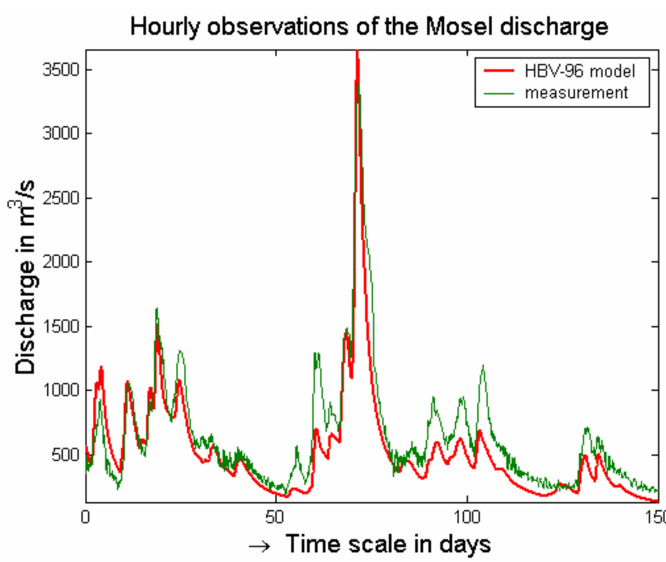

Fig. 2. Measured discharges of the Mosel and a rainfall-runoff results of the HBV-96 hydrological model in the period 1998 – 1999.

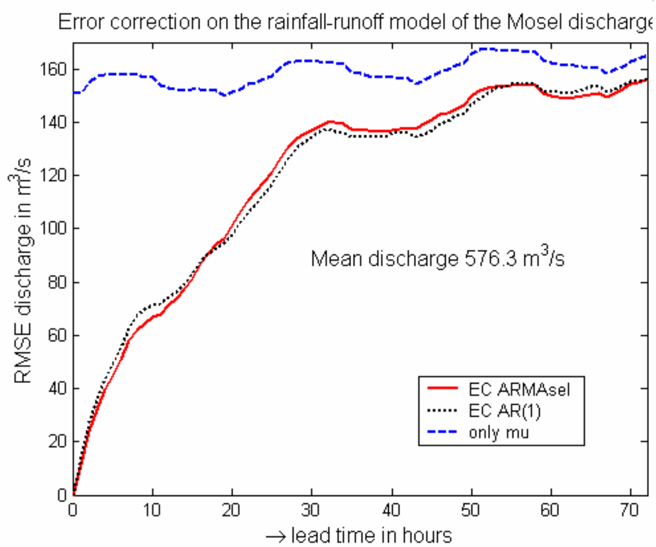

Fig. 3. Average forecast accuracy for a period of 3 days for the Mosel catchment; 150 updates have been made with one day intervals, with a model estimated from the error signal in the last three weeks. The error correction is made with the automatically selected ARMAsel model, with the estimated AR(1) model and with only the mean value over the last three weeks.

As prediction of the discharge, it is possible to use the HBV-96 model output at time *t* plus an error correction that is based on the observed difference between model and measurement until time *t*. Also the model output itself could be extrapolated to the future and rainfall forecasts can be incorporated to improve those model extrapolations [3]. However, this paper is only concerned with the best error correction and does not discuss the simulation rainfall-runoff model. The future output of the given simulation model is used to compute the root mean square error in the error correction figures, which give the average differences between the hourly predictions for the next 3 days and the observed differences between model and measurement for the same lead times.

Fig. 3 shows that the dynamic error correction gives an important improvement of the accuracy during the first day. For prediction further ahead, only a slight difference is found with the correction that uses only the mean value. As a reference number, the root mean square error RMSE between the model discharge and the measurements without any error correction would be 199.6 for the data used in Fig. 3. It is notable that the error correction with the AR(1) model and with the selected ARMAsel model are rather close in this example. Results similar to AR(1) are found with all fixed order AR models from AR(1) until AR(10). The selected AR orders varied between 2 and 8 in the 150 individual updates of Fig. 3. It turns out that the AR(1) parameter or reflection coefficient is very dominant in the response. The AR(2) parameter is also statistically significant and the higher order parameters are on the boundary of statistical significance for the number of past observations used in the estimation of the update model, which was 504.

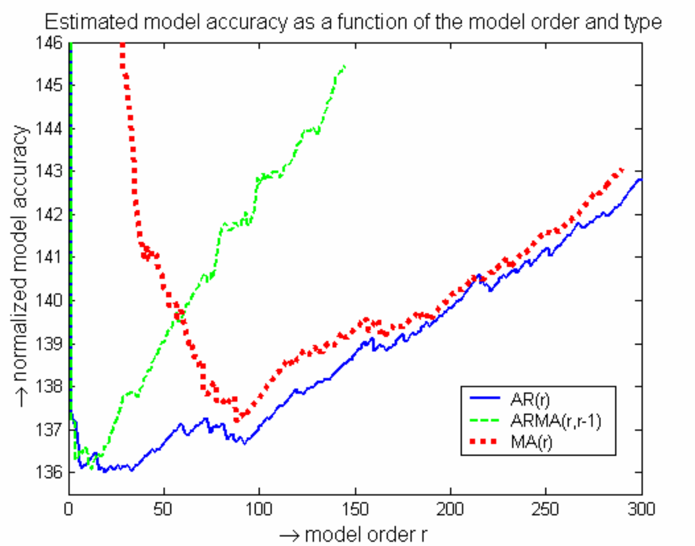

Fig. 4. Estimated one step ahead squared prediction error as the model accuracy of a number of candidate time series models estimated from 4292 observations of the difference between the measured and the HBV-96 model discharges of the Mosel. The AR(7) model was selected for those data. The estimated normalized accuracy of the AR(0) model was 21424 and of the AR(1) model was 191. It is obvious that many AR and ARMA models are suitable candidates as model for the error signal, with almost the same accuracy of about 136. However, low order MA models are poor.

The ARMAsel program can give additional information about the estimated accuracy of the squared one step ahead prediction of all candidate time series models that have been estimated with ARMAsel [8]. The estimated accuracies of the time series models in Fig. 4 shows that many global AR and ARMA models give a reasonable accuracy. It should be realized that the characteristics of the data, like in Fig. 2, are not really stationary. This will have a strong impact on the predictions further ahead. Probably as a result hereof, also predictions with an AR(1) model give good results in Fig. 3, comparable with the ARMAsel result. Only 504 observations are used for each predicting model in Fig. 3, whereas Fig. 4 is based on many more observations. More parameters become statistically significant then. It has also been verified that predictions with a global AR(30) model give almost the same accuracy as predictions with local AR(1) or AR(2) models. If more than about 100 parameters are estimated, the quality of AR and MA models is approximately the same in Fig.4.

Fig. 5 gives the results of error correction for the Nahe catchment. It is obvious that the behavior is quite different from the Mosel. The horizon for accurate predictions is only about 15 hours here. Furthermore, the performance of the AR(1) correction is much worse than that of ARMAsel. Also Fig. 6 with the model accuracies looks quite different from Fig. 4 with the Mosel results. This shows that ARMAsel with selection of the best model gives satisfactory error corrections for the various circumstances that have been studied. Also results for the Main and the Neckar catchment have been tested, with the same conclusion. Using fixed order AR(1) or AR(2) will be good in some examples, poor in others.

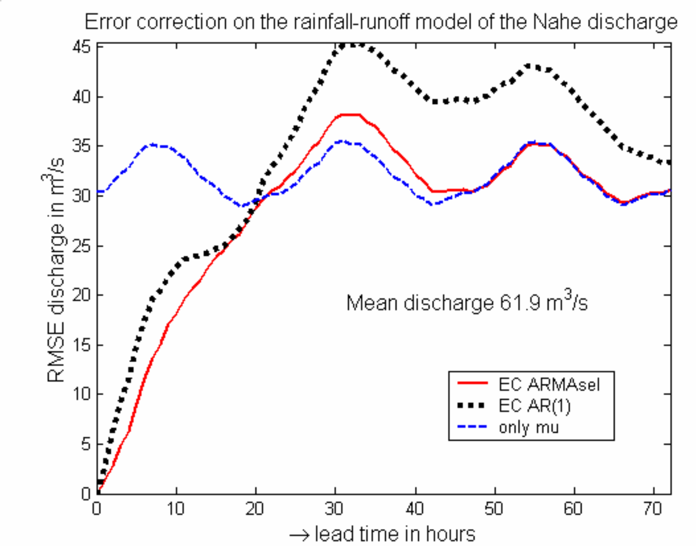

Fig. 5. Average forecast accuracy for a period of 3 days for the Nahe catchment; 150 updates have been made with one day intervals, with a varying model based on the last 504 hours. The error correction is made with the automatically selected ARMAsel model, with the estimated AR(1) model and with only the mean value over the last 504 hours.

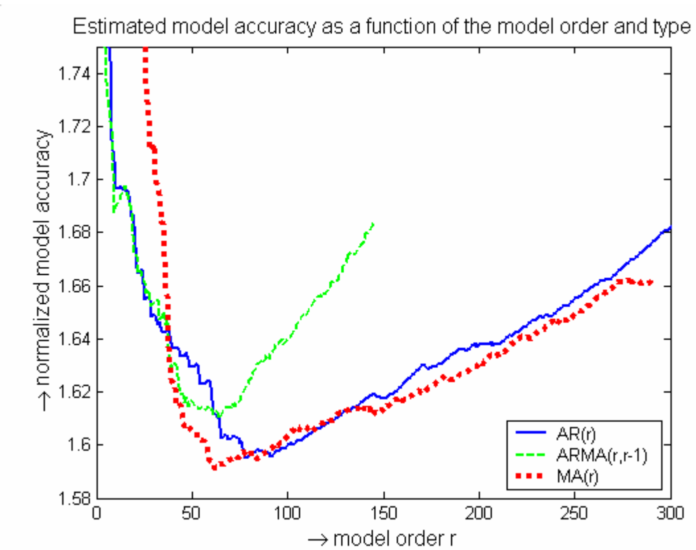

Fig. 6. Estimated one step ahead model accuracy of candidate time series models estimated from 4292 observations of the difference between measured and simulated discharges of the Nahe. The MA(62) model was selected for those data. The estimated accuracy of the AR(0) model was 846, of the AR(1) model was 7.3 and of the AR(3) model was 1.8. AR(1) is not a good choice for those data.

Especially the behavior of the AR models is quite different for the Mosel and Nahe data in Fig. 4 and Fig. 6, respectively. The best global model for the Nahe is the MA(62) model. That means that the autocorrelation function is completely damped out after 62 hours. The fact that prediction in the Mosel model is possible for a longer lead time than in the Nahe follows also from the fact that the autocorrelation function of the Nahe discharge data dies out much quicker in Fig. 7.

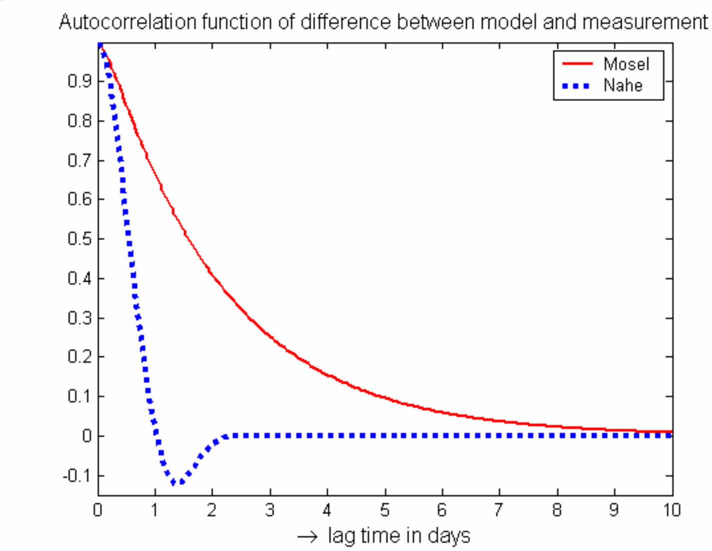

Fig. 7. Estimated autocorrelation functions of the difference between the HBV-96 model and the measured discharges of the Mosel and Nahe catchments.

It is good if the average error over a long period is small, but it would still be better for early flood forecasting if the extreme values are predicted well. Therefore, the error correction has been tested on the interval from week 67 and 76 in Fig. 2. Results are given in Fig. 8. With those strong varying discharges, the lead time giving still accurate predictions is smaller than in Fig.2 and the remaining error is more than 50 % greater. Using the selected model from ARMAsel is significantly better here than using the fixed AR(1) model. The RMS error without correction is 362 m^3/s in those 10 days. Hence, the error is reduced with more than 50 % for a lead time of almost a day, which may lead to a much a better quality for the overall flood warning system.

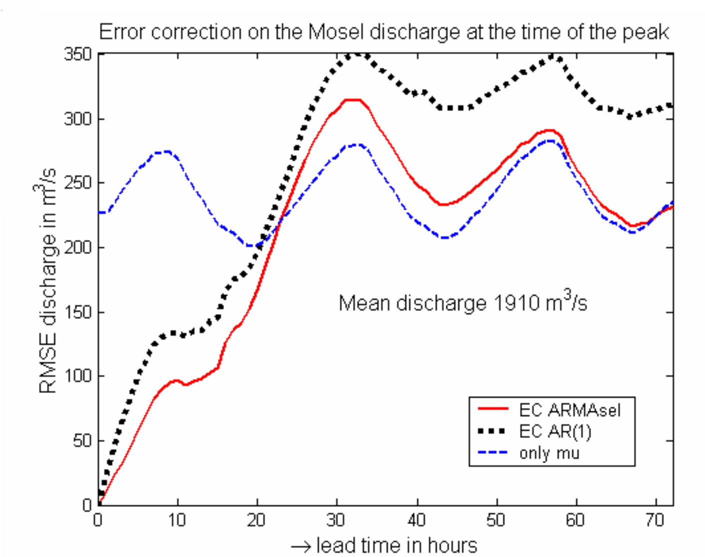

Fig. 8. Average forecast accuracy for a period of 3 days for the Mosel catchment; 10 updates have been made with one day intervals, starting at day 67 in Fig. 2 with a varying model based on the last three weeks. The evaluated section is has the highest discharge.

Table 2. Influence of the history length in hours, used for estimation with ARMAsel on the RMSE accuracy of the Mosel discharge error correction, averaged over the first 12 hours of the prediction. Predictions have been made over the 150 day period and over a short period of 10 days with the peak discharge .

History length in hours	150 day period	10 days, peak period
10	83.5	216.4
20	79.4	203.5
50	62.8	129.5
100	57.8	108.0
200	53.1	80.4
300	51.5	71.9
400	50.4	67.1
500	49.6	***66.3***
750	***49.3***	72.6
1000	50.8	85.8
2000	50.7	85.6
3500	50.5	80.8

An AR(1) model gives often a reasonable error correction and it can be estimated if 10 or more observations are available. Therefore, error correction can use a very short history and still be successful. Which length of the history used for computing the error correction model, however, will give the best error correction? The history length has been varied to study the influence on the error correction accuracy. Table 2 gives the average RMSE of the first 12 hours of the predictions, like in Fig. 3 for the whole 150 day period and in Fig. 8 for the 10 day peak period, both as a function of the history length used for parameter estimation in each daily update. The error correction has always been made with the automatically selected ARMAsel model.

Table 2 shows that a using very short period as history for estimation of the error correction model is not advisable. The difference is greater for the peak period than for the longer period. The RMSE is much greater than can be obtained with a longer history. A history length of around 500 hours is the best. The average RMSE in the discharge over a long period is steadily becoming almost a constant if more data are used for the computation of the ARMAsel model that is used for error correction. However, the modeling of the peak discharges has a different performance. The accuracy is becoming less if a longer history is used to estimate the model for error correction. Here, a history length of 500 hours or three weeks gives the best predictions.

VI. CONCLUDING REMARKS

The ARMAsel algorithm has a good performance for small samples as well as for a very large number of observations. History lengths from 10 hours to several months can be used for estimation without any problem.

In stationary conditions, the prediction accuracy becomes somewhat better if a longer history of past errors is used to compute the time series model. A length between 300 and 750 hours gives a good response in quickly changing circumstances as well as in more stationary conditions.

The autocorrelation function and the prediction horizon are rather different for catchments modelled with the same hydrological model. The use of ARMAsel automatically adapts the error correction to the dynamics of the catchment that is involved.

ACKNOWLEDGEMENTS

The hydrological model of the German part of the Rhine basin and the Mosel river has been developed by the Federal Institute of Hydrology in Germany (BfG) for the Institute for Inland Water Management and Waste Water Treatment in the Netherlands (RIZA).

REFERENCES

[1] A.J. Jakeman and G.M. Hornberger, “How much complexity is warranted in a rainfall-runoff model?,” *Water Resources Research*, vol. 29, pp. 2637-2649, August 1993

[2] P.C. Young, “Advances in real-time flood forecasting,” *Phil. Trans. Royal SOC. London*, vol. A 360, pp. 1433-1450, 2002.

[3] H. Madsen, G. Wilson and H.C. Ammentorp, “Comparison of different automated strategies for calibration of rainfall-runoff models,” *Journal of Hydrology*, vol. 261, pp. 48-59, 2002.

[4] H. Madsen, M.B. Butts, S.T. Khu and S.Y. Liong, “Data assimilation in rainfall-runoff forecasting,” *4th Int. Conf. On Hydroinformatics*, Cedar Rapids, Iowa, USA, 8 pp., July 2000.

[5] A.Y. Shamseldin, A.E. Nasr and K.M. O’Connor, “Comparison of different forms of the multi-layer feed-forward neural network method used for river flow forecasting,” *Hydrology and Earth System Sciences*, pp. 671-684, 2002.

[6] R.J. Abrahart and L. See, “Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments,” *Hydrological Processes*, vol. 14, pp. 2157-2172, 2000.

[7] P.M.T. Broersen, “Automatic spectral analysis with time series models,” *IEEE Trans. Instrum. Meas.*, vol. 51, pp. 211-216, April 2002.

[8] P.M.T. Broersen. ARMASA, Matlab toolbox, freely available at http://www.mathworks.com/matlabcentral/ then select: File Exchange, Signal Processing, Spectral Analysis, ARMASA.

[9] H. Akaike, “A new look at the statistical model identification,” *IEEE Trans. Autom. Control*, vol. AC-19, pp. 716-723, 1974.

[10] J. Rissanen. (1978), “Modelling by shortest data description,” *Automatica*, vol. 41, pp. 465-471, 1978.

[11] P.M.T. Broersen, “Finite sample criteria for autoregressive order selection,” *IEEE Trans. on Signal Process.*, vol. 48, pp. 3550-3558, Dec. 2000.

[12] M.B. Priestley, *Spectral Analysis and Time Series.* London, U.K: Academic Press, 1981.

[13] P.J. Brockwell and R.A. Davies, *Time Series: Theory and Methods*. Springer Verlag, New York, 1987.

[14] R.H. Jones, “Maximum likelihood fitting of ARMA models to time series with missing observations.,” *Technometrics*, vol. 22, pp. 389-395, 1980.

[15] M.G.F. Werner, M. van Dijk and J. Schellekens, “DELFT-FEWS: An open shell flood forecasting system,” *Proceedings 6th Int. Conf. on Hydroinformatics*, World Scientific Publishing Company, Singapore, pp. 1205-1212, 2004.

[16] H.J. Barneveld and D.G. Meijer, “SOBEK-model Andernach-Lobith. Model construction, calibration and verification.” *HKV-Geodan report* prepared for RIZA, 1997.

[17] G. Lindstrom, B. Johannson, M. Persson, M. Gardelin and S. Bergstrom, “Development and test of the distributed HBV-96 hydrological model”, *J. Hydrol*., vol. 201, pp. 272-288, 1997.

[18] Federal Institute of Hydrology (BfG). “Hydrological modelling in the Rhine basin, Part II -Report on hourly modelling”, pp. 1-37. 2001.