Hydrol. Earth Syst. Sci. Discuss., 6, 1677–1706, 2009
www.hydrol-earth-syst-sci-discuss.net/6/1677/2009/

Hydrology and Earth System Sciences Discussions

Papers published in *Hydrology and Earth System Sciences Discussions* are under open-access review for the journal *Hydrology and Earth System Sciences*

A novel approach to parameter uncertainty analysis of hydrological models using neural networks

D. L. Shrestha[1], **N. Kayastha**[2], **and D. P. Solomatine**[1,3]

[1]UNESCO-IHE Institute for Water Education, Delft, The Netherlands
[2]MULTI Disciplinary Consultants Ltd, Kathmandu, Nepal
[3]Water Resources Section, Delft University of Technology, The Netherlands

Received: 17 February 2009 – Accepted: 19 February 2009 – Published: 4 March 2009

Correspondence to: D. L. Shrestha (d.shrestha@unesco-ihe.org)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

In this study, a methodology has been developed to replicate time consuming Monte Carlo (MC) simulation by using an Artificial Neural Network (ANN) for assessment of model parametric uncertainty. First, MC simulation of a given process model is run. Then an ANN is trained to approximate the functional relationships between the input variables of the process model and the synthetic uncertainty descriptors estimated from the realizations. The trained ANN model encapsulates the underlying characteristics of the parameter uncertainty and can be used to predict uncertainty descriptors for the new data vectors. This approach was validated by comparing the uncertainty descriptors in the verification data set with those obtained by MC simulation. The method is applied to estimate parameter uncertainty of a lumped conceptual hydrological model, HBV, for the Brue catchment in UK. The results are quite promising as the prediction intervals estimated by ANN are reasonably accurate. The proposed techniques could be useful in real time applications when it is not practicable to run a large number of simulations for complex hydrological models and when the forecast lead time is very short.

1 Introduction

Monte Carlo (MC) simulation is the most widely used method for uncertainty analysis. It involves random sampling from the distribution of parameters inputs and successive model runs until a desired statistically significant distribution of outputs is obtained. The main advantage of the MC simulation is its general applicability, however methods of this type require a large number of samples (or model runs), and their applicability is sometimes limited to simple models. In case of computationally intensive models, the time and resources required by these methods could be prohibitively expensive. A version of the MC simulation method was introduced under the term "generalised likelihood uncertainty estimation" (GLUE) by Beven and Binley (1992). The GLUE concept

HESSD

6, 1677–1706, 2009

has become a popular method in analysing the parameter uncertainty in hydrological modelling. (On the appropriateness of attributing the GLUE methodology to Bayesian methods see Mantovan and Todini (2006), and the subsequent interesting discussion in the J. Hydrol. in 2007.)

Although the MC simulation is conceptually simple and very flexible, it is well recognized that MC-based simulation still lack a well-established convergence criteria to terminate the simulations at a desired level of accuracy. A number of researches were conducted to reduce the number of MC simulation runs and to identify the convergence criteria (e.g., McKay et al., 1979; Rossel et al., 2001; Ballio and Guadagnini, 2004).

In addition, MC based method for uncertainty analysis of the outputs of such models is straightforward, but becomes impractical in real time applications when there is no time to perform the uncertainty analysis because the large number of model runs is required. For such situations alternative methods were developed, referred to as the moment propagation technique, which are able to directly calculate first and second moments without application of MC simulation (see, e.g., Rosenblueth, 1981; Harr, 1989; Protopapas and Bras, 1990; Melching, 1995; Kunstmann et al., 2002).

We propose to use ANN to approximate the uncertainty descriptors of a model output (this latter model will be further referred as *M*). The idea of using statistical and, in general, machine learning models, to improve model accuracy is not new. Typically, information about model errors is used to train data-driven error correctors (Abebe and Price, 2004) or to build more sophisticated data-driven models of model uncertainty (Shrestha and Solomatine, 2006, 2008). In this study, we extend this idea towards building a model (referred as *V*) encapsulating the information about the realizations of the process (hydrologic) model *M* output generated by MC simulations. Instead of predicting a single value of the output of *M* as done in the most error correction procedures, we aim at predicting the distribution of the output of *M* generated by MC based simulations. Thus, the method allows one to predict the uncertainty bounds of the model *M* prediction without re-running the MC simulations when new input data is observed and fed into *M*.

In this study, the approach is tested on estimating the two quantiles of the MC simulation of the model output. Thus, the model parameter uncertainty is measured by only two quantiles of the probability distribution that constitutes the prediction interval of the model output corresponding to some confidence level (say, e.g. 90%). The experiments were carried out with ANN. The HBV hydrological model of the Brue catchment in the United Kingdom is used as a case study.

2 Main application of ANN in hydrological modelling

ANN is a popular technique used to discover a dependency between inputs and outputs of a physical system from the available data. By data we understand the known samples combinations of inputs and corresponding outputs. As such a dependency ("model") is discovered, it can be used to predict the future system's outputs from the known input values.

ANN has been extensively used, especially in rainfall-runoff (R-R) modelling (Minns and Hall, 1996; Dawson and Wilby, 2001; Abrahart and See, 2000; Govindaraju and Rao, 2000). Apart from ANN, other machine learning techniques have been also used: for example, fuzzy rules based system (Bardossy et al., 1995; Klir and Yuan, 1996), model trees (Solomatine and Dulal, 2003), support vector machines (Dibike et al., 2001). However, the application of such techniques to estimate the uncertainty of physically based or machine learning based R-R modelling is very limited. Abebe and Price (2004) used ANN to forecast the surge prediction accuracy along the Dutch coast in North Sea. Shrestha and Solomatine (2006) used machine learning techniques to estimate non-parametric uncertainty of river flow forecasting by ANN and other machine learning techniques in the Sieve river basin, Italy. Shrestha and Solomatine (2008) and Solomatine and Shrestha (2009) used machine learning techniques to estimate uncertainty of the simulated river flows by a conceptual R-R model to various case studies. The details of ANN can be found in Bishop (1995) and Haykin (1999), and the overviews of ANN applications within hydrology can be found in Maier and Dandy

(2000), and Dawson and Wilby (2001).

3 Methodology

3.1 Basic idea

There are a number of assumptions to consider. First, we assume that the uncertainty of a hydrological model output depends on the forcing input data and the model states (e.g., rainfall, antecedent rainfall, soil moisture etc.). We also assume that uncertainty associated with the prediction of a hydrological variable, e.g. runoff, has similar magnitude for similar hydrological conditions. By the hydrological conditions we understand here the combination of the state of input data and the state variables, which are forcing or driving to generate the runoff in the catchment. For example, one can see that the prediction of extreme events such as peak flows is more difficult if compared to the low flows. Consequently, uncertainty of the flow prediction in the peak flow is higher as compared to those for low flow.

Instead of building a model of the error in the process model output, as it is done in the most of the error updating procedures (e.g., Abebe and Price, 2003), in the presented approach a predictive model for the parameters of the distribution of the process model output is built (this distribution is generated by MC simulations). Thus, our method allows one for predicting the uncertainty bounds of the model prediction without running the MC simulations in real time application.

3.2 Definition of the process model *M*

Consider a deterministic model *M* of a real world system predicting a system output variable y given the input data vector $\boldsymbol{x}$, initial condition of the state variables s_o and the vector of the parameters θ. The model *M* could be physically based, conceptual, or even data driven. In this paper we assumed model *M* is a conceptual hydrological

model. The system response can be represented by equation:

$$y = M(\boldsymbol{x}, s, \theta) + \varepsilon = \hat{y} + \varepsilon \quad (1)$$

where ε is the the model error between the observed response y and the corresponding model response $\hat{y}$. Before running the model M, the components of the model, i.e. input data $\boldsymbol{x}$, initial conditions s_o, parameters vector θ and the model structure itself have to be specified, while the output or model response $\hat{y}$ and the state variable s are computed by running the model. These components may be uncertain in various ways to various degrees; the consequences of these uncertainties will be propagated into the model states and the outputs. In this paper, however, only uncertainty associated with parameters vector θ is considered.

3.3 Monte Carlo simulation

The MC simulation is performed by running the hydrological model M multiple times either changing the input data $\boldsymbol{x}$ or parameters vectors or even the structure of the model or combination of them. In this paper we assume that the model structure and the input data is certain (correct), so mathematically this can be expressed as:

$$\hat{y}_{t,i} = M(\boldsymbol{x}, \theta_i); \quad t = 1, 2, \ldots, n; \quad i = 1, 2, \ldots, s \quad (2)$$

where θ_i is the set of parameters sampled for i^{th} run of MC simulation, $\hat{y}_{t,i}$ is the model output of the t^{th} time step for i^{th} run, n is the number of time steps and s is the number of simulations.

3.4 Estimation of the prediction interval

The statistical properties (such as moments and quantiles) of the model output for each time step t are estimated from the realizations $\hat{y}_{t,i}$. One of the ways to judge about the uncertainty of model output is to use the error variance: large variance of the model

error typically indicates that the model prediction is uncertain. In most cases, however, variance does not sufficiently describe the uncertainty, and more informative quantities such as prediction intervals are used.

Prediction interval is comprised of upper and lower limits between which a future unknown value is expected to lie with the prescribed probability. These limits are typically the quantiles of the model output distribution. In each simulation, model output is given a different weight (to be defined later), so a quantile can be found using the following equation:

$$P(\hat{y}_t < \hat{Q}(p)) = \sum_{i=1}^{n} w_i | \hat{y}_{i,t} < \hat{Q}(p) \tag{3}$$

where, $\hat{y}_t$ is the model output at time step t, $\hat{y}_{t,i}$, is the value of model outputs at time t simulated by the model $M(\boldsymbol{x}, \theta_i)$ at simulation i, $\hat{Q}(p)$ is p% quantile, w_i is the weight given to the model output at simulation i. Quantiles obtained in this way are conditioned on the inputs to the model, model structure, and the weight vector w_i.

In order to compute the prediction interval of the model simulation for the given confidence level α ($0<\alpha<1$), two quantiles (1-α)/2*100% and (1+α)/2*100% are estimated from the $\hat{y}_{t,i}$. Following Shrestha and Solomatine (2006), these two quantiles will be called the lower prediction limit PL^L and the upper prediction limit PL^U:

$$\hat{Q}(p) = PL^L, \quad \text{where}, p = (1-\alpha)/2 \tag{4}$$

$$\hat{Q}(p) = PL^U, \quad \text{where}, p = (1+\alpha)/2 \tag{5}$$

The prediction interval $PI = [PL^L \; PL^U]$ is apportioned into two parts given the output $\bar{y}$ of the calibrated (optimal) model as:

$$PI^L = \bar{y} - PL^L, \quad PI^U = PL^U - \bar{y} \tag{6}$$

where PI^L is the distance between the model output and the lower prediction limit, PI^U is the distance between the model output and the lower prediction limit. PI^L and PI^U

are referred to lower and upper prediction intervals, respectively (although these are not intervals but distances).

3.5 ANN model V for estimation of prediction intervals (PIs)

To build the model V that maps the input data and the state variables to the prediction interval of the model output that is generated by MC simulations, ANN will be used. Experience suggests that the model residuals (errors) may show non-stationary bias, variance skewness and autocorrelation over one or more time steps (Beven and Freer, 2001). This characteristic of the model output distribution motivates us to build a statistical (machine learning) model to approximate not only the mean and the variance but also the prediction interval of the output.

Model V encapsulating the functional relationship between the input data $\boldsymbol{x}$ and the prediction interval PI will take the following form:

$$PI^L = V_L(X_V) + \xi_L = \widehat{PI}^L + \xi_L, \quad PI^U = V_U(X_V) + \xi_U = \widehat{PI}^U + \xi_U \tag{7}$$

where PI^L and PI^U are lower and upper prediction intervals computed from MC data; $\widehat{PI}^L$ and $\widehat{PI}^U$ are lower and upper prediction intervals estimated by ANN; ξ_L, ξ_U are the residual error in estimating the lower and upper prediction intervals, respectively.

Model V, after being trained, encapsulates the underlying dynamics of the uncertainty measures of the MC simulations and maps the input to these measures. The model V can be of various types, from linear to non-linear regression function such as an ANN. The choice of the model depends on the complexity of the problem to be handled and the availability of data. Once the model V is trained on the calibration data, it can be employed to estimate the uncertainty measures such as PIs for the new input data vector that was not used in the model building process. Once the PI^L and PI^U are estimated, the prediction interval of the model output is computed by adding the model output as follows:

$$PL^L = \bar{y} - \widehat{PI}^L, \quad PL^U = \bar{y} + \widehat{PI}^U \tag{8}$$

3.6 Selection of the input variables for model V

In order to train model V, data set X_V should be constructed on the basis of the set $D = \{\boldsymbol{x}_t, y_t\}$, (where, $\boldsymbol{x}_t$ = input data and y_t = observed data) of the hydrological model. Since the nature of models M and V is different, in most cases for choosing the adequate variables for X_V (possibly lagged), additional analysis of relationships between the output of V and the variables constituting D is needed. Such analysis is typically based on correlation and average mutual information. For example, if model M is a conceptual hydrological model, it would typically use rainfall (R_t) and evapotranspiration (E_t) as input variables to simulate the output variable runoff (Q_t). However, the uncertainty model V whose aim is to predict the probability distribution of error of the simulated runoff will be trained with the possible combination of rainfall and evapotranspiration or effective rainfall and, past values (lagged) of them including the lagged values of runoff.

3.7 Models performance indicators

The uncertainty model V can be validated in two ways: a) measuring its predictive capability; b) measuring the statistics of the uncertainty. The former approach measures the accuracy of uncertainty models in approximating the quantiles of the probability distribution of the model error generated by MC simulations. The later approach measures the goodness of the uncertainty models as uncertainty estimators.

Two performance measures such as coefficient of correlation (r) and the root mean squared error ($RMSE$) are used to measure the predictive capability of the uncertainty model. Beside these numerical measures, the graphical plots such as scatter and time plot of the quantile of the model error obtained from the MC simulation and their predicted values are used.

Goodness of the uncertainty models as uncertainty measures is evaluated by using the so-called prediction interval coverage probability ($PICP$) and mean prediction interval (MPI) (Shrestha and Solomatine, 2006). The $PICP$ measures the probability

that the observed values lies within the estimated PIs and it is estimated as:

$$PICP = \frac{1}{n}\sum_{t=1}^{n} C \tag{9}$$

$$\text{with } C = \begin{cases} 1, PL_t^L \le y_t \le PL_t^U; \\ 0, \text{otherwise}. \end{cases}$$

MPI estimates the average width of the PIs and gives an indication of how high is the uncertainty:

$$MPI = \frac{1}{n}\sum_{t=1}^{n}(PL_t^U - PL_t^L) \tag{10}$$

Theoretically, the value of *PICP* should be close to the prescribed degree of confidence. If there is no uncertainty, then the value of *MPI* is zero.

4 Application

4.1 Study area

The Brue catchment located in the South West of England, UK is selected for the application of the methodology. The catchment has a drainage area of 135 km^2 with the average annual rainfall of 867 mm and the average river flow of 1.92m^3/s, for the period from 1961 to 1990. The discharge is measured at Lovington. The hourly potential evapotranspiration was computed using the modified Penman method recommended by FAO (Allen et al., 1998). Splitting of available data set is based on Shrestha and Solomatine (2008), one year hourly data from 1994/06/24 05:00 to 06/24/1995 04:00 was selected for calibration of HBV hydrological model, running MC simulations to generate data for uncertainty model *V* and training model *V*. Data from 24/06/1995

05:00 to 31/05/1996 13:00 was used for the verification (testing) of the hydrological model, running MC simulations to generate data for validating uncertainty model V and validating model V. Each of the two data sets represents almost a full year of observations, and it appeared that the statistical properties of these sets are similar.

4.2 Conceptual hydrological model

A simplified version of HBV model (Fig. 2) was used to simulate river flows. Input data are observations of precipitation, air temperature, and estimates of potential evapotranspiration. The detailed description of the model can be found in Lindström et al. (1997).

4.3 Experimental setup

The nine parameters listed on Table 1 are used in HBV model. The model was first calibrated using the global optimization routine – adaptive cluster covering algorithm, ACCO (Solomatine et al., 1999) to find the best set of parameters, and then followed by manual adjustments of the parameters. The ranges of parameters for automatic calibration and parameter uncertainty analysis were set based on a range of calibrated values from other model applications (e.g., Braun and Renner, 1992) and the information about the catchment.

The model was calibrated using Nash and Sutcliff coefficient of efficiency (CE) as a performance measure of HBV model. For the calibration period CE was 0.96. The model was validated by simulating the flows for the independent validation data set, and CE was 0.83. Figure 3 shows the observed and simulated hydrograph in both calibration and verification period. The rainfall is also shown on the figure.

4.4 MC simulation and its convergence analysis

The parameters of HBV model are sampled from the uniform distribution with the ranges given in Table 1. The model is run for each random parameter set and like-

lihood measure is computed for each model run. CE is used as the basis to calculate likelihood. We investigated the number of behavioural samples retained out of 74 467 MC samples for different values of rejection threshold. It was observed that only 1/3 of simulations (25 000 samples) are accepted for threshold value of 0, whereas less than 1/10 of simulations are retained for a threshold value of 0.7.

We have also tested the convergence of MC simulations to know the number of samples required to get the reliable results. Mean and standard deviation of CE were used to analyse the convergence of MC simulations (Eq. 11 and 12). Other statistics for convergence test can be found in Ballio and Guadagnini (2004).

$$ME_k = \frac{1}{k}\sum_{i=1}^{k}(CE_i) \tag{11}$$

$$SDE_k = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(CE_i - ME_k)^2} \tag{12}$$

where CE_i is the coefficient of model efficiency of i^{th} MC run, ME_k and SDE_k are the mean and standard deviation of efficiency of model runs upto k^{th} runs, respectively. The Fig. 4 depicts the two statistics – mean and standard deviation of CE – used to analyse the convergence of MC simulations. It is observed that both statistics are stable after 5000–10 000 simulations, so 10 000 MC simulations are reasonable to consider in this case study.

4.5 Sensitivity of parameters

The interaction between parameter values results in the broad region of acceptable simulation (corresponding to the different parameter values) and it is the combination of parameter values that produces the acceptable or non-acceptable simulations within the chosen model structure. Such result reveals little about the sensitivity of the model

predictions to the individual parameters, except where some strong change in the likelihood measure is observed in a certain range of a particular parameter. More detailed analysis based on the extension of regional sensitivity analysis (Freer et al., 1996; Spear and Hornberger, 1980) was performed. Figure 5 shows the cumulative distributions for 10 equal sets (by number) of the MC simulations. Each parameter population is ranked from best to worst in terms of the chosen likelihood function and the ranked population is then divided into ten bins of equal size. The cumulative likelihood distribution of each group is then plotted. Parameter sensitivity can be evaluated by assessing the spread of cumulative distribution function of each group is then plotted. Parameters with the strong deviation of the distribution function are recognized as sensitive parameters. Three parameters *ALFA*, *K* and *MAXBAS* were found to be the most sensitive parameters, while the two parameters *K*4 and *CFLUX* shows no sensitive at all. The rest of the parameters show the moderate sensitivity.

4.6 ANN for emulating MC simulation

Once having the uncertainty results generated by MC simulation, ANN is trained to learn functional relationship between the uncertainty results and the input data. The GLUE method has been used for parameter uncertainty estimation of HBV model. The threshold value of 0.0 (measure by *CE*) is selected to classify simulation either behavioural or non-behavioural. 90% uncertainty bounds are calculated using the 5% and 95% quantiles of the MC simulation realizations.

Corresponding 90% lower and upper PIs are calculated using the model output simulated by optimal model parameter sets found by Shrestha and Solomatine (2008). Hence the computed upper and lower PIs are conditional on contemporary value of the model simulation. Next step was to select the most relevant input variables to build predictive ANN model.

4.6.1 Selection of input variables

To select the input variables several approaches can be used (see, e.g., Solomatine and Dulal, 2003; Guyon and Elisseeff, 2003; Bowden et al., 2005). The input variables for model V are constructed from the rainfall, evapotranspiration and observed discharge. Experimental results show that evapotranspiration alone does not have significant influence on the prediction interval. Thus it was decided not to include the evapotranspiration as a separate variable, but rather to use effective rainfall (rainfall minus evapotranspiration for rainfall greater than evapotranspiration and zero otherwise).

Figure 6 shows the correlation coefficient and the average mutual information (AMI) of R_t and its lagged variables with the lower and upper PIs. It is observed that the correlation coefficient is minimum at the zero hour lag time and increases with the lags upto 9 h (Fig. 6a) for lower PI. While the optimal lag time (time at which the correlation coefficient and/or AMI is maximum) is 7 h in the case of upper PI (Fig. 6b). At this optimal lag time, information about the PIs is contained maximally in the variable R_t. Such findings are also supported by the AMI analysis. Additionally, correlation and AMI between the PIs and observed discharge were analysed. The results show that the immediate and the recent discharges (with the lag of 0, 1, 2) have very high correlation with the PIs. So it was also decided to use the observed discharge as the input to the model V.

Several structures of the input data including lagged variables were considered. The principle of parsimony was followed to avoid the use of a large number of inputs, so the aggregates, such as the moving averages or derivatives of the inputs that would have hydrological meaning were considered. Typically, the rainfall depth at hourly time step partially exhibits random behaviour, and is not very representative of the rainfall phenomenon during a short period. Hence, we used the mean rainfall value of the lagged variables R_{t-5},R_{t-6},R_{t-7},R_{t-8}, and R_{t-9} as the rainfall input, which is denoted as R_{t-9a}. Furthermore, derivative of the flow indicates whether the flow situation is either normal or base flow (zero or small derivative), or can be characterized as the rising

HESSD
6, 1677–1706, 2009

limb of the flood event (high positive derivative), or the recession limb (high negative derivative. Therefore, in addition to the flow variable Q_{t-1}, rate of change of flow at time $t-1$ is computed by deducting the flow at time $t-2$ from flow at $t-1$ and denoted by ΔQ_{t-1}.

The ANN model is based on 8745 data records from 24/06/1994 05:00 to 24/06/1995 04:00 (hourly) and its structure is given by:

$$\mathrm{PI} = V(R_{t-9a}, Q_{t-1}, \Delta Q_{t-1}) \tag{13}$$

where, PI is prediction intervals PI^L or PI^U; R_{t-9a} is moving average of R_{t-5},R_{t-6},R_{t-7},R_{t-8}, R_{t-9}; and ΔQ_{t-1} is $Q_{t-1} - Q_{t-2}$ (characterizes the derivative of previous discharge).

4.6.2 Model training

The same data set used for calibration and verification of HBV model were used for training and verification of model V respectively. However, for proper training of the ANN model the calibration data set is segmented into two sets; 15% of data sets for cross validation (CV) and 85% for training.

CV data set was used to identify the best structure of ANN. In this paper, a multilayer perceptron network was used; optimization was performed by the Levenberg-Marquardt algorithm. The hyperbolic tangent function was used for the hidden layer with linear transfer function at the output layer. The maximum number of epoch was fixed to 1000. Trial and error method is adopted to detect the optimal number of neurons in the hidden layer, testing a number of neurons from 1 to 10. It was observed that six neurons give the lowest error on CV set.

5 Results and discussions

Figure 7 shows the scatter plot of observed and simulated discharge in verification period. For many data points HBV model is quite accurate but its error (uncertainty) is quite high during the peak flows. This can be explained by the fact that the version of the HBV model used in this study is the lumped model and one cannot expect high accuracy from it.

ANN-based uncertainty model V trained on the data generated by MC simulations, was tested on the verification data set; its performance is shown in Fig. 8. Figure 9 shows the hydrograph with the 90% uncertainty bounds predicted by ANN together with the MC simulation uncertainty bounds in the verification period. It can be said that ANN reproduces the MC simulations uncertainty bounds reasonably well, in spite of the low correlation of the input variables with the PIs. Although some errors can be noticed, the predicted uncertainty bounds follow the general trend of MC uncertainty bounds. Noticeably the model fails to capture the observed flow during one of the peak events (bottom left figure). Note however, that the results of ANN model and MC simulations are visually closer to each other than both of them to the observed data.

Detailed analysis reveals that estimated uncertainty bounds contain 77.00% (*PICP*) of the observed runoffs, which is very close to the MC simulation result (77.24%). The average width of prediction intervals (*MPI*) estimated by ANN is narrower (1.93 m^3/s) as compared to the value obtained with MC simulations (2.09 m^3/s). Further analysis of the results reveals that 14.74% of the observed data are below the lower uncertainty bounds whereas 8.01% of data are above the upper bounds.

The predictive capability of ANN model in estimating lower and upper PIs are compared in both calibration and verification period and it appears that the correlation coefficient and *RMSE* for PI^L is higher than those of PI^U. This can be explained by the fact that PI^U corresponds to the higher values of flow (where the HBV model is less accurate) and has higher variability, which makes its prediction a difficult task.

6 Conclusions

This paper presents a method to replicate the results of Monte Carlo simulations in the form of a predictive ANN model. The method is computationally efficient and can be used in real time application when the large number of model runs required, and is applicable to various kinds of hydrological models.

The ANN models is first trained on the data generated by MC simulations to encapsulate the relationship between the hydrometeorological variables and the characteristics of the model output probability distribution (prediction interval), and then the trained models are used to estimate the prediction interval for the new input data. It is worth mentioning that MC simulations are done off-line only to generate the data to train the ANN, while the trained ANN models are employed to estimate the uncertainty in real time application without running the MC simulations any more.

In this study, two separate ANN models are used to estimate the two quantiles (5% and 95%) forming the 90% prediction interval. However the methodology can be extended to predict several quantiles of the model outputs, that is, in fact, estimating the shape of the probability distribution of the model output generated by MC simulations. The conceptual hydrological model HBV was applied to the Brue catchment in United Kingdom and used as a case study. The results demonstrate that the prediction of uncertainty with ANN generate interpretable uncertainty estimates, and this is an indicator that the presented method can be a valuable tool for assessing uncertainty of various predictive models. The proposed method can be used to replicate the results of various versions of the MC methods, e.g. Markov Chain Monte Carlo and Latin hypercube sampling. Furthermore, this method can be applied in the context of other sources of uncertainty – input, structure, or combined.

Further studies aim at testing other machine learning techniques (possibly including instance-based learning), and applying the presented methodology to other hydrological (process) models in various case studies.

Acknowledgements. This work was partly supported by the European Communitys Sixth

Framework Program through the grant to the budget of the Integrated Project FLOODsite, Contract GOCE-CT-2004-505420, and by the Delft Cluster Research Programme of the Dutch Government (project “Safety against flooding”)

References

Abebe, A. and Price, R.: Managing uncertainty in hydrological models using complementary models, Hydrol. Sci. J., 48, 679–692, 2003. 1681

Abebe, A. and Price, R.: Forecasting surge prediction accuracy on the dutch coast using Neural Nwtworks, in: Proceedings of the 6th International Conference on Hydroinformatics: Singapore, 21–24 June 2004, World Scientific, 2004. 1679, 1680

Abrahart, R. and See, L.: Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments, Hydrol. Proc., 14, 2157–2172, 2000. 1680

Allen, G., Pereisa, S., Raes, D., and Smith, M.: Crop evapotraspiration, Guideline for computing crop water requirements-FAO, Irrigation and drainage paper, 56, 1998. 1686

Ballio, F. and Guadagnini, A.: Convergence assessment of numerical Monte Carlo simulations in groundwater hydrology., Water Resour. Res., 40, W04603, doi:10.1029/2003WR002876, 2004. 1679, 1688

Bardossy, A., Duckstein, L., and Bogardi, I.: Fuzzy rule-based classification of atmospheric circulation patterns, Int. J. Climat., 15, 1087–1097, 1995. 1680

Beven, K. and Binley, A.: Future of distributed models: Model calibration and uncertainty prediction., Hydrol. Proc., 6, 279–298, 1992. 1678

Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, J. Hydrol., 249, 11–29, 2001. 1684

Bishop, C.: Neural Networks for Pattern Recognition. Clarendon-Press, 1995. 1680

Bowden, G., Dandy, G., and Maier, H.: Input determination for neural network models in water resources applications. Part 1background and methodology, J. Hydrol., 301, 75–92, 2005. 1690

Braun, L. and Renner, C.: Application of a Conceptual Runoff Model in Different Physiographic Regions of Switzerland, Hydrol. Sci. J., 37, 217–232, 1992. 1687

Dawson, C. and Wilby, R.: Hydrological modelling using artificial neural networks, Prog. Phys. Geogr., 25, 80, 2001. 1680, 1681
Dibike, Y., Velickov, S., Solomatine, D., and Abbott, M.: Model Induction with Support Vector Machines: Introduction and Applications, J. Comp. Civil Eng., 15, 208–216, 2001. 1680
Freer, J., Beven, K., and Ambroise, B.: Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach, Water Resources Res., 32, 2161–2173, 1996. 1689
Govindaraju, R. and Rao, A.: Artificial Neural Networks in Hydrology, Kluwer Academic Publishers, 348 pp., 2000. 1680
Guyon, I. and Elisseeff, A.: An introduction to variable and feature selection, J. Mach. Learn. Res., 3, 1157–1182, 2003. 1690
Harr, M.: Probabilistic estimates for multivariate analyses, Appl. Math. Model., 13, 313–318, 1989. 1679
Haykin, S.: Neural Networks: A Comprehensive Foundation. Prentice-Hall, Upper Saddle River, NJ, USA, 1999. 1680
Klir, G. J. and Yuan, B.: Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: selected papers by: Zadeh, L. A., World Scientific Publishing Co. Inc., River Edge, NJ, USA, 840 pp., 1996. 1680
Kunstmann, H., Kinzelbach, W., and Siegfried, T.: Conditional first-order second-moment method and its application to the quantification of uncertainty in groundwater modeling, Water Resour. Res., 38, 6–6, 2002. 1679
Lindström, G., Johansson, B., Persson, M., Gardelin, M., and Bergström, S.: Development and test of the distributed HBV-96 hydrological model, J. Hydrol., 201, 272–288, 1997. 1687, 1699
Maier, H. and Dandy, G.: Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications, Environ. Model. Software, 15, 101–124, 2000. 1680
Mantovan, P. and Todini, E.: Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology, J. Hydrol., 330, 368–381, 2006. 1679
McKay, M., Conover, W., and Beckman, R.: A comparison of three methods for selecting values of input variables in the analysis ofoutput froma computer code, Technometrics, 21, 239–245, 1979. 1679
Melching, C.: Reliability Estimation in Computer Models of Watershed Hydrology, Water Res. Publ., 69–118, 1995. 1679

Minns, A. and Hall, M.: Artificial neural networks as rainfall-runoff models, Hydrol. Sci. J., 41, 399–417, 1996. 1680
Nash, J. and Sutcliff, J.: River flow forcasting through conceptual models – Part I- A discussion of principles, J. Hydrol., 10, 282–290, 1970. 1687
Protopapas, A. and Bras, R.: Uncertainty Propagation With Numerical Models for Flow and Transport in the Unsaturated Zone, Water Resour. Res., 26, 2463–2474, 1990. 1679
Rosenblueth, E.: Two-point estimates in probability, Appl. Math. Model., 5, 329–335, 1981. 1679
Rossel, R. A. V., Goovaerts, P., and McBratney, A. B.: Assessment of the production and economic risks of site-specific liming using geostatistical uncertainty modeling, Environmetrics, 12(8), 699–711, 2001 1679
Shrestha, D. and Solomatine, D.: Machine learning approaches for estimation of prediction interval for the model output, Neural Networks, 19, 225–235, 2006. 1679, 1680, 1685
Shrestha, D. and Solomatine, D.: Data-driven approaches for estimating uncertainty in rainfall-runoff modelling, International Journal of River Basin Management, 6, 109–122, 2008. 1679, 1680, 1686, 1689
Solomatine, D. and Shrestha, D.: A novel method to estimate model uncertainty using machine learning techniques, Water research resources, 45, W00B11, doi:10.1029/2008WR006839, 2009. 1680
Solomatine, D. and Dulal, K.: Model trees as an alternative to neural networks in rainfall–runoff modelling/Arbres de modèles comme alternative aux réseaux de neurones en modélisation pluie–débit, Hydrological Sciences Journal/Journal des Sciences Hydrologiques, 48, 399–411, 2003. 1680, 1690
Solomatine, D., Dibike, Y., and Kukuric, N.: Automatic calibration of groundwater models using global optimization techniques, Hydrol. Sci. J., 44, 879–894, 1999. 1687
Spear, R. and Hornberger, G.: Eutrophication in Peel Inlet-II. Identification of Critical Uncertainties Via Generalized Sensitivity Analysis, Water Res., 14, 43–49, 1980. 1689

Table 1. The uniform prior interval of HBV model parameters are used for calibration of the model and for analysis of parameter uncertainty by MC simulations method.

Parameter	Description	Uniform prior range	Optimum value
FC	Maximum soil moisture content	100–300	160.33
LP	Limit for potential evapotranspiration	0.5–0.99	0.527
ALFA	Response box parameter	0–4	1.54
BETA	Exponential parameter in soil routine	0.9–2	1.963
K	Recession coefficient for upper tank	0.0005–0.1	0.001
K4	Recession coefficient for lower tank	0.0001–0.005	0.004
PERC	Percolation from upper to lower response box	0.01–0.09	0.089
CFLUX	Maximum value of capillary flow	0.01–0.05	0.038
MAXBAS	Transfer function parameter	8–15	12.00

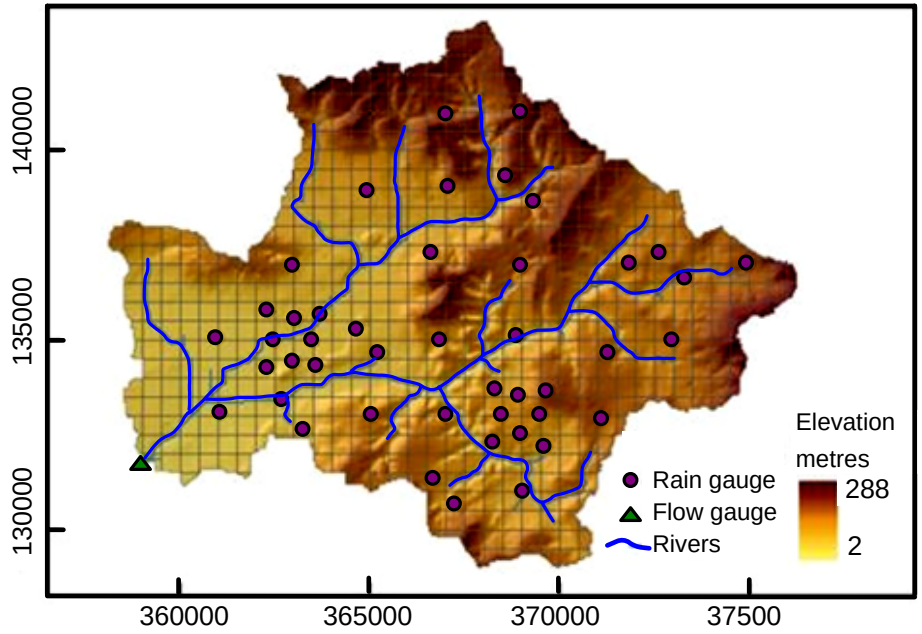

Fig. 1. The Brue catchment showing (the horizontal and vertical axes refer to the easting and northing in British national grid reference co-ordinates).

HESSD
6, 1677–1706, 2009

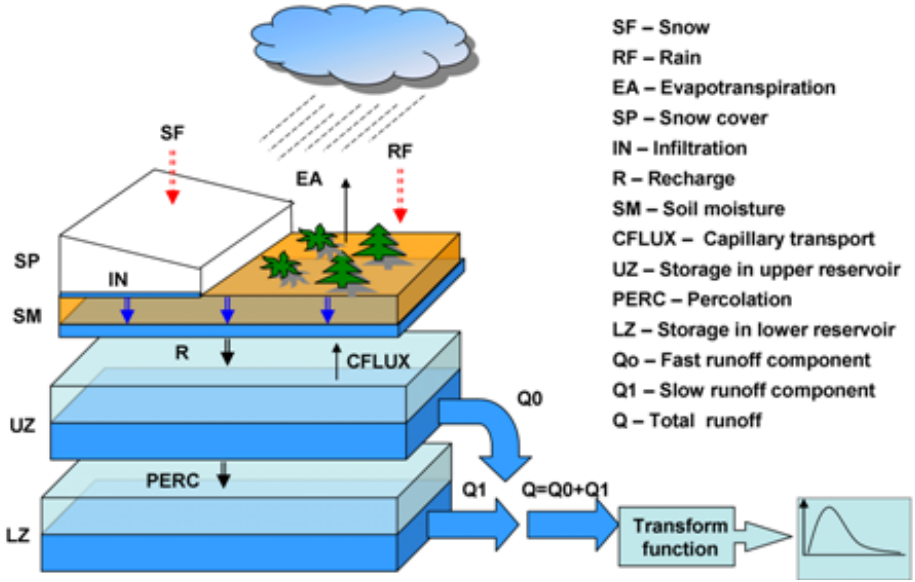

Fig. 2. Schematic representation of HBV-96 model (Lindström et al., 1997) with routine for snow (upper), soil (middle) and response (bottom)

HESSD
6, 1677–1706, 2009

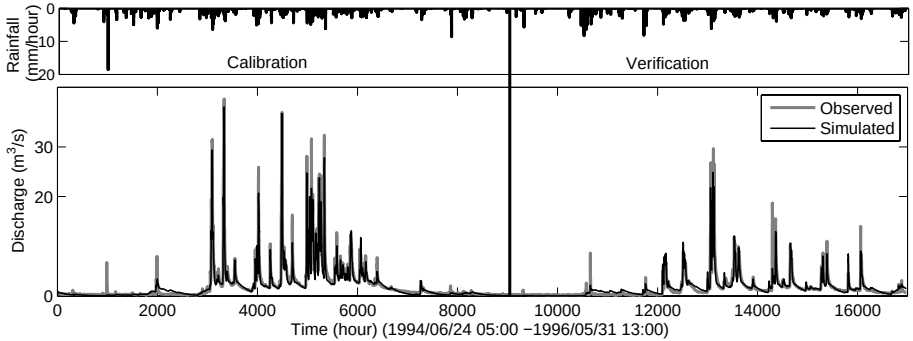

Fig. 3. Observed and simulated discharge hydrograph.

HESSD
6, 1677–1706, 2009

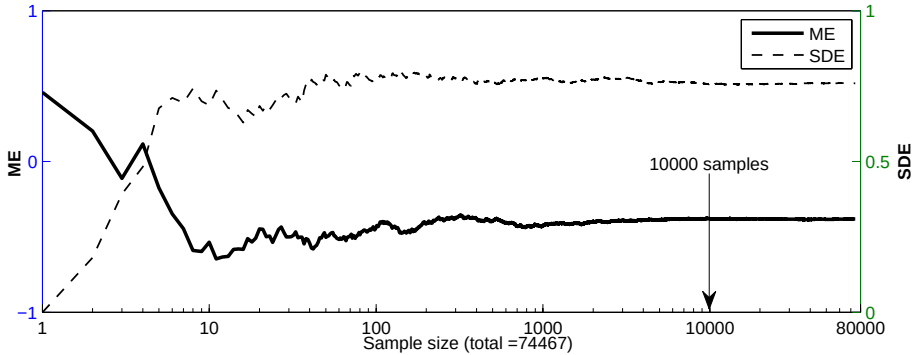

Fig. 4. The convergence of mean and standard deviation of the coefficient of model efficiency. Note that x-axis is log scale to see initial variation.

HESSD
6, 1677–1706, 2009

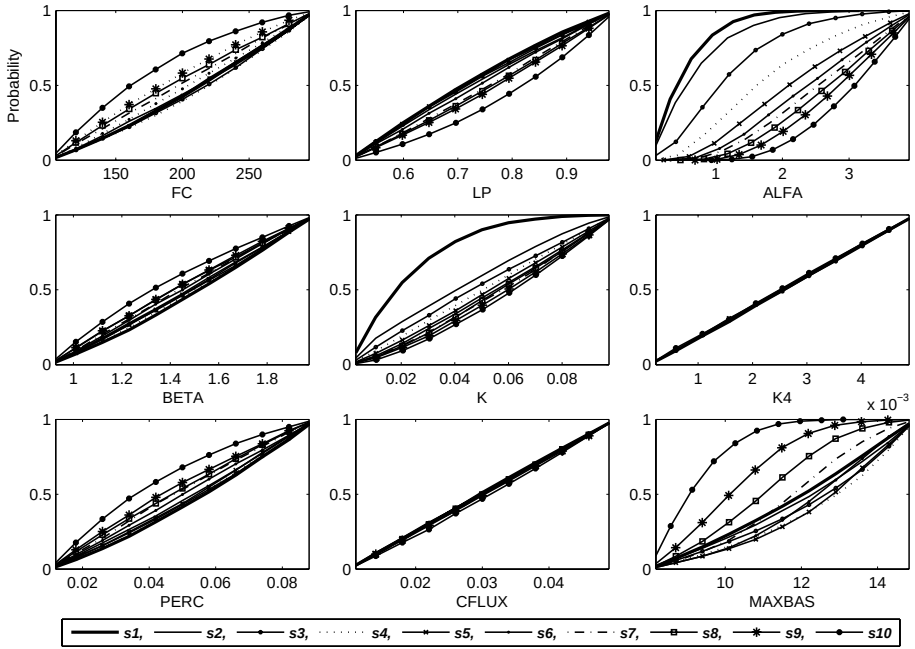

Fig. 5. Sensitivity of individual parameters expressed as cumulative distribution values in 10 equal sets *s*1 through *s*10 (by number) of MC simulations and each line represents cumulative likelihood distribution of each set from best (*s*1) to worst (*s*10) in term of the chosen likelihood. Parameters with strong deviation of the distribution function are recognized as sensitive parameters (*ALFA*, *K* and *MAXBAS*).

HESSD
6, 1677–1706, 2009

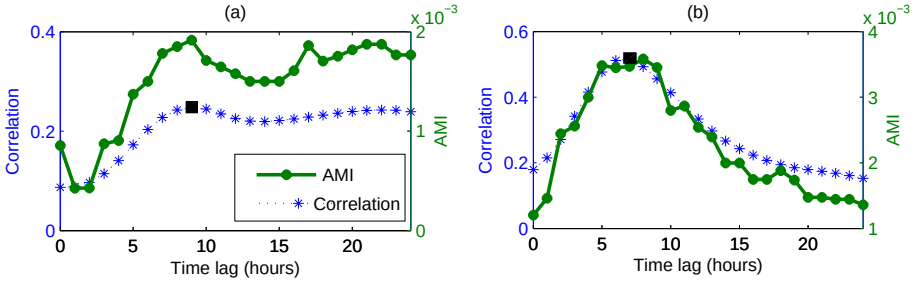

Fig. 6. Linear correlation and AMI between rainfall **(a)** lower prediction intervals; and **(b)** upper prediction intervals for different lags time. Black squares show the maximum correlation.

HESSD
6, 1677–1706, 2009

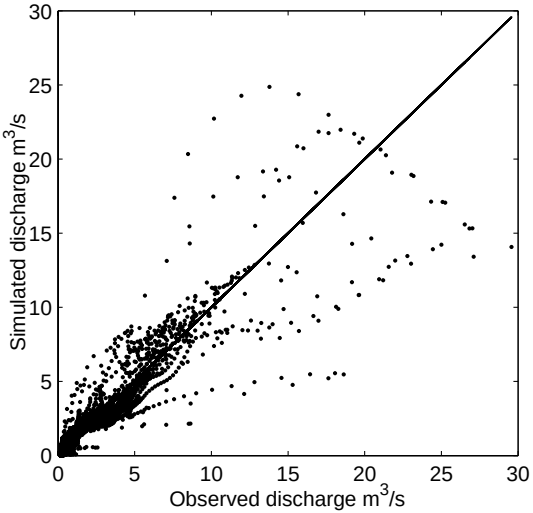

Fig. 7. Scatter plot of observed and simulated discharge from HBV model.

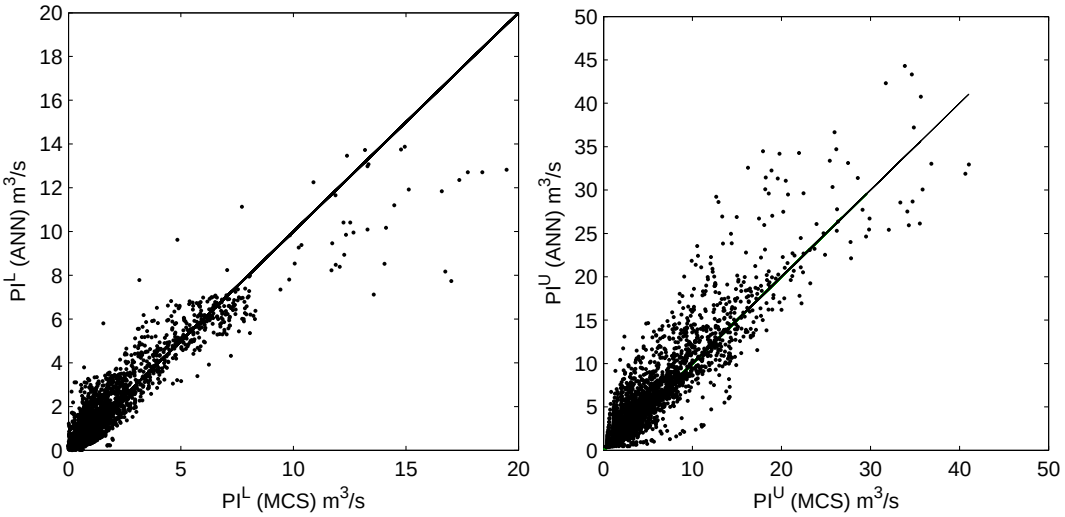

Fig. 8. Scatter plots showing the performance of the ANN-based uncertainty model in the verification data set. x-axes show the perdiction intervals obtained by MC simulations and y-axes show the prediction intervals estimated by ANN.

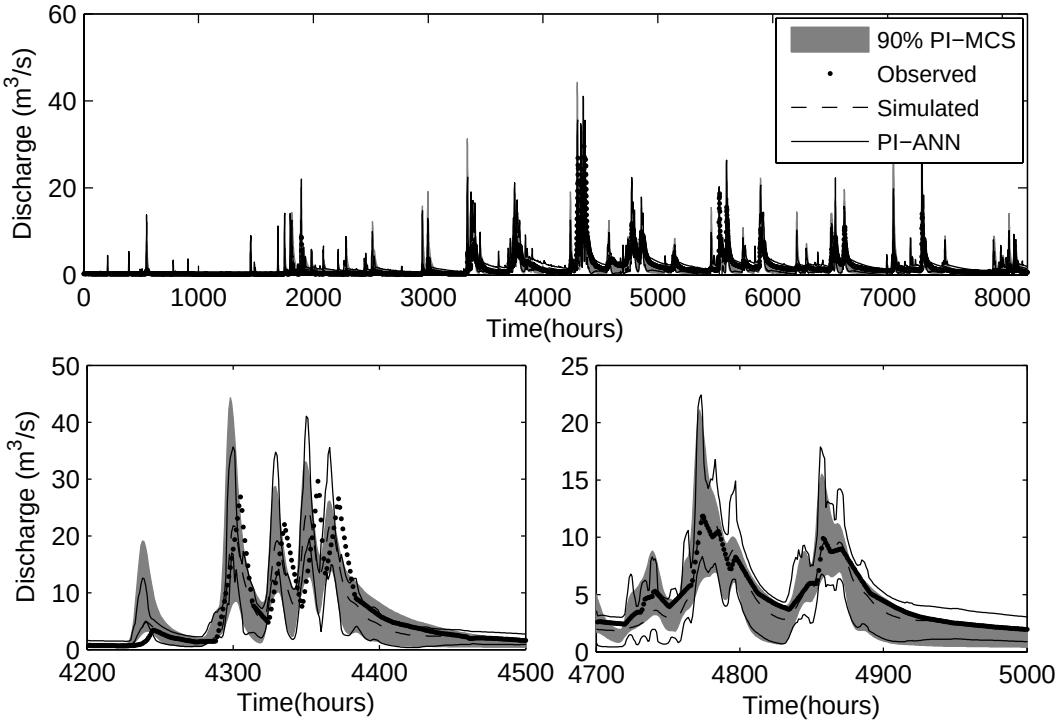

Fig. 9. Hydrograph of 90% prediction bounds estimated by MC simulation and ANN in verification period (24/06/1995 05:00 to 31/05/1996 13:00), the black dot indicates observed discharges and the dark grey shaded area denotes the prediction uncertainty that results from MC simulation. The black line denotes the prediction uncertainty estimated by ANN and dash line indicates simulated discharges from optimal model parameters.