

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

D. L. Shrestha¹, N. Kayastha², and D. P. Solomatine^{1,3}

¹UNESCO-IHE Institute for Water Education, Delft, The Netherlands

²MULTI Disciplinary Consultants Ltd, Kathmandu, Nepal

³Water Resources Section, Delft University of Technology, The Netherlands

Received: 17 February 2009 – Published in Hydrol. Earth Syst. Sci. Discuss.: 4 March 2009

Revised: 1 July 2009 – Accepted: 7 July 2009 – Published: 21 July 2009

Abstract. In this study, a methodology has been developed to emulate a time consuming Monte Carlo (MC) simulation by using an Artificial Neural Network (ANN) for the 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 MC 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 the MC simulation. The method is applied to estimate the parameter uncertainty of a lumped conceptual hydrological model, HBV, for the Brue catchment in the United Kingdom. The results are quite promising as the prediction intervals estimated by the 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.

methods have been developed to propagate the uncertainty through the hydrological models and to derive meaningful uncertainty bounds of the model simulations. These methods range from analytical and approximation methods (see, e.g., Tung, 1996; Melching, 1992) to Bayesian and Monte Carlo (MC) sampling based methods (e.g., Kuczera and Parent, 1998; Beven and Binley, 1992; Vrugt et al., 2003; Thieman et al., 2001), methods based on the analysis of model errors (e.g., Montanari and Brath, 2004; Shrestha and Solomatine, 2008; Solomatine and Shrestha, 2009), and methods based on fuzzy set theory (see, e.g., Maskey et al., 2004). The majority of these methods deal only with a single source of uncertainty and consider model uncertainty to be mostly produced by parameter uncertainty assuming that the model structure is correct and the input data is free from errors. Only recently new techniques have been emerging such as data assimilation techniques (Vrugt et al., 2005; Moradkhani et al., 2005), multi model averaging techniques (see, e.g., Ajami et al., 2007; Vrugt and Robinson, 2007; Georgakakos et al., 2004), Bayesian approaches (Kavetski et al., 2006; Kuczera et al., 2006), and efficient Markov chain Monte Carlo (MCMC) techniques (Haario et al., 2006; Vrugt et al., 2008) to explicitly treat two or more sources of uncertainty such as input, parameter and structure uncertainty.

MC simulation technique is a widely used method for uncertainty analysis in hydrological modelling and allows the quantification of the model output uncertainty resulting from uncertain model parameters, input data or model structure. The approach involves random sampling from the distributions of the uncertain inputs and the model is run successively until a desired statistically significant distribution of outputs is obtained. The main advantage of the MC

1 Introduction

The uncertainty analysis of hydrological models in recent years received special attention. Several uncertainty analysis



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

simulation based uncertainty analysis is its general applicability; however, methods of this type require a large number of samples (or model runs), so their applicability may be limited to simple models. In the 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). GLUE is one of the popular methods for analysing parameter uncertainty in hydrological modelling and has been widely used over the past ten years to analyse and estimate predictive uncertainty, particularly in hydrological applications (see, e.g., Freer et al., 1996; Beven and Freer, 2001; Montanari, 2005). Users of GLUE are attracted by its simple understandable ideas, relative ease of implementation and use, and its ability to handle different error structures and models without major modifications to the method itself. Despite its popularity, there are theoretical and practical issues related with the GLUE method as is reported in the literature. For instance, Mantovan and Todini (2006) argue that GLUE is inconsistent with the Bayesian inference processes such that it leads to an overestimation of uncertainty, both for the parameter and the predictive uncertainty estimation. For issues regarding inconsistency and other criteria, readers are referred to the citation above and the subsequent discussions in the Journal of Hydrology in 2007 and 2008. A practical problem with the GLUE method is that, for models with a large number of parameters, the sample size from the respective parameter distributions must be very large in order to achieve a reliable estimate of model uncertainties (Kuczera and Parent, 1998). In order to reduce the number of samples (i.e., model runs) a surrogate model of the model response surface can be applied (see, e.g., Khu and Werner, 2003). Blasone et al. (2008) used adaptive MCMC sampling within the GLUE methodology to improve the sampling of the high probability density region of the parameter space. Another practical issue is that in many cases the percentage of observations falling within the prediction limits provided by GLUE is much lower than the given confidence level used to produce these prediction limits (see, e.g., Montanari, 2005). Xiong and O'Connor (2008) modified the GLUE method to improve the efficiency of the GLUE prediction limits in enveloping the observed discharge.

The MC based method for uncertainty analysis of the outputs of hydrological models is very flexible, conceptually simple and straightforward, but becomes impractical in real time applications when there is little time to perform the uncertainty analysis because of the large number of model runs required. For such situations alternative approximation methods have been developed, and referred to as moment propagation techniques, which are able to calculate directly the first and second moments without the application of an MC simulation (see, e.g., Rosenblueth, 1981; Harr, 1989; Protopapas and Bras, 1990; Melching, 1995; Kunst-

mann et al., 2002). A number of research studies have been conducted to reduce the number of MC simulation runs effectively, for instance, Latin Hypercube sampling (see, e.g., McKay et al., 1979), and, more recently, the delayed rejection adaptive Metropolis method (Haario et al., 2006), and the differential evolution adaptive Metropolis method, DREAM (Vrugt et al., 2008). In these Metropolis algorithm-based uncertainty analysis methods, the comparison of the statistics of multiple sample chains in parallel would provide a formal solution to assess how many model runs are required to reach convergence and obtain stable statistics of the model output and parameters. In spite of the improved efficiency of the MC methods, it is well recognized that traditional MC based simulation still lacks a well-established convergence criterion to terminate the simulations at a desired level of accuracy (e.g., Ballio and Guadagnini, 2004).

We propose to use artificial neural networks (ANNs) to emulate the uncertainty 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 to as V) encapsulating the information about the realizations of the process (e.g., hydrological) model M output generated by the MC simulations. Instead of predicting a single value of the model error, as done in the most error correction procedures, we aim to predict the distribution of the output of M generated by the 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 newly observed input data is fed into M , while the MC based uncertainty analysis methods require a fresh set of the MC runs for each analysis. For instance, GLUE will typically require a fresh set of the MC runs from the behavioural models to produce the prediction intervals (PIs) of the model output for each time step with the new data input.

The proposed technique to emulate the complex model by a simple model is an example of surrogate modelling, or meta-modelling - an approach widely used when running the complex model could be computationally expensive. For example, O'Hagan (2006) used the Gaussian process emulator to emulate a complex simulation model. Li et al. (2006) introduced an approach to meta-modelling whereby a sequential technique is used to construct and simultaneously update mutually dependent meta-models for multi-response, high-fidelity deterministic simulations. Young and Ratto (2008) proposed a dynamic emulation model to emulate a complex high order model by a low order data based mechanistic model. The novelty of our approach can be described as:

1. Our method explicitly builds an emulator of the MC uncertainty results rather than of a single simulation model.
2. Our emulator is based on a machine learning technique, while other techniques are Bayesian (e.g., O'Hagan, 2006), use non-linear differential equations (e.g., the data based mechanistic model of Young, 1998).

In this study, we consider only parameter uncertainty of the process model and employ the GLUE method to analyse it. It is worthwhile to note that our method can be used with other uncertainty analysis methods and can be applied for analysing other sources of uncertainty as well. The approach is tested on estimating two quantiles of the MC realizations of the process model. Thus, the model parameter uncertainty is measured by only two quantiles of the probability distribution that constitutes the PIs of the model output corresponding to some confidence level (e.g. 90%). The HBV hydrological model of the Brue catchment in the United Kingdom is used as a case study. An ANN is used as a surrogate model to estimate the PIs of the HBV model outputs.

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.

ANNs have been extensively used in hydrological modelling in past ten years, particularly in rainfall-runoff 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 systems (Vernieuwe et al., 2005; Jacquin and Shamseldin, 2006; Klir and Yuan, 1996), model trees (Solomatine and Dulal, 2003), and support vector machines (Dibike et al., 2001). However, the application of such techniques to estimate the uncertainty of physically based or data-driven hydrological models is very limited. Abebe and Price (2004) used an 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 an 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 rainfall-runoff model to various case studies. More information about ANNs can be found in Bishop (1995) and Haykin (1999), and the overview of ANN applications within hydrology can be found in Maier

and Dandy (2000), and Dawson and Wilby (2001). In this paper, however, ANNs are not used for hydrological prediction, but to encapsulate the data generated by MC simulations.

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 the uncertainty associated with the prediction of the hydrological variables, e.g. runoff, has similar magnitude for similar hydrological conditions. By the hydrological conditions we understand here the combination of the state of the input data and the state variables, which are forcing or driving the generation of the runoff in the catchment. For example, one can see that, compared to the low flows, it is more difficult to predict extreme events such as peak flows. Consequently, uncertainty of the prediction of the peak flow is higher compared to those for low flows.

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 the MC simulations). Thus, our method allows for predicting the uncertainty bounds of the model prediction without running the MC simulations in a real time application. Indeed, MC simulations are emulated by ANN models.

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 \mathbf{x} , the initial condition of the state variables s_0 and the vector of the parameters θ . The model M could be physically based, conceptual, or even data driven. In this paper we assumed the model M is a conceptual hydrological model. The system response can be represented by:

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

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 \mathbf{x} , initial conditions s_0 , 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 \mathbf{x} or the parameters vectors or even the structure of the model, or a combination of them. In this paper we assume that the model structure and the input data are certain (correct), so mathematically this can be expressed as:

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

where θ_i is the set of parameters sampled for the i^{th} run of the MC simulation, $\hat{y}_{t,i}$ is the model output of the t^{th} time step for the 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 way to judge the uncertainty of the model output is to use the error variance: a large variance of the model error typically indicates that the model prediction is uncertain. In most cases, however, the variance does not sufficiently describe the uncertainty, and more informative quantities such as prediction intervals are used.

A prediction interval (PI) 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, the model output is given a different weight (to be defined later), so a quantile can be found using the following equation:

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

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 $M(\mathbf{x}, \theta_i)$ at simulation i , $\hat{Q}(p)$ is p^{th} [0, 1] 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, the model structure, and the weight vector w_i . In the GLUE method, w_i is the likelihood weight (see Sect. 4.4).

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

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

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

The PI is apportioned into two parts given the output \bar{y} of the calibrated (optimal) model as:

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

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 upper prediction limit. Following Shrestha and Solomatine (2006), PI^L and PI^U represent the lower and upper PI, respectively (although these are not intervals but distances).

3.5 ANN models V for estimation of prediction intervals

To build the model V that maps the input data and the state variables to the PIs of the model output that is generated by the MC simulations, an ANN will be used. Experience suggests that the model residuals (errors) may show a 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 non-linear regression (ANN) model to approximate not only the mean and the variance, but also the prediction interval of the output.

The model V encapsulating the functional relationship between the input data \mathbf{x} and the lower and upper PI takes the form:

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

where PI^L and PI^U are the lower and upper PI computed from the MC simulations data; $V_L(X_V)$ and $V_U(X_V)$ are lower and upper PI estimated by ANN; ξ_L , ξ_U are the residual error in estimating the lower and upper PI, respectively.

Model V , after being trained, encapsulates the underlying dynamics of the uncertainty descriptors (in this case lower and upper PI) of the MC simulations and maps the input data to these descriptors. The model V can be of various types, from linear to non-linear regression models such as an ANN. The choice of 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 descriptors such as PIs for the new input data vector that was not used in any of the model building process. Once the PI^L and PI^U are estimated, the prediction limits of the model output is given by rearranging Eq. (6):

$$PL^L = \bar{y} - V_L(X_V), \quad PL^U = \bar{y} + V_U(X_V) \quad (8)$$

3.6 Selection of the input variables for model V

In order to train the model V , data set X_V should be constructed on the basis of the set $D = \{\mathbf{x}_t, y_t\}$, $t = 1, 2, \dots, n$, (where, \mathbf{x}_t is the input data vector, y_t is the observed data, n is the number of data) of the hydrological model. Since the natures of models M and V are different, in most cases for choosing the relevant input variables for X_V additional analysis of relationships between the PIs and the variables constituting D is needed. Such analysis is typically based on correlation and average mutual information (AMI). While correlation analysis is used to find the linear relationship between

the variables, mutual information analysis is used to determine linear or non-linear the dependencies between the variables. The mutual information is a measure of information available from one set of data having knowledge of another set of data. The average mutual information AMI (Fraser and Swinney, 1986) between two variables X and Y is given by

$$\text{AMI} = \sum_{i,j} P_{XY}(x_i, y_j) \log_2 \left[\frac{P_{XY}(x_i, y_j)}{P_X(x_i)P_Y(y_j)} \right] \quad (9)$$

where $P_X(x)$ and $P_Y(y)$ are the marginal probability density functions of X and Y , respectively, and $P_{XY}(x, y)$ is the joint probability density functions of X and Y . A high value of the AMI would indicate a strong dependence between two variables.

For example, if the model M is a conceptual hydrological model, it would typically use rainfall and evapotranspiration as input variables to simulate the output variable runoff. However, the uncertainty model V whose aim is to predict the probability distribution of the MC simulations will be trained with the possible combination of the lagged variables (i.e. past values) of rainfall and evapotranspiration or effective rainfall and runoff. It is worthwhile to mention that the input data structure may not be the same for the models V^L and V^U of the lower and upper PI, respectively.

3.7 Models performance indicators

The uncertainty model V can be validated in two ways: a) by measuring its predictive capability; and b) by 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 output generated by the MC simulations. The latter approach measures the goodness of the uncertainty models as uncertainty estimators.

Two performance measures such as the coefficient of correlation, CoC 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 hydrograph plot of the quantile of the model output obtained from the MC simulation and their predicted values can be analysed to judge the accuracy of the predictive uncertainty model V (in this study ANN model).

The goodness of the uncertainty models is evaluated based on uncertainty measures using the so-called prediction interval coverage probability, PICP and the mean prediction interval, MPI (Shrestha and Solomatine, 2006). The PICP measures the probability that the observed values lie within the estimated PIs and it is estimated as:

$$\text{PICP} = \frac{1}{n} \sum_{t=1}^n C \quad (10)$$

$$\text{with } C = \begin{cases} 1, & \text{PL}_t^L \leq y_t \leq \text{PL}_t^U; \\ 0, & \text{otherwise.} \end{cases}$$

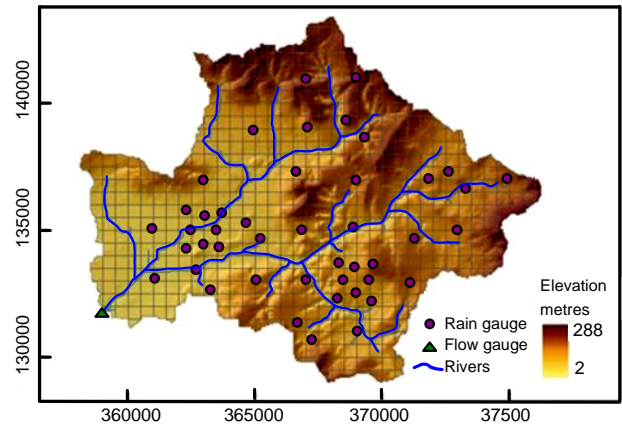


Fig. 1. The Brue catchment showing dense rain gauges network (reproduced from Shrestha and Solomatine (2008) with permission from the International Association for Hydraulic Research). The horizontal and vertical axes refer to the easting and northing in British national grid reference co-ordinates.

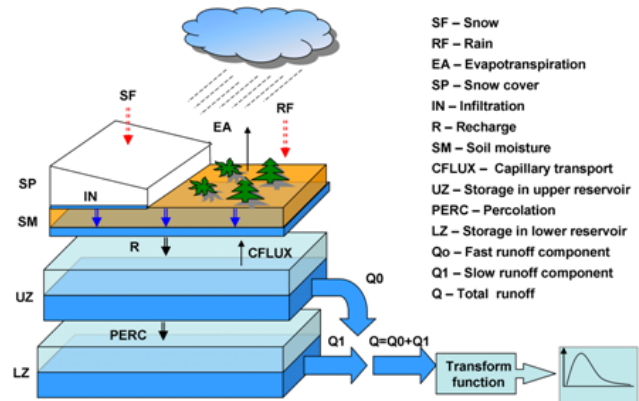


Fig. 2. Schematic representation of the HBV model with routines for snow, soil and runoff response (reproduced from Shrestha and Solomatine (2008) with permission from the International Association for Hydraulic Research).

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

$$\text{MPI} = \frac{1}{n} \sum_{t=1}^n (\text{PL}_t^U - \text{PL}_t^L) \quad (11)$$

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

Table 1. The statistical properties of the runoff data.

Statistical properties	Complete data	Calibration data	Verification data
Period (day/month/year hour)	24/06/1994 05:00 – 31/05/1996 13:00	24/06/1994 05:00 – 24/06/1995 04:00	24/06/1995 05:00 – 31/05/1996 13:00
Number of data	16977	8760	8217
Average (m ³ /s)	1.91	2.25	1.53
Minimum (m ³ /s)	0.15	0.17	0.14
Maximum (m ³ /s)	39.58	39.58	29.56
Standard deviation (m ³ /s)	3.14	3.68	2.37

4 Application

4.1 Study area

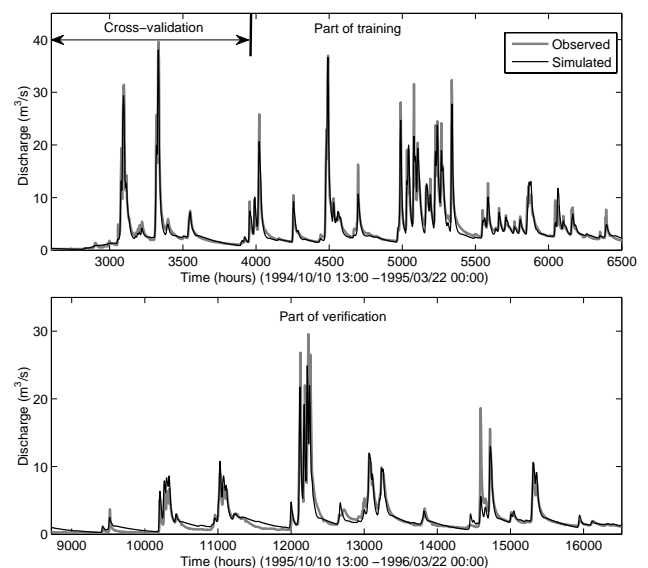
The Brue catchment located in the South West of England, UK is selected for the application of the methodology. The catchment has a drainage area of 135 km² with the average annual rainfall of 867 mm and the average river flow of 1.92 m³/s, for the period from 1961 to 1990. The discharge is measured at Lovington. The hourly potential evapotranspiration was computed using the modified Penman method recommended by FAO (Allen et al., 1998). Splitting of the available data set is based on Shrestha and Solomatine (2008), one year hourly data from 24 June 1994, 05:00:00 to 24 June 1995, 04:00:00 was selected for calibrating the HBV hydrological model, running MC simulations to generate data for uncertainty model *V* and training model *V*. Data from 24 June 1995, 05:00:00 to 31 May 1996, 13:00:00 was used for validating the hydrological model, running MC simulations to generate data for validating uncertainty model *V*. Each of the two data sets represents almost a full year of observations, and their statistical properties are shown in Table 1.

4.2 Conceptual hydrological model

A simplified version of the HBV model (Fig. 2) was used. 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 of the HBV model are listed in Table 1. 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, which was followed by manual adjustments of the parameters. The ranges of parameters for automatic calibration and for parameter uncertainty analysis were based on a range of calibrated values from other model applications (e.g., Braun and Renner, 1992) and the hydrologic descriptions of the catchment. The ranges were extended when the solutions were found

**Fig. 3.** Observed and simulated discharge hydrographs in a part of calibration (training+cross-validation) and verification period.

near the border of the parameter ranges, and re-calibration of the model was done with the extended ranges of the parameters. Sometimes automatic calibration gives the parameter values which do not represent the physical process well in all situations. Therefore, manual fine tuning of the parameters followed the automatic procedure by visual comparison of the observed and simulated hydrograph.

The model was calibrated using Nash-Sutcliffe model efficiency, CoE (Nash and Sutcliffe, 1970) as a performance measure of the HBV model. For the calibration period, the CoE was 0.96. The model was validated by simulating the flows for the independent verification data set, and the CoE was 0.83. Figure 3 shows the observed and simulated hydrograph in both calibration and verification period.

4.4 MC simulation and its convergence analysis

The parameters of the HBV model are sampled from the uniform distribution with the ranges given in Table 1. The model

Table 2. The ranges and calibrated value of the Hydrologiska Byråns Vattenbalansmodell (HBV) model parameters. The uniform ranges of parameters are used for calibration of the HBV model using the adaptive cluster covering algorithm and for analysis of the parameter uncertainty by MC simulations method.

Parameter	Description and unit	Ranges	Calibrated value
FC	Maximum soil moisture content (mm)	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 (/hour)	0.0005–0.1	0.001
K4	Recession coefficient for lower tank (/hour)	0.0001–0.005	0.004
PERC	Percolation from upper to lower response box (mm/hour)	0.01–0.09	0.089
CFLUX	Maximum value of capillary flow (mm/hour)	0.01–0.05	0.038
MAXBAS	Transfer function parameter (hour)	8–15	12.00

is run for each random parameter set and the likelihood measure is computed for each model run. We use the sum of the squared errors as the basis to calculate the likelihood measure (see Freer et al., 1996) in the form:

$$L(\theta_i/D) = \left(1 - \frac{\sigma_e^2}{\sigma_{\text{obs}}^2}\right)^N \quad (12)$$

where $L(\theta_i/D)$ is the likelihood measure for the i^{th} model conditioned on the observations D , σ_e^2 is the associated error variance for the i^{th} model, σ_{obs}^2 is the observed variance for the period under consideration, N is a user defined parameter. We set N to 1; in this case Eq. (12) is equivalent to CoE.

We investigated the number of behavioural samples retained out of 74 467 MC samples for different values of the rejection threshold. It was observed that only 1/3 of simulations (25 000 samples) are accepted for the threshold value of 0, whereas less than 1/10 of simulations are retained for the threshold value of 0.7.

We have also tested the convergence of the MC simulations to know the number of samples required to obtain reliable results. Since we have used CoE as an objective function to calibrate the model and as a likelihood measure in GLUE, we used the same metric to determine the convergence of the MC simulations. The mean and standard deviation of CoE were used to analyse the convergence of the MC simulations:

$$\text{ME}_k = \frac{1}{k} \sum_{i=1}^k (\text{CoE}_i) \quad (13)$$

$$\text{SDE}_k = \sqrt{\frac{1}{k} \sum_{i=1}^k (\text{CoE}_i - \text{ME}_k)^2} \quad (14)$$

where CoE_i is the CoE of the i^{th} MC run, ME_k and SDE_k are the mean and standard deviation of the model efficiency of the model up to the k^{th} run, respectively. Other statistics for

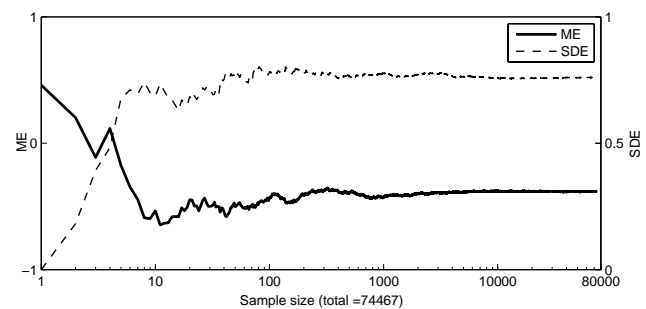


Fig. 4. The convergence of the mean (ME) and the standard deviation (SDE) of Nash-Sutcliffe model efficiency. Note that x-axis is log scale to see initial variation.

testing convergence can be found in Ballio and Guadagnini (2004). Figure 4 depicts the two statistics – the mean and standard deviation of the CoE – 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.

We also carried out experiments with other 9 performance metrics (e.g., RMSE, absolute error, percentage bias and others) for the convergence test. It is observed that the results are consistent with the CoE metric. Furthermore we also analysed the convergence of the runoff predictions at some arbitrary points (e.g., in peak flow, medium flow and base flow), and the results are quite consistent with the previous global performance metrics.

4.5 ANN for emulating MC simulation

Once the uncertainty results generated by the MC simulations are obtained, the ANN is trained to learn the functional relationship between the uncertainty results and the input data. The GLUE method has been used for the parameter uncertainty estimation of the HBV model. The threshold value

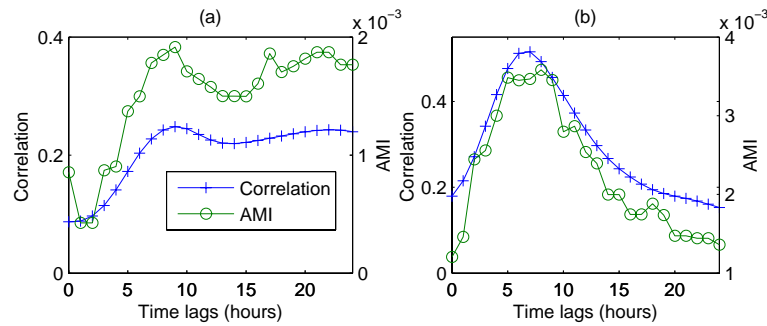


Fig. 5. Linear correlation and average mutual information (AMI) between effective rainfall (a) lower prediction interval; and (b) upper prediction interval for different time lags.

of 0 (measured by CoE) is selected to classify the simulation as either behavioural or non-behavioural. 90% uncertainty bands are calculated using the 5% and 95% quantiles of the predicted output likelihood weighted distribution resulting from the MC simulation realizations.

The corresponding 90% lower and upper PIs (see Eq. 6) are calculated using the model output simulated by the optimal model parameter sets found by Shrestha and Solomatine (2008). Hence the computed lower and upper PIs are conditional on contemporary value of the model simulation. The next step was to select the most relevant input variables to build a predictive ANN model.

It is worth mentioning that we have not studied the uncertainty of the ANN model. However, we have tried to ensure its high accuracy, and hence lower uncertainty: the number of hidden neurons was optimized, and different initializations of ANN weights and biases were tried. For sure, more could be done to investigate the uncertainty of the resulting ANN, for example, with the help of Bayesian approaches (see, e.g., Khan and Coulibaly, 2006).

4.5.1 Selection of input variables

To select the input variables for the ANN model, several approaches can be used (see, e.g., Solomatine and Dulal, 2003; Guyon and Elisseeff, 2003; Bowden et al., 2005). The input variables for the model V are constructed from the rainfall, evapotranspiration and observed discharge. Experimental results show that the evapotranspiration alone does not have a significant influence on the PIs. 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). The following conventions are used throughout this manuscript while defining the input variables:

$RE_{t-\tau}$: effective rainfall at time $t-\tau$;

$Q_{t-\tau}$: discharge at time $t-\tau$; where τ is lag time (0, 1, 2, ..., τ_{\max}).

Figure 5 shows the correlation coefficient and the AMI (see Eq. 10) of RE_t and its lagged variables with the lower

and upper PIs. It is observed that the correlation coefficient is minimum at a zero hour lag time and increases as the lag increases up to 7 h (Fig. 5b) for the upper PI. The correlation plot (see Fig. 5a) for the lower PI is different than the upper PI. The optimal lag time (the time at which the correlation coefficient and/or AMI is maximum) is 9 h. Such findings are also supported by the AMI analysis. At optimal lag time, the variable RE_t provides the maximum amount of information about the PIs. Additionally, the correlation and AMI between the PIs and the observed discharge were analysed. The results show that the immediate and the recent discharges (with the lag of 0, 1, 2) have a very high correlation with the PIs. So it was also decided to use the past values of the observed discharge as additional input to the model V .

Based on the above analysis, several structures for the input data for the ANN model were considered. The principle of parsimony was followed to avoid the use of a large number of inputs, so the aggregates, such as moving averages or derivatives of the inputs that would have a hydrological meaning were considered. Typically, the rainfall depth at a shorter (e.g., an hourly) time step partially exhibits random behaviour, and is not very representative of the rainfall phenomenon during the short period. Considering the maximum dependence to 7 h and 9 h lag time, we used two mean effective rainfall values:

1. RE_{t-7a} – the mean of RE_{t-3} , RE_{t-4} , RE_{t-5} , RE_{t-6} , RE_{t-7} ; and
2. RE_{t-9a} – the mean of RE_{t-5} , RE_{t-6} , RE_{t-7} , RE_{t-8} , RE_{t-9} .

Furthermore, the 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 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} , the rate of flow change at time $t-1$ is also considered.

Table 3 presents five possible combinations of input structure for the two ANN models. Note that these two models

Table 3. Input data structures of ANN models to reproduce Monte Carlo uncertainty results of the HBV model.

ANN models	Lower prediction interval	Upper prediction interval
ANN11	$RE_{t-9a}, Q_{t-1}, \Delta Q_{t-1}$	$RE_{t-9a}, Q_{t-1}, \Delta Q_{t-1}$
ANN22	$RE_{t-7a}, Q_{t-1}, \Delta Q_{t-1}$	$RE_{t-7a}, Q_{t-1}, \Delta Q_{t-1}$
ANN33	$RE_{t-7}, RE_{t-8}, RE_{t-9}, Q_{t-1}, \Delta Q_{t-1}$	$RE_{t-7}, RE_{t-8}, RE_{t-9}, Q_{t-1}, \Delta Q_{t-1}$
ANN44	$RE_{t-9a}, Q_{t-1}, Q_{t-2}$	$RE_{t-9a}, Q_{t-1}, Q_{t-2}$
ANN12	$RE_{t-9a}, Q_{t-1}, \Delta Q_{t-1}$	$RE_{t-7a}, Q_{t-1}, \Delta Q_{t-1}$

are trained independently for the lower and upper PIs; however, it is possible to use a single ANN model to produce two outputs given that it has same input structures (e.g., ANN models ANN11, ANN22, ANN33, ANN44 in Table 3).

The structure of the two ANN models to estimate the lower and upper PI, for instance in ANN11 configuration, takes the following form:

$$PI_t^L = V_L(RE_{t-9a}, Q_{t-1}, \Delta Q_{t-1}) \quad (15)$$

$$PI_t^U = V_U(RE_{t-9a}, Q_{t-1}, \Delta Q_{t-1}) \quad (16)$$

where, PI_t^L and PI_t^U are the lower and upper PI for t^{th} time step, and ΔQ_{t-1} is $Q_{t-1} - Q_{t-2}$ (characterizes the derivative of the previous discharge). Note that we have not used Q_t as input to the ANN model because during the model application this variable is not available (indeed, the prediction of this variable is done by the HBV model and the ANN model assesses the uncertainty of the prediction). Furthermore, we would like to stress that in this study the uncertainty of the model output is assessed when the model is used in simulation mode. However, this method can also be used in forecasting mode, provided that the process model is also run in forecasting mode.

4.5.2 Model training

The same data sets used for the calibration and verification of the HBV model were used for training and verification of the model V , respectively. However, for proper training of the ANN models the calibration data set is segmented into two subsets: 15% of data sets for cross-validation (CV) and 85% for training (see Fig. 3).

The CV data set was used to identify the best structure of the ANN. In this paper, a multilayer perceptron network with one hidden layer was used; optimization was performed by the Levenberg-Marquardt algorithm. The hyperbolic tangent function was used for the hidden layer with the linear transfer function at the output layer. The maximum number of epochs was fixed at 1000. The trial and error method was 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 seven and eight neurons for the lower and upper PI respectively give the lowest error on the CV data set.

4.5.3 Comparison of computational time

The standard MC runs of 25 000 simulations with the HBV model for hourly resolution data of one year takes about 9 h CPU time in a standard PC (Pentium Dual core 1.8 GHz, 1.8 GHz, 2 GB RAM) (there is the exchange of data between the MC code and the model via a file). Training of the ANN model in this study took about 30 min of CPU time using Levenberg-Marquardt optimization algorithm including a couple of iterations to get good results. Time required to choose the relevant input variables and the data preparation depends on the modeller's experience. Generally it may take a few hours to a day, but this is done once. Afterwards the trained ANN model is used in operation to predict the uncertainty, while in conventional MC based uncertainty analysis the process based model has to be run for a large numbers of times for each time step when prediction is needed. This prohibits the use of the MC simulation in many real-time applications, especially for computationally intensive models where run time of the model might be hundred times greater than that of the model considered in this study.

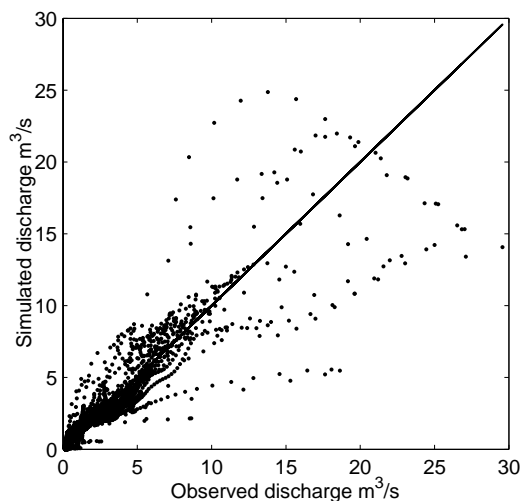
5 Results and discussions

Figure 6 shows the scatter plot of observed and simulated discharge by HBV model in the verification period. For many data points the HBV model is quite accurate but its error (uncertainty) is quite high during the peak flows. It is observed that residuals are autocorrelated and heteroscedastic (see, e.g., Shrestha and Solomatine, 2008). 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. Furthermore input data might be erroneous.

The ANN based uncertainty model V trained on the data generated by the MC simulations, was tested in the verification data set; its performance is shown in Table 4. All ANN models are characterized by similar values of the coefficient of correlation (CoC) in the CV data set producing a lower PI. In the verification data set, ANN11 or ANN12 (note that configurations ANN11 and ANN12 have the same ANN model for the lower PI), ANN44 have the highest CoC value. In producing the upper PI, ANN22 or ANN12 (configurations ANN22 and ANN12 have the same ANN model

Table 4. Performances of different ANN models measured by the coefficient of correlation (CoC), the prediction interval coverage probability (PICP) and the mean prediction interval (MPI).

ANN configurations	CoC for lower prediction interval			CoC for upper prediction interval			PICP	MPI
	Training	CV	Verification	Training	CV	Verification	Verification	Verification
ANN11	0.91	0.95	0.86	0.80	0.86	0.80	77.00	2.09
ANN22	0.92	0.95	0.83	0.83	0.88	0.79	81.78	2.24
ANN33	0.91	0.95	0.79	0.75	0.84	0.76	73.70	1.90
ANN44	0.90	0.96	0.86	0.79	0.88	0.80	73.99	1.93
ANN12	0.91	0.95	0.86	0.83	0.88	0.79	76.55	2.09

**Fig. 6.** Scatter plot of observed and simulated discharge by HBV model in the verification period.

for the upper PI) and ANN44 give the highest CoC value in the CV data set. Figure 7 presents the scatter plots of one of the ANN models to produce the lower and upper prediction limits in the verification period. It appears that the CoC values obtained when predicting the lower prediction limits are higher than those for the upper prediction limits. This can be explained by the fact that the upper prediction limits correspond to the higher values of flow (where the HBV model is less accurate) and have higher variability, which makes the prediction a difficult task.

We also compared the percentage of the observation data falling within the PIs (i.e., PICP) produced by the MC simulations and the ANN models. The value of PICP obtained by the MC simulations is 77.24%. Note that we specified the confidence level of 90% to produce these PIs and theoretically one would expect a PICP value of 90%. However it has been reported that the PICP obtained by the MC simulations is normally much lower than the specified confidence level used to produce the PIs. This low value of PICP is consistent with the results reported in the literature (see, e.g., Montanari, 2005; Xiong and O'Connor, 2008). The low efficiency

of the PIs obtained by the MC simulations in enveloping the real-world discharge observations might be mainly due to the following three reasons among others:

- The uncertainty in the model structure, the input (such as rainfall, temperature data) and output discharge data are not considered in the MC simulations.
- We assumed a uniform distribution and ignored the parameter correlation. These assumptions could be wrong.
- We employed the GLUE method. The width of the uncertainty bound obtained by the GLUE method varies with the rejection threshold and the likelihood measure to a great extent.

Among the ANN models, ANN22 gives the best result corresponding to a PICP value (i.e. 81.78% against the required 90%). However ANN11 and ANN12 gives PICP values very close to the MC simulation results. The average width of the PIs (i.e., MPI) by ANN22 is wider; this is one of the reasons to include a little bit more observed data inside the PIs. ANN33 gives the lowest MPI value with the lowest PICP value. The MPI values of ANN11 and ANN12 are very close (almost equal) to the MC simulation results.

Figure 8 shows the comparison of the 90% PIs estimated by the MC simulations with 5 different ANN configurations of the input data in the verification period. We also compared the PIs of ANN12 with those of ANN11 (Fig. 8a and b), ANN22 (Fig. 8c and d), ANN33 (Fig. 8e and f), and ANN44 (Fig. 8g and h). It is observed that the upper PI on peaks are overestimated by ANN11 compared to ANN12. Note that the lower PIs of ANN11 and ANN12 coincide because the input structures of the ANN11 and ANN12 for the lower PIs are the same. One can see from Fig. 8c and d that ANN22 and ANN12 give similar results with a slight underestimation of the lower PI. In this comparison, the upper PI of ANN22 and ANN12 coincide because of the same input structures for the upper PI. One can see a noticeable difference between ANN33 and ANN12 for both the upper and the lower PI. Most of the upper PIs on the peaks are overestimated by ANN33 and the lower PIs on most of peaks are considerably underestimated. This result is not surprising if one analyses

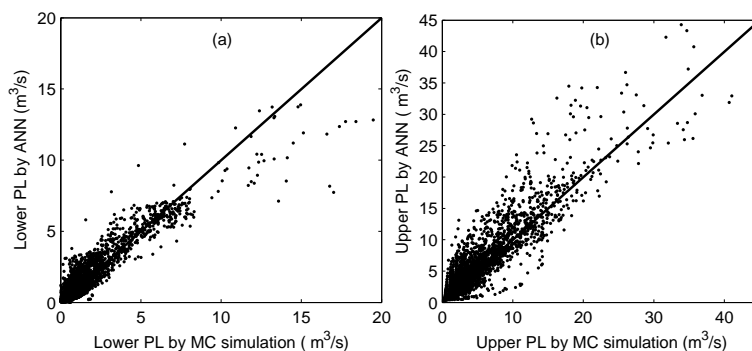


Fig. 7. Scatter plots showing the performance of the ANN model to reproduce 90% prediction limits (PL) of MC simulations in the verification data set. **(a)** 90% lower PL **(b)** 90% upper PL. X-axes show the prediction limits obtained by MC simulations and Y-axes show the prediction limits estimated by ANN.

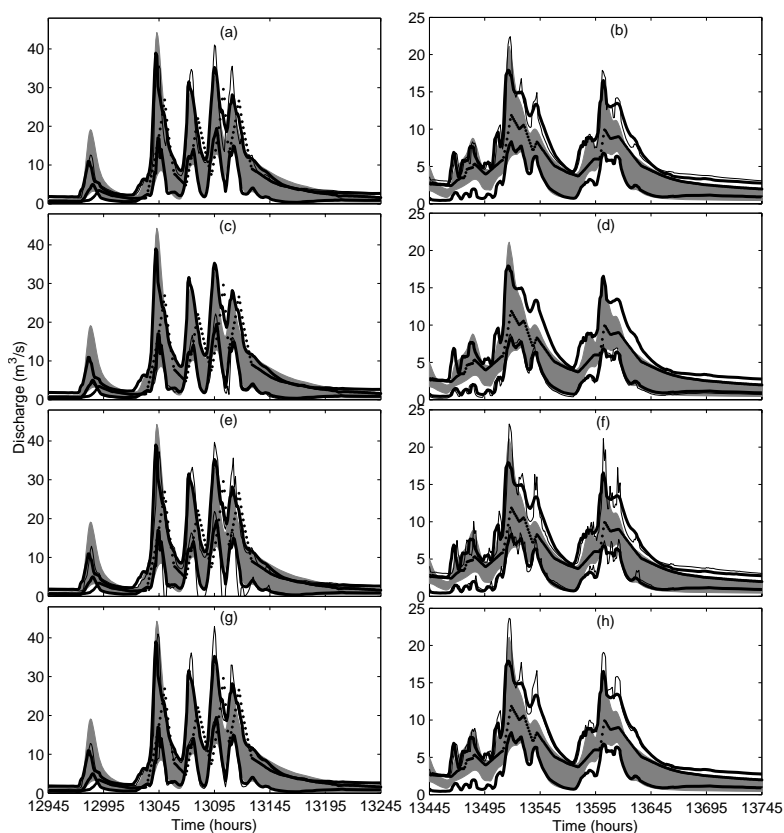


Fig. 8. A comparison of the 90% PIs estimated by the MC simulations (darker shaded region) and ANN12 (thick lines) with **(a and b)** ANN11, **(c and d)** ANN22, **(e and f)** ANN33, and **(g and h)** ANN44 in parts of the verification period. The black dots indicate observed discharges and the thin lines denote the 90% PIs obtained by ANN11, ANN22, ANN33, ANN44.

the structure of the input data. The obvious reason is that the input data consists of instantaneous values of RE_{t-7} , RE_{t-8} , RE_{t-9} , while in the other models we used a moving average value of effective rainfall, and thus the smooth function of the ANN is produced. Figure 8g and h shows that inclusion of the input ΔQ_{t-1} does not improve the accuracy as expected. However overestimation of the upper PIs on the peaks is due

to the combined effect of considering RE_{t-9} (while upper PI have maximum correlation at 7 h) and of not using ΔQ_{t-1} .

From the above analysis it can be concluded that model ANN12 is relatively better: it produces 90% PIs close to PIs obtained by the MC simulation. It can be said that in general most of the ANN models (except ANN33) reproduce the MC simulations uncertainty bounds reasonably well except

for some peaks, 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 the MC uncertainty bounds. Noticeably, the model fails to capture the observed flow during one of the peak events (Fig. 8a, c, e, and g). Note however, that the results of the ANN model and the MC simulations are visually closer to each other than both of them to the observed data.

It is also observed from Fig. 8 that the model prediction uncertainty caused by parameter uncertainty is rather large. There could be several reasons for this including:

- The GLUE method does not strictly follow the Bayesian inference process (Mantovan and Todini, 2006), which leads to an overestimation of the model prediction uncertainty.
- In the GLUE method, the uncertainty bound very much depends on the rejection threshold to distinguish between the behavioural and non-behavioural models. In this study we used a quite low value of the rejection threshold (CoE value of 0) which produces relatively wider uncertainty bounds.
- We considered only parameter uncertainty assuming that the model structure and the input data are correct. As mentioned at the beginning of this section, the scatter plot reveals that this assumption is not really correct.

6 Conclusions

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

There are basically three main steps involved in the proposed method:

1. Running an MC based uncertainty analysis method to generate data for the ANN uncertainty model;
2. Preprocessing and analysing the data to select relevant input variables for the ANN uncertainty model; and
3. Building and training the ANN model.

The ANN models are first trained on the data generated by the MC simulations to encapsulate the relationship between the hydrometeorological variables and the characteristics of the model output probability distribution (e.g., quantiles or prediction intervals, PIs), and then the trained ANN models are used to estimate the uncertainty descriptors such as PIs for the new input data. The 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 a 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% PIs. 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 the MC simulations. The conceptual hydrological model HBV was applied to the Brue catchment in the United Kingdom and used as a case study. The results demonstrate that the ANN approximates uncertainty of the model prediction with reasonable accuracy, and this is an indicator that the presented method can be a valuable tool for assessing uncertainty of various predictive process models. The proposed method can be used to emulate the results of various MC based uncertainty analysis methods such as Markov chain Monte Carlo sampling, Latin Hypercube sampling and others. 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”). The authors are also grateful to Roland Price for providing valuable remarks on the manuscript. The authors thank Jasper A. Vrugt for providing valuable comments. Fabrizio Fenicia and other two anonymous referees are also acknowledged for providing helpful and constructive reviews.

Edited by: N. Verhoest

References

- Abebe, A. J. and Price, R. K.: Managing uncertainty in hydrological models using complementary models, *Hydrol. Sci. J.*, 48, 679–692, 2003.
- Abebe, A. J. and Price, R. K.: Forecasting surge prediction accuracy on the Dutch coast using neural networks, in: *Proceedings of the 6th International Conference on Hydroinformatics: Singapore, 21–24 June 2004*, World Scientific, 2004.
- Abrahart, R. J. 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.
- Ajami, N. K., Duan, Q., and Sorooshian, S.: An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, *Water Resour. Res.*, 43, W01403, doi:10.1029/2005WR004745, 2007.

- Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration, guideline for computing crop water requirements, FAO, Irrigation and drainage paper, 56, 300 pp., 1998.
- 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.
- Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Proc.*, 6, 279–298, doi:10.1002/hyp.3360060305, 1992.
- 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.
- Bishop, C. M.: Neural networks for pattern recognition, Clarendon-Press, Oxford, 504 pp., 1995.
- Blasone, R. S., Vrugt, J. A., Madsen, H., Rosbjerg, D., Robinson, B. A., and Zyvoloski, G. A.: Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov chain Monte Carlo sampling, *Adv. Water Resour.*, 31, 630–648, doi:10.1016/j.advwatres.2007.12.003, 2008.
- Bowden, G. J., Dandy, G. C., and Maier, H. R.: Input determination for neural network models in water resources applications, Part 1 - background and methodology, *J. Hydrol.*, 301, 93–107, doi:10.1016/j.jhydrol.2004.06.020, 2005.
- Braun, L. N. and Renner, C. B.: Application of a conceptual runoff model in different physiographic regions of Switzerland, *Hydrol. Sci. J.*, 37, 217–232, 1992.
- Dawson, C. W. and Wilby, R. L.: Hydrological modelling using artificial neural networks, *Prog. Phys. Geogr.*, 25, 80–108, doi:10.1177/030913330102500104, 2001.
- Dibike, Y. B., Velickov, S., Solomatine, D. P., and Abbott, M. B.: Model induction with support vector machines: Introduction and applications, *J. Comput. Civil Eng.*, 15, 208–216, 2001.
- Fraser, A. M. and Swinney, H. L.: Independent coordinates for strange attractors from mutual information, *Phys. Rev. A*, 33(2), 1134–1140, 1986.
- Freer, J., Beven, K., and Ambrose, B.: Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach, *Water Resour. Res.*, 32, 2161–2173, 1996.
- Georgakakos, K., Seo, D.-J., Gupta, H. V., Schaake, J., and Butts, M. M.: Towards the characterization of streamflow simulation uncertainty through multimodel ensembles, *J. Hydrol.*, 298, 222–241, doi:10.1016/j.jhydrol.2004.03.037, 2004.
- Govindaraju, R. S. and Rao, A. R.: Artificial neural networks in hydrology, Kluwer Academic Publishers, Amsterdam, 348 pp., 2000.
- Guyon, I. and Elisseeff, A.: An introduction to variable and feature selection, *J. Mach. Learn. Res.*, 3, 1157–1182, 2003.
- Haario, H., Laine, M., Mira, A., and Saksman, E.: DRAM: efficient adaptive MCMC, *Stat. Comput.*, 16, 339–354, doi:10.1007/s11222-006-9438-0, 2006.
- Harr, M.: Probabilistic estimates for multivariate analyses, *Appl. Math. Model.*, 13, 313–318, doi:10.1016/0307-904X(89)90075-9, 1989.
- Haykin, S.: Neural networks: A comprehensive foundation. Prentice-Hall, Upper Saddle River, NJ, USA, 1999.
- Jacquín, A. P. and Shamseldin, A. Y.: Development of rainfall-runoff models using Takagi-Sugeno fuzzy inference systems, *J. Hydrol.*, 329(1–2), 154–173, 2006.
- Kavetski, D., Kuczera, G., and Franks, S. W.: Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, *Water Resour. Res.*, 42, W03407, doi:10.1029/2005WR004368, 2006.
- 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.
- Khan, M. S. and Coulibaly, P.: Bayesian neural network for rainfall-runoff modeling, *Water Resour. Res.*, 42, W07409, doi:10.1029/2005WR003971, 2006.
- Khu, S.-T. and Werner, M. G. F.: Reduction of Monte-Carlo simulation runs for uncertainty estimation in hydrological modelling, *Hydrol. Earth Syst. Sci.*, 7, 680–692, 2003, <http://www.hydrol-earth-syst-sci.net/7/680/2003/>.
- Kuczera, G. and Parent, E.: Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, *J. Hydrol.*, 211, 69–85, 1998.
- Kuczera, G., Kavetski, D., Franks, S. W., and Thyer, M.: Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, *J. Hydrol.*, 331, 161–177, doi:10.1016/j.jhydrol.2006.05.010, 2006.
- 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, 1035, doi:10.1029/2000WR000022, 2002.
- Li, G., Azarm, S., Farhang-Mehr, A., and Diaz, A.: Approximation of multiresponse deterministic engineering simulations: a dependent metamodeling approach, *Struct. Multidiscip. O.*, 31, 260–269, doi:10.1007/s00158-005-0574-5, 2006.
- 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.
- Maier, H. R. and Dandy, G. C.: Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications, *Environ. Modell. Softw.*, 15, 101–124, 2000.
- Mantovan, P. and Todini, E.: Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology, *J. Hydrol.*, 330, 368–381, doi:10.1016/j.jhydrol.2006.04.046, 2006.
- Maskey, S., Guinot, V., and Price, R. K.: Treatment of precipitation uncertainty in rainfall-runoff modelling: a fuzzy set approach, *Adv. Water Resour.*, 27, 889–898, 2004.
- McKay, M. D., Conover, W. J., and Beckman, R. J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, 21, 239–245, 1979.
- Melching, C. S.: An improved-first-order reliability approach for assessing uncertainties in hydrologic modeling, *J. Hydrol.*, 132, 157–177, 1992.
- Melching, C. S.: Reliability estimation in computer models of watershed hydrology, *Water Res. Publ.*, 69–118, 1995.
- Minns, A. W. and Hall, M. J.: Artificial neural networks as rainfall-runoff models, *Hydrol. Sci. J.*, 41, 399–417, 1996.
- Montanari, A. and Brath, A.: A stochastic approach for assessing the uncertainty of rainfall-runoff simulations, *Water Resour. Res.*, 40, W01106, doi:10.1029/2003WR002540, 2004.
- Montanari, A.: Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty

- of rainfall-runoff simulations, *Water Resour. Res.*, 41, W08406, doi:10.1029/2004WR003826, 2005.
- Moradkhani, H., Hsu, K.-L., Gupta, H. V., and Sorooshian, S.: Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, *Water Resour. Res.*, 41, W05012, doi:10.1029/2004WR003604, 2005.
- Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models – Part I – A discussion of principles, *J. Hydrol.*, 10, 282–290, 1970.
- O'Hagan, A.: Bayesian analysis of computer code outputs: A tutorial, *Reliab. Eng. Syst. Safe.*, 91, 1290–1300, doi:10.1016/j.ress.2005.11.025, 2006.
- 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.
- Rosenblueth, E.: Two-point estimates in probability, *Appl. Math. Model.*, 5, 329–335, doi:10.1016/S0307-904X(81)80054-6, 1981.
- Shrestha, D. L. and Solomatine, D. P.: Machine learning approaches for estimation of prediction interval for the model output, *Neural Networks*, 19, 225–235, doi:10.1016/j.neunet.2006.01.012, 2006.
- Shrestha, D. L. and Solomatine, D. P.: Data-driven approaches for estimating uncertainty in rainfall-runoff modelling, *Int. J. River Basin Management*, 6, 109–122, 2008.
- Solomatine, D. P., Dibike, Y. B., and Kukuric, N.: Automatic calibration of groundwater models using global optimization techniques, *Hydrol. Sci. J.*, 44, 879–894, 1999.
- Solomatine, D. P. and Dulal, K. N.: Model trees as an alternative to neural networks in rainfall-runoff modelling, *Hydrol. Sci. J.*, 48, 399–411, 2003.
- Solomatine, D. P. and Shrestha, D. L.: A novel method to estimate model uncertainty using machine learning techniques, *Water Resour. Res.*, 45, W00B11, doi:10.1029/2008WR006839, 2009.
- Thiemann, M., Trosset, M., Gupta, H. V., and Sorooshian, S.: Bayesian recursive parameter estimation for hydrologic models, *Water Resour. Res.*, 37, 2521–2535, 2001.
- Tung, Y.-K.: Uncertainty and reliability analysis, in: *Water Resources Handbook*, edited by: Mays, L. W., McGraw-Hill Book Company, 7.1–7.64, 1996.
- Vernieuwe, H., Georgieva, O., De Baets, B., Pauwels, V. R. N., Verhoest, N. E. C., and De Troch, F. P.: Comparison of data-driven Takagi-Sugeno models of rainfall discharge dynamics, *J. Hydrol.*, 302(1–4), 173–186, 2005.
- Vrugt, J. A., Diks, C., Gupta, H. V., Bouten, W., and Verstraten, J. M.: Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resour. Res.*, 41, W01017, doi:10.1029/2004WR003059, 2005.
- Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S.: A Shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, 39, 1201, doi:10.1029/2002WR001642, 2003.
- Vrugt, J. A. and Robinson, B. A.: Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging, *Water Resour. Res.*, 43, W01411, doi:10.1029/2005WR004838, 2007.
- Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., and Robinson, B. A.: Treatment of input uncertainty in hydrologic modeling: doing hydrology backwards with Markov chain Monte Carlo simulation, *Water Resour. Res.*, 44, W00B09, doi:10.1029/2007WR006720, 2008.
- Xiong, L. and O'Connor, K.: An empirical method to improve the prediction limits of the GLUE methodology in rainfall-runoff modeling, *J. Hydrol.*, 349, 115–124, doi:10.1016/j.jhydrol.2007.10.029, 2008.
- Young, P. C.: Data-based mechanistic modelling of environmental, ecological, economic and engineering systems, *Environ. Modell. Softw.*, 13, 105–122, doi:10.1016/S1364-8152(98)00011-5, 1998.
- Young, P. C. and Ratto, M.: A unified approach to environmental systems modeling, *Stoch. Env. Res. Risk A.*, 10, 008–0271, doi:10.1007/s00477-008-0271-1, 2008.