Stoch Environ Res Risk Assess (2017) 31:993–1010
DOI 10.1007/s00477-017-1408-x

ORIGINAL PAPER

Quantification of parametric uncertainty of ANN models with GLUE method for different streamflow dynamics

Hakan Tongal[1] · **Martijn J. Booij**[2]

Published online: 28 March 2017

Abstract This study describes the parametric uncertainty of artificial neural networks (ANNs) by employing the generalized likelihood uncertainty estimation (GLUE) method. The ANNs are used to forecast daily streamflow for three sub-basins of the Rhine Basin (East Alpine, Main, and Mosel) having different hydrological and climatological characteristics. We have obtained prior parameter distributions from 5000 ANNs in the training period to capture the parametric uncertainty and subsequently 125,000 correlated parameter sets were generated. These parameter sets were used to quantify the uncertainty in the forecasted streamflow in the testing period using three uncertainty measures: percentage of coverage, average relative length, and average asymmetry degree. The results indicated that the highest uncertainty was obtained for the Mosel sub-basin and the lowest for the East Alpine sub-basin mainly due to hydro-climatic differences between these basins. The prediction results and uncertainty estimates of the proposed methodology were compared to the direct ensemble and bootstrap methods. The GLUE method successfully captured the observed discharges with the generated prediction intervals, especially the peak flows. It was also illustrated that uncertainty bands are sensitive to the selection of the threshold value for the Nash–Sutcliffe efficiency measure used in the GLUE method by employing the Wilcoxon–Mann–Whitney test.

Keywords Generalized likelihood uncertainty estimation (GLUE) · Parametric uncertainty · Bootstrap · Artificial neural networks · Uncertainty measures · Rhine Basin

✉ Hakan Tongal
hakantongal@sdu.edu.tr

Martijn J. Booij
m.j.booij@utwente.nl

[1] Department of Civil Engineering, Engineering Faculty, Süleyman Demirel University, Isparta, Turkey

[2] Department of Water Engineering and Management, Faculty of Engineering Technology, University of Twente, Enschede, The Netherlands

1 Introduction

Reliable and timely forecasting of streamflow with reasonable estimates of predictive uncertainty is required for hydrological activities with regard to controlling, planning, and operating of water resources and for reducing risks in decision making (Zhang et al. 2009). Sustainable water resources management is usually based on evaluating various scenarios with possible outcomes. Over- or under-estimation of uncertainty associated with model outputs could lead to over-design or alleviation of measures to be taken for possible situations. Artificial neural networks (ANNs) have frequently been used for streamflow forecasting (e.g., Jothiprakash and Magar 2012; Pramanik and Panda 2009; Tongal and Berndtsson 2016; Wang et al. 2009b; Zeroual et al. 2016), since they have the capability to reproduce the highly nonlinear nature of streamflow dynamics (Toth and Brath 2007) and do not necessitate a priori assumptions related to normality and homoscedasticity. However, as indicated by Kasiviswanathan and Sudheer (2013), development of an ANN model is stochastic in nature, and different results can be obtained in various model configurations that weaken the reliability of the outputs of ANNs. Maier and Dandy (2000) concluded that a major limitation of ANNs is that the uncertainty associated with predictions is seldom quantified. One possible reason of this could be that ANNs have significant degrees of freedom in

their development. Therefore, few studies have considered the quantification of uncertainty inherent in the obtained forecasting results from ANNs. Nevertheless, the quality of predictions related to the hydrological applications of ANN could be severely impacted without accounting for uncertainty in predictions that could limit their usability (Kasiviswanathan and Sudheer 2013; Kingston et al. 2005). Thus, to improve the reliability and credibility of outputs of ANNs, the quantification of the uncertainty associated with the results produced with ANNs is of utmost importance (Srivastav et al. 2007).

Generally, three major sources of uncertainties affect the reliability of the model estimated discharge: input uncertainty resulting from errors associated with data, model structure uncertainty influenced by an imperfect model structure, and uncertainty in model parameters, i.e. parametric uncertainty (Chen et al. 2013). Bayesian techniques (Kasiviswanathan et al. 2016; Khan and Coulibaly 2006; Kingston et al. 2005; Zhang et al. 2009) and bootstrap methods (Boucher et al. 2010; Kumar et al. 2015; Srivastav et al. 2007; Tiwari and Chatterjee 2010) are one of the most employed approaches for construction of prediction intervals of ANN models. The Bayesian approach is based on a stringent theoretical framework, however, it is computationally expensive, relies on some hypotheses that have to be carefully analyzed, and requires the calculation of the Hessian matrix of the cost function (Kasiviswanathan and Sudheer 2013; Papadopoulos et al. 2001). Recently, Bayesian neural networks (BNNs) have been applied for hydrological modeling. Amongst others, Kingston et al. (2005) proposed Bayesian training of ANN models to account for parametric uncertainty in the predictions by combining traditional ANNs with the adaptive Metropolis and Gibbs sampling algorithms to obtain a large number of parameter sets (i.e. weights and biases). While the proposed methodology successfully generated the prediction limits that indicate the level of uncertainty, it is computationally intensive and requires complex implementation and coding. Also, as the authors noted, the results "may be biased by the initial weights, which is not statistically optimal." Zhang et al. (2009) applied an evolutionary Monte Carlo training algorithm for four BNNs and estimated uncertainty limits of streamflow simulation. However, as the authors concluded, the 95% uncertainty intervals generated by all four BNNs failed to include 95% or more of observed streamflow data. This was partially attributed to inadequate recognition of all uncertainty sources and improper definition of error characteristics associated with different uncertainty sources. Additionally, there is a question of the effectiveness of the sampling algorithm of BNNs for generating models observing complex posterior distributions (Zhang et al. 2009). On the contrary, the use of the bootstrap method is highly attractive due to its simplicity and the ease of implementation. For instance, it involves relatively few assumptions that led to a relevant error distribution (Selle and Hannah 2010) and does not necessitate the calculation of the Hessian matrix involved in the Bayesian technique (Kasiviswanathan and Sudheer 2013). Srivastav et al. (2007) quantified the parametric uncertainty by using the bootstrap technique through training of ANN models by resampling of input–output patterns with replacement. Tiwari and Chatterjee (2010) followed a similar methodology to quantify the parametric uncertainty of ANN outputs.

The results of the studies mentioned above indicated that despite obtaining a good generalization in forecasting, the developed models failed to capture the peak flow characteristics of hydrographs that limits the usability of the method in a flood forecasting context. Furthermore, in this approach, it is implicitly assumed that resampling of input–output patterns with or without replacement would lead to parametric uncertainty resulting from the stochastic nature of input data which may not always be valid. The obtained uncertainty in outputs simply could stem from the sampling uncertainty. Thus, to overcome the mentioned disadvantages of the Bayesian and bootstrap approaches, we will show that the rarely used (i.e. in ANN modeling) generalized likelihood uncertainty estimation (GLUE) method (Beven and Binley 1992) for the quantification of the parametric uncertainty of ANNs could be efficient and promising.

The GLUE method acknowledges multiple parameter sets rather than one optimal solution that gives acceptable simulations of the considered system. Many hydrological studies have employed the GLUE method to quantify uncertainty and evaluate models (see Breinholt et al. 2013; Tian et al. 2014; Uniyal et al. 2015), however, there are few applications of the GLUE method to ANN models. Rogiers et al. (2012) employed the GLUE method to quantify the uncertainty associated with the hydraulic conductivity predictions obtained from an ANN model that uses the entire grain-size distribution. Yu et al. (2015) proposed the incorporation of the moving least-square with entropy (MLS-E) method for stochastic sampling in the GLUE method for uncertainty analysis of flood inundation modeling and concluded that the proposed method outperformed moving least squares, quadratic surface response, and ANN in both the predicted confidence interval and the most likely value of water depths. Noori et al. (2010) applied uncertainty analysis for daily carbon monoxide concentration in the atmosphere by using an ANN and an adaptive neuro-fuzzy inference system based on Monte Carlo simulations. The GLUE method has been criticized for not being formally Bayesian and resulting in statistically incoherent and unreliable parameter and predictive distributions (Mirzaei et al. 2015) due to the subjective choice of a likelihood measure and the

corresponding threshold value to distinguish "behavioural" and "non-behavioural" parameter sets (i.e. the parameter sets having a likelihood measure above a threshold value are considered as "behavioural" and vice versa). Additionally, Jin et al. (2010) claimed that the GLUE methodology is not appropriate for parameter sets that have a correlation among the parameters. However, Beven and Freer (2001) argued that "any effects of model nonlinearity, covariation of parameter values and errors in model structure, input data or observed variables, with which the simulations are compared, are handled implicitly within this procedure [GLUE]." Besides the disadvantages mentioned above, the stochastic nature of the development of an ANN model including a high number of parameters could be responsible for a relatively small number of GLUE applications to ANN models in hydrology.

In this study, we will show that a modified version of the GLUE method provides a suitable framework for quantifying parametric uncertainty because of the ease of implementation and the fact that no adjustment of the existing code of the simulation model is needed. The answers to the following research questions will be addressed: (1) could GLUE be efficient for quantifying the "pure" parametric uncertainty of ANN models without bootstrapping of input data? (2) are there any advantages of using the GLUE method over the bootstrap technique? (3) does the uncertainty in forecasted streamflow show significant variation among three sub-basins of the Rhine Basin, namely East Alpine, Main, and Mosel sub-basins, having different hydro-climatological characteristics? and (4) how sensitive are the GLUE results to selected threshold values of the likelihood function?

The remainder of the paper is organized as follows. Section 2 briefly presents the study area and data. The methodology employed is described in Sect. 3. Results of the analyses are presented and discussed in Sect. 4, and final conclusions are drawn in Sect. 5.

2 Study area and data

The Rhine River has a length of 1233 km and a catchment area of 185,300 km^2 with densely populated and highly industrialized areas in Western Europe. Being of great economic and environmental importance for the riparian countries, the Rhine River is used for various aims, such as agriculture, hydropower generation, industry and domestic water use (Middelkoop et al. 2001). The Rhine River, originating from the Swiss Alps, feeds one of the largest lakes of Europe, Lake Constance. It forms part of the French–German border and receives water from several tributaries such as the Neckar, Main, and Mosel before flowing into the North Sea (Hurkmans et al. 2007). In general, the discharge regime of the Rhine is dominated by rainfall and snow melt from the Alpine glaciers.

In this study, we have selected the East Alpine, Main, and Mosel sub-basins (Disse and Engel 2001) as case studies because these basins have different climatological and geographical characteristics and carrying vast volumes of water. The Main and Mosel sub-basins are rainfall-dominated areas while snow storage and snowmelt strongly affect the discharge regime of the East Alpine. Spatial characteristics and generated discharges of these tributaries can be seen below (Tongal et al. 2013) (Table 1).

Daily river flow, precipitation, evapotranspiration, and temperature data covering the period from January 1993 to December 2001 were obtained from the German Federal Institute of Hydrology in Koblenz (Germany) for the East Alpine (*station #6935055*), Main (*station #6335304*) and Mosel (*station #6336050*). Selecting a representative training data set which covers the dynamical patterns to eliminate the impact of data splitting-variability is essential for the forecasting performance of ANNs (Melesse et al. 2011). The data set was split into a training data set from 1 January 1993 to 31 December 1999 and a testing data set from 1 January 2000 to 31 December 2001. The training dataset was further divided into two sub-groups of which 80% is assigned to a calibration set, and the remaining 20% assigned to a validation set. To avoid over-fitting of the training data, an early stopping approach that uses the validation set was implemented.

3 Methodology

3.1 Artificial neural networks

Artificial Neural Networks are mathematical structures having interconnected neurons to map the relationship between input and output variables or to process received signal(s) to generate output(s). Among different types of ANNs based on their topology, activation functions, and training algorithms, multi-layer perceptron feed-forward neural networks are one of the most frequently used ANNs (e.g., Lohani et al. 2011; Singh et al. 2009). Practically, an ANN model with one-hidden-layer including sufficient neurons is capable of mapping nonlinear dynamical relationships (Badrzadeh et al. 2016) and provide enough complexity to map the relationships between input and output patterns (Hornik et al. 1989). An ANN model with two hidden layers generally gives less accurate results than its single hidden layer counterpart (De Villiers and Barnard 1993) only when the number of observations is limited and in this situation having more additional parameters can cause over-fitting. Thus, the one-hidden-layer feed-forward neural network was used in this study. In a typical feed-

Table 1 Spatial characteristics and average annual generated discharges for the period 1993–2001 of three sub-basins of the River Rhine

Sub-basin	Area (km^2)	Altitude range (m)	Mean annual discharges ($m^3\ s^{-1}$)
East Alpine	16,051	143–3270	395
Main	24,833	83–939	192
Mosel	27,262	59–1326	364

forward neural network with input, hidden, and output layers, every neuron in each layer is connected to a neuron located in an adjacent layer using different weights (Ghavidel and Montaseri 2014). Each neuron receives signals from the neurons of the predecessor layer with weighted connections except the input layer to produce the argument to the transfer function which generates the output of a neuron. The number of neurons in the input and output layers is a result of the modeled system/problem, and the number of neurons in the hidden layer should be determined. In this study, the number of neurons in the hidden layer was determined via a trial-and-error procedure. A typical ANN model can be expressed in a functional form as follows:

$$y_k = f_o\left[\sum_j w_{jk}\cdot f_h\left(\sum_i \tilde{w}_{ij}x_i + \tilde{b}_j\right) + b_k\right] \quad (1)$$

where x is an input vector, w_{jk} is the weight of the link between the jth hidden neuron to the kth output neuron, $\tilde{w}_{ij}$ is the connection weight from the ith neuron in the input layer to the jth neuron in the hidden layer, b_k is bias of kth output neuron, $\tilde{b}_j$ is the threshold value or bias of jth hidden neuron, f_h and f_o represent the activation function for hidden and output neuron, respectively (Badrzadeh et al. 2013). A combination of a hyperbolic tangent function and a linear transfer function in the hidden and output layer was used since this is advantageous when extrapolating beyond the range of training data (Maier and Dandy 2000).

Optimal weights and biases can be found by minimizing the error between the estimated and observed outputs. Converging fast, being robust and efficient for training small to medium-size networks, and being less easily trapped in local minima (Aqil et al. 2007; Badrzadeh et al. 2013), the Levenberg-Marquart algorithm (Hagan and Menhaj 1994) with the back-propagation algorithm was used for optimization. In this two-stage algorithm, the input signal is propagated through the layers to calculate the output(s) in the first step, and then the weights and biases are adjusted to minimize the network error in the second step.

3.2 GLUE methodology

The assessment and knowledge of uncertainty are of great significance and essential for a meaningful interpretation of the model results (Warmink and Booij 2015). A general framework of uncertainty analysis consists of contexting and framing, classification, importance assessment, quantification and propagation of uncertainty through a model (Demirel et al. 2013a). Walker et al. (2003) briefly characterized uncertainty sources as input, model structure, parameter, and model technical uncertainty. In this study, due to the high degree of freedom accompanying ANN model development (Kasiviswanathan and Sudheer 2013), and a significant part of the prediction uncertainty of an ANN model can result from the parameters (Kingston et al. 2005) only the parametric uncertainty of ANNs was considered.

The generalized likelihood uncertainty estimation (GLUE) method (Beven and Binley 1992) is an extension of the generalized sensitivity analysis of Spear and Hornberger (1980) and Hornberger and Spear (1981) within the context of the estimation and propagation of uncertainty by Bayesian Monte Carlo simulations. Based on the equifinality thesis which emphasizes that there are many acceptable behavioural parameter sets that give relatively equal performance according to a specific likelihood function, the modified GLUE method for the parametric uncertainty of an ANN model will be applied in this study in five steps:

1. Determining the model structure: relevant input parameters with their significant lagged observations have been selected with the help of the stepwise regression method (see Ghavidel and Montaseri 2014) for the daily forecasting of discharge. The number of neurons in the hidden layer has been determined via a trial-and-error process. Efficiencies of the constructed ANN models in streamflow forecasting were checked by the mean of the root mean square error (RMSE) normalized by the mean discharge values. The mean of the normalized RMSE values was obtained during the training and testing periods by running each model 100 times.
2. Reasonably wide ranges for the parameters are selected based on experience or prior distributions, and a large number of parameter sets are drawn from the prior distributions. GLUE can be applied to ANN only if the distributions of the weights and biases are known a priori as well as the sampling procedure considers the correlation between the weights and biases (Srivastav et al. 2007). Thus, appropriate distributions should be fitted to weights and biases, and correlated variables

should be drawn from these distributions. The uncertainty in ANN forecasts due to model parameters is a function of the type and shape of the probability distributions of the parameters (weights and biases). Generally, a high number of parameter sets is generated based on the prior knowledge about the distributions of parameters. However, since these distributions are not often known a priori, the uniform distribution with quite wide ranges is used instead. As mentioned by Srivastav et al. (2007), this will lead to biased outputs from ANN models since the parameters of ANN models do not have a physical meaning, and therefore, ranges of parameters values can not be determined a priori. Since the properties of the probability distributions of ANN parameters can be determined from the variability of the parameters during the calibration period (Srivastav et al. 2007), prior distributions of the parameter sets were identified in the training period of 01.01.1993–31.12.1999 with the calibrated 5000 ANN models. This number is sufficiently large enough since the obtained type of probability distributions and their parameter values remained unchanged between the calibrated 1000 ANN models to 5000 ANN models (not shown here) and reported studies related to obtaining predictive uncertainty from ANN models usually employed 500 ANN models (Kasiviswanathan and Sudheer 2013) to 1000 ANN models (Talebizadeh et al. 2010). The Bayesian Information Criterion (BIC) was used to identify the best-representing parameter distribution:

$$BIC = -2\ln(L_{\max}) + k\ln(n) \tag{2}$$

where n is the number of data points, k is the number of parameters to be estimated, and $L_{\max}$ is the maximized value of the log-likelihood function for the estimated distribution.

3. Defining a likelihood measure $l(\theta)$ to distinguish a parameter set either as "behavioural" or "non-behavioural" through a comparison of the associated likelihood value with a pre-determined threshold value. In this study, the widely-used Nash–Sutcliffe efficiency (NSE) in GLUE analyses (Demirel et al. 2013a) was used as a likelihood function which is defined as:

$$NSE = 1 - \frac{\sum_{t=1}^{n}\left(Q_t^m(\theta) - Q_t\right)^2}{\sum_{t=1}^{n}\left(Q_t^m(\theta) - \bar{Q}_t\right)^2} \tag{3}$$

where n is the total number of time steps, $\bar{Q}_t$ is the mean of observed discharges, $Q_t^m(\theta)$ and Q_t are the modeled discharges with parameter set θ and observed discharges at a time step t, respectively.

4. Estimating the likelihood weight (w_i) for each behavioural parameter set by, $w_i = l(\theta)/\sum_{k=1}^{N} l(\theta_k)$ where N is the number of behavioural parameter sets.
5. Finally, the prediction uncertainty of streamflow is determined by the quantiles of the cumulative distribution using the weighted behavioural parameter sets obtained from Step 4, for the given certainty level α. Upper and lower prediction limits can be determined by quantiles of $(1-\alpha/2)$ and $(\alpha/2)$ where α is assumed 0.05 in this study.

3.3 Indices for assessing the prediction bounds

The prediction bounds corresponding to a confidence level were evaluated based on the quantitative measures percentage of coverage (POC) (Kasiviswanathan and Sudheer 2016), average relative length (ARIL) (Jin et al. 2010), and average asymmetry degree (AAD) (Xiong et al. 2009). The POC assess the number of observed streamflows that falls within the prediction band where 100% indicates that all observed values are covered by the prediction bands. While Kasiviswanathan and Sudheer (2016) considered a POC value of an ideal prediction band as 100%, Zhang et al. (2009) preferred the uncertainty interval with a POC coefficient close to the expected proportion (i.e., 95%) which also is preferred in this study. The ARIL shows the bandwidth of the prediction bands for the discharge. The AAD shows the asymmetrical behaviour of prediction bands and could take values between 0 and 1. An AAD value between 0 and 0.5 means that observed values mainly lie between the corresponding upper and lower bounds. If it is equal to 0, which is the ideal case, then the observed value is equal to the middle point value of the prediction bounds at each step. If it is larger than 0.5, the majority of observed values either lie above or below the upper and lower bounds of the hydrograph, respectively. Formulas are presented in Appendix 2.

4 Results and discussion

4.1 Determination of input and model structures

In this study, typical one-hidden-layer feed-forward ANN models were constructed for forecasting one-day ahead streamflows based on precipitation (P), evapotranspiration (ET), temperature (T), and discharge (Q) for three Rhine sub-basins, namely East Alpine, Main, and Mosel. The main steps when building a neural network are choosing significant input variables with significant lags and identifying the optimal network structure. The forward

stepwise regression method was applied to determine the most significant input combination. In this method, input variables are subsequently added to the single best input variable up to the allowed number of antecedent values to obtain an improvement in the model performance, i.e. adjusted-R^2. The results showed that the adjusted-R^2 value, which takes into account the number of parameters of the model and evaluates whether the added new parameter improves the model more than would be expected by chance, was not statistically changed by adding more inputs than the input combination of Q_{t-1},Q_{t-2},P_t,P_{t-1} for the East Alpine, Q_{t-1},Q_{t-6},P_{t-1},T_{t-7} for the Main, and Q_{t-1},Q_{t-2},Q_{t-3},P_{t-1},T_{t-6} for the Mosel sub-basins. Since there is not a systematic or standard procedure to determine the number of hidden neurons, the number of hidden neurons was determined by a trial-and-error procedure similarly to Chiang et al. (2004), Chang et al. (2007), Jain et al. (2004) and Wang et al. (2009a) by varying the number from 2 to 10. This is because too many hidden neurons may lead to over-fitting, reduce the generalization capabilities and increase the training time without significant improvement of model performance (Chang et al. 2007; Ranjithan et al. 1993). The final structures of the ANN models are 4-4-1, 4-5-1, and 5-7-1 for the East Alpine, Main, and Mosel sub-basins with a total of 25, 31, and 50 parameters (weights and biases), respectively.

Efficiencies of the constructed ANN models in streamflow forecasting were checked with the mean of the RMSE normalized by the mean discharge values and the NSE obtained during the training and testing periods (Table 2) by running each model 100 times. The initial weights of each ANN model were randomly generated between -1 and $+1$.

4.2 Determination of distributions of model parameters

Regarding the results obtained from the optimal ANN models, it can be concluded that the obtained model structures are appropriate for forecasting of streamflows. Parametric uncertainty is related to the accuracy and representativeness of the model parameters, and in terms of "equifinality", different parameter sets can represent the behaviour of the modeled system equally. Prior distributions of the parameter sets were determined in the training period of 01.01.1993–31.12.1999 with the calibrated 5000 ANN models.

Fig. 1 *Box* plots of the parameters obtained from 5000 ANNs for the ▶ **a** East Alpine, **b** Main, and **c** Mosel sub-basins. The median values are shown as a *thick line* within each *box*. Corresponding descriptive statistics and parameters of the distributions are available in the Appendix

The obtained distributions and parameters can be seen in Fig. 1 for the three sub-basins. In these figures, for the sake of a clear illustration, in cases where the whiskers would extend more than three times the interquartile range (i.e., the difference between the third and the first quartile), they are truncated, and the remaining outliers are indicated by black circles.

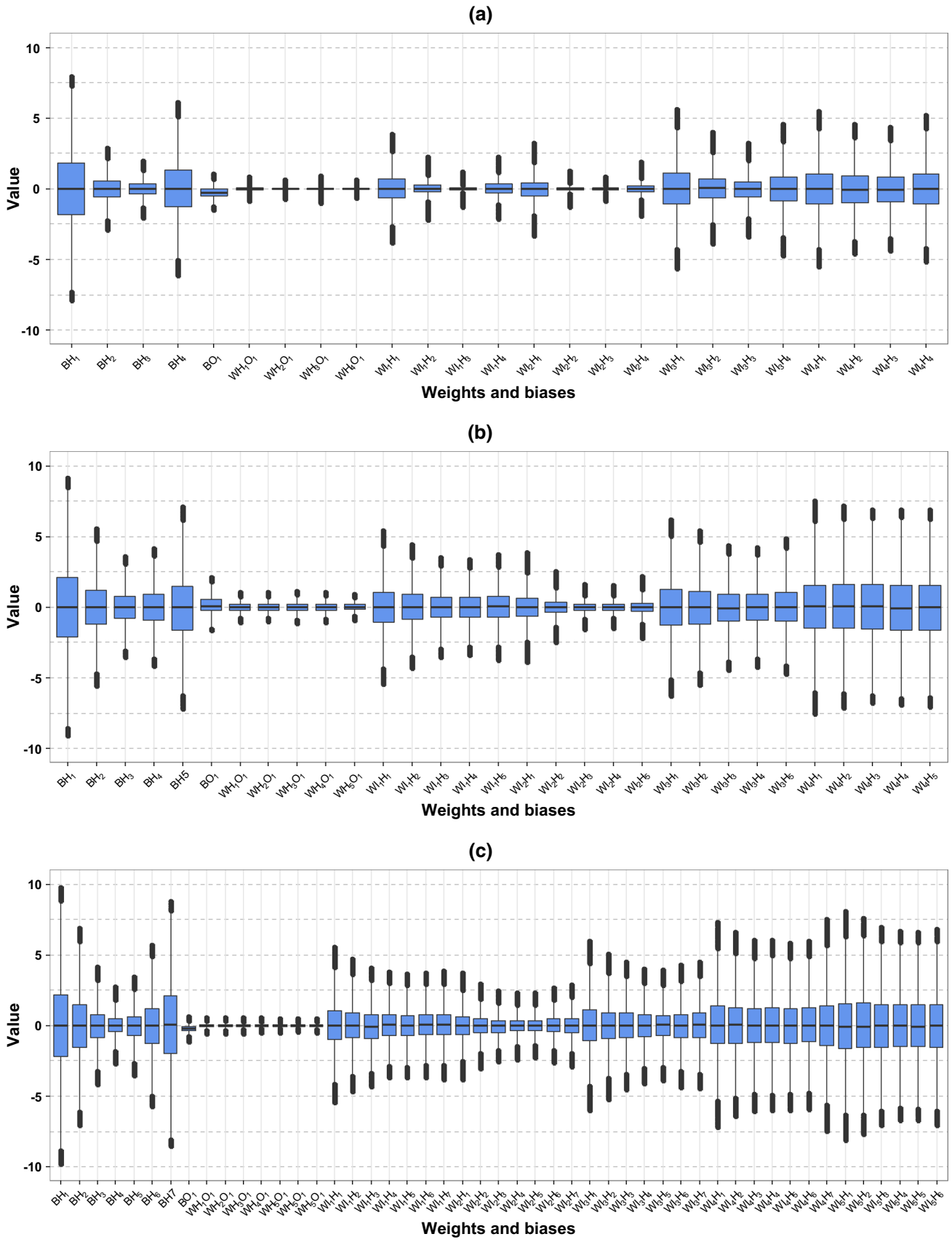
Most of the parameters are found to be following the *t* location-scale distribution (Elipot et al. 2016) with varying location, scale, and shape values except three out of 25 parameters for the East Alpine sub-basin and seven out of 50 parameters for the Mosel sub-basin that follow a logistic distribution. Here, BH_1(Fig. 1) is the bias parameter of hidden neuron#1, WI_1H_1 indicates the weight between input neuron#1 and hidden neuron#1, and WH_1O_1 is the weight between hidden neuron#1 and output neuron#1. As can be seen from the tables in the Appendix and Fig. 1, defining prior ranges for the weights and biases could be unfeasible because of the rather wide ranges and variations among the variables. Similarly to the normal distribution, the *t* and logistic distributions are symmetric and bell-shaped but have heavier tails meaning that they are more prone to producing values that fall far from their means. The main reason for obtaining distributions with heavy tails may be the large variability of flow values. For instance, flow values vary between 162 to 1167 $m^3\ s^{-1}$, 3.0 to 998 $m^3\ s^{-1}$, and 37.40 to 4020 $m^3\ s^{-1}$ for the East Alpine, Main, and Mosel sub-basins, respectively.

4.3 Uncertainty estimation of ANN models due to parametric uncertainty

Behavioural parameter sets can be explored with the help of Monte Carlo simulation methods. Usually, due to lack of prior distributions, parameter sets are uniformly generated from feasible ranges (Chen et al. 2013; Zhang and Li 2015). Latin hypercube sampling generates well-distributed parameters in a multi-dimensional space and

Table 2 Performances of the optimal ANN models during the training and testing periods

Sub-basins	Training			Testing		
	East Alpine	Main	Mosel	East Alpine	Main	Mosel
$RMSE/\bar{Q}$	0.026	0.256	0.239	0.016	0.241	0.208
NSE	0.989	0.874	0.963	0.995	0.870	0.938


(a)
Value
Weights and biases
(b)
Value
Weights and biases
(c)
Value
Weights and biases

reduces the number of runs (Jung et al. 2012). However, random sampling in the high-dimensional parameter space with significant correlations among the variables may not guarantee that even with a large number of simulations, the best parameter set can be found. Thus, in this paper, based on the correlation matrix obtained from the weights and biases of the ANN models, we generated 125,000 correlated normal random parameters from a multivariate normal distribution. Then, these normal random parameters were placed into the probability density function of the normal distribution to find the probabilities. Finally, in total 125,000 parameter sets were drawn from the estimated probability distributions based on these calculated probabilities without undergoing further training. To investigate the influence of the threshold value used in GLUE on the results, threshold values for NSE of 0.6 and 0.65 were selected (Gong et al. 2011). Predictive uncertainty in forecasted streamflow resulting from parametric uncertainty represented by 95% uncertainty intervals can be seen in Fig. 2, and the corresponding POC, ARIL, and AAD uncertainty measures are given in Table 3.

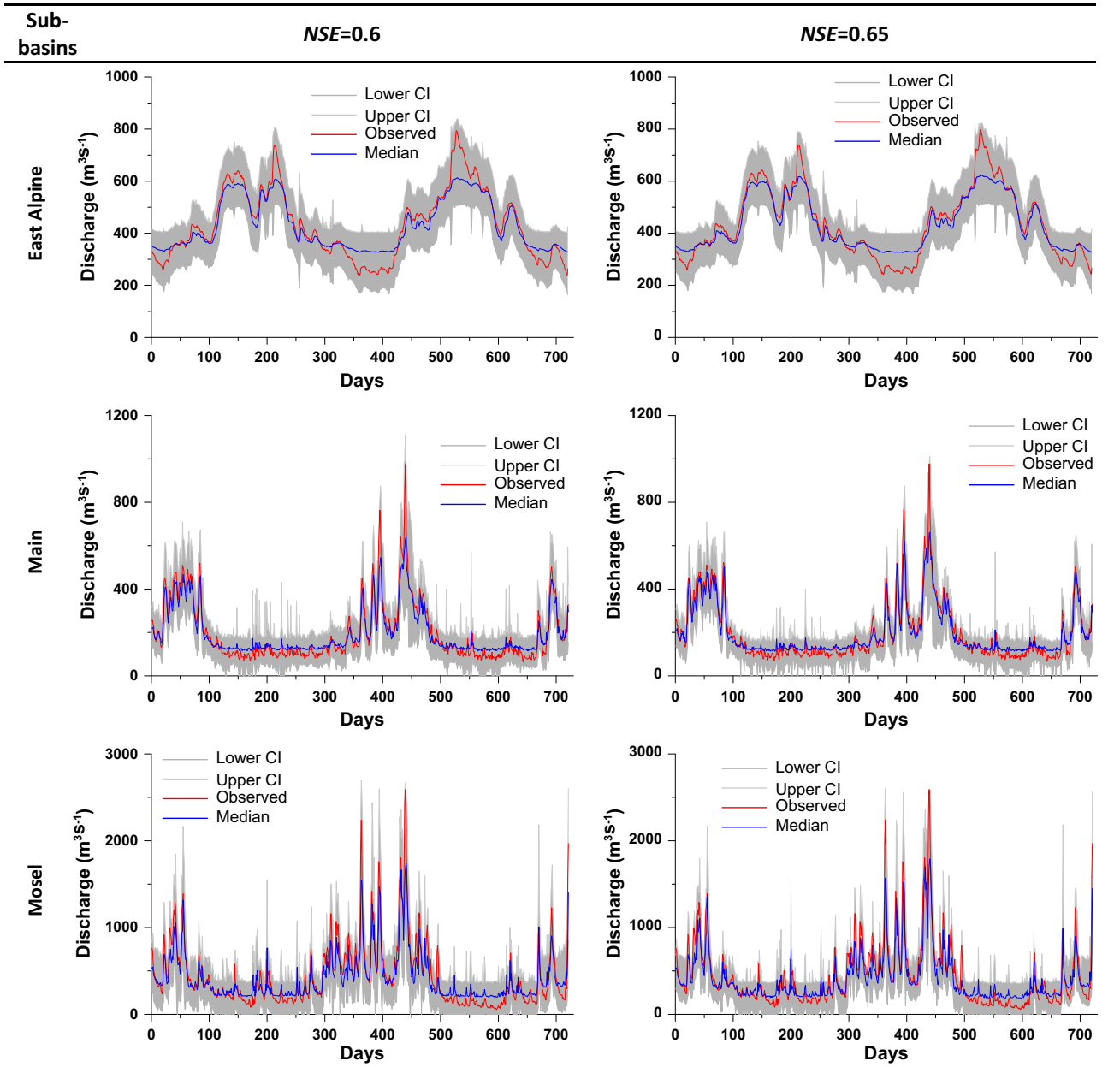


Fig. 2 Uncertainty in forecasted daily discharge due to parametric uncertainty of ANN models quantified by the GLUE method for a threshold value of 0.6 and 0.65

Significant information about the influence of parameters on the model outputs can be obtained through wide parameter ranges, but a large number of model simulations may result in a small number of behavioural parameter sets (Gong et al. 2011). A total of 1235, 87, and 72 behavioural parameter sets were obtained when the threshold value was equal to 0.6, and for the threshold value of 0.65, a total of 915, 49, and 41 behavioural parameter sets have been obtained for the East Alpine, Main, and Mosel sub-basins, respectively. Possible reasons for obtaining a relatively small number of behavioural parameter sets for the Main and Mosel sub-basins could be the complex dynamics of river flows that are difficult to be captured by an ANN model and a high number of correlated parameters with different ranges and distributions (Fig. 1). However, when using threshold values of 0.6 and 0.65, sufficient parameter sets ($\geq$40) were obtained (Gong et al. 2011). Furthermore, a decrease in the number of behavioural parameter sets was found as the threshold values increased, similarly to Gong et al. (2011) and Breinholt et al. (2013).

In the estimation of uncertainty bands and measures in Fig. 2 and Table 3, the intervals at the 95% confidence level were considered by finding the 2.5th and 97.5th percentiles of the associated distribution of the simulation results. According to Fig. 2, all models consistently well predict the dynamical behaviour of streamflow during the testing period. It is well known that ANNs are less capable of capturing peak flows, however, the models sufficiently generated 95% confidence intervals that capture extreme flows. To show the superiority of the proposed method for estimating the uncertainty bands, the bootstrap method that is based on resampling with replacement and the direct ensemble method were applied (Fig. 3; Table 4). In the former method, a set of 100 bootstrap samples of the available training data set was generated and ANN models were trained using these samples, then the performance of the trained ANN models was assessed through the observation pairs that are not included in a bootstrap sample in the testing phase. More information can be found in Tiwari and Chatterjee (2010). As indicated by Lee and Kang (2016), there is not a theoretical guideline to produce ensemble output for ANN models. To avoid including additional sources of uncertainty such as sampling uncertainty associated with the observations, the direct ensemble method employs the original 5000 ANN models that were trained with different random initial weights. These models were used for the evaluation of the ensemble output and to obtain uncertainty estimates (see Boucher et al. 2010; Lee and Kang 2016).

Table 3 Comparison of uncertainty estimates for different threshold values

Sub-basins	NSE = 0.6			NSE = 0.65		
	POC (%)	ARIL	AAD	POC (%)	ARIL	AAD
East Alpine	100	0.531	0.076	100	0.500	0.079
Main	99.31	1.319	0.115	98.89	1.144	0.133
Mosel	99.03	1.975	0.139	96.39	1.711	0.175

As can be seen from Fig. 3 and Table 4, most of the observed values from the direct ensemble method lie outside the confidence intervals, and very narrow uncertainty bands were obtained compared to the proposed GLUE method and the bootstrap method, similarly to Kumar et al. (2015) and Tiwari and Chatterjee (2010). In contrast to the first order uncertainty analysis and the bootstrap method conducted by Kasiviswanathan and Sudheer (2013), and Srivastav et al. (2007) respectively, peak flows were well captured by the GLUE and bootstrap methods. The bootstrap method is based on resampling of data with replacement and training an individual network on each resampled instance of the original training data set with different random initial weights by assuming this would lead to parametric uncertainty. In the ensemble method, only the uncertainty stemming from the training algorithm was considered since there is not any execution involved in the training phase such as shuffling data similarly to the bootstrap method. Only the outputs from 5000 ANNs were pooled together in order to generate an ensemble of hydrographs where uncertainty can be represented. Thus, we can conclude that the proposed method can be useful for the quantification of parametric uncertainty in ANN models.

For the East Alpine, Main, and Mosel sub-basins, the estimated percentages of coverage are 100, 99.31, and 99.03% for a threshold value of 0.6, and 100, 98.89, and 96.39% for a threshold value of 0.65, respectively. A better uncertainty estimate will result in a smaller average width with more values captured within the prediction band—maximum POC (Kasiviswanathan and Sudheer 2016). However, we preferred a better uncertainty interval with a POC coefficient close to the expected proportion (i.e., 95%) similarly to Zhang et al. (2009). While the obtained POC values for the GLUE method are 100, 98.89, and 96.39% for a threshold value of 0.65, POC values for the bootstrap method are 85.58, 96.53, and 91.40%. It is evident that the different approaches yielded various 95% uncertainty intervals. For instance, obtained uncertainty intervals from the bootstrap and direct ensemble methods are narrower than the GLUE method. However, the bootstrap method yielded better uncertainty intervals for the East Alpine basin and slightly better for the Main basin by considering the POC values. To some extent, one of the main drawbacks of the GLUE method is the subjectivity in choosing the likelihood function (Uniyal et al. 2015; Zhang et al. 2015). In our study, for the GLUE method, selection

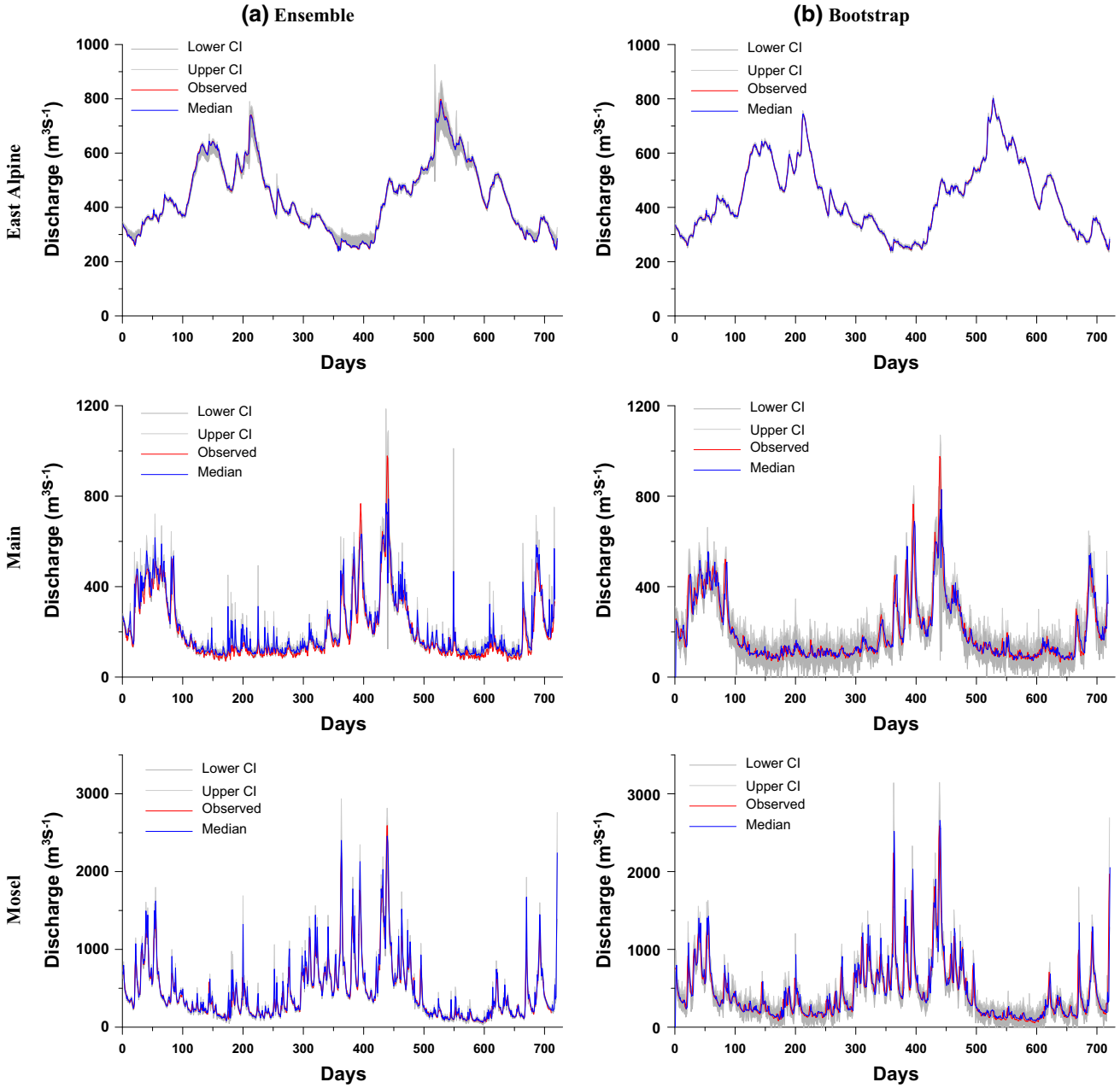


Fig. 3 Uncertainty in forecasted daily discharges from the ensemble and bootstrap methods

Table 4 Comparison of uncertainty estimates for the ensemble and bootstrap methods

Sub-basins	Ensemble			Bootstrap		
	POC (%)	ARIL	AAD	POC (%)	ARIL	AAD
East Alpine	82.80	0.075	0.295	85.58	0.036	0.270
Main	46.03	0.318	0.597	96.53	0.979	0.168
Mosel	55.62	0.286	0.533	91.40	0.980	0.209

of 0.6 and 0.65 as threshold values is directly related to the sufficient number of behavioural parameter sets (≥ 40) to construct the uncertainty bands (Gong et al. 2011). To investigate the performance of the GLUE analysis, threshold values of 0.8 and 0.89 were also used (not shown) only for the East Alpine basin since it has a sufficient number of behavioural parameters. 217 and 41 behavioural parameter sets were obtained for the East Alpine basin, respectively. While narrower uncertainty bands have been

achieved with these values of NSE, no significant performance gain was obtained for the uncertainty indices except ARIL. The POC, ARIL, and AAD values are 100%, 0.378, and 0.103 for a threshold value of 0.8 and 99.44%, 0.255, and 0.142 for a threshold value of 0.89. Thus, one possible reason for obtaining the relatively small number of behavioural parameters and wider uncertainty intervals could be the proposed sampling methodology. In future work, the efficiencies of other sampling algorithms such as evolutionary Monte Carlo algorithm (Liang 2005)–an advanced Markov Chain Monte Carlo method–will be investigated.

The highest uncertainty arising from parametric uncertainty was obtained for Mosel sub-basin and the lowest for the East Alpine sub-basins. This can also be seen from the other uncertainty measures, i.e. ARIL, and AAD. The highest ARIL value was obtained for the Mosel and the lowest for the East Alpine for threshold values of 0.6 and 0.65. AAD values less than 0.5 imply a high degree of symmetry around observed values, in contrary values larger than 0.5 denote more asymmetrical uncertainty bands. The highest values of AAD were obtained for the Mosel for both threshold values (Table 3), but they are all smaller than 0.2 indicating a high degree of symmetry. Our results are in line with Jin et al. (2010) who employed a simple conceptual water balance model with the use of GLUE and a Bayesian method to quantify the parameter uncertainty of monthly runoff. The authors found ARIL and POC values of 1.012 and 91% for GLUE and 1.169 and 89% for the Bayesian method and these are close to ours. Furthermore, Kasiviswanathan and Sudheer (2016) obtained values of 98.14, 46.58, and 97.17% for POC using the bootstrap, Bayesian, and prediction interval methods and employing an ANN model by considering both rainfall and discharge values as input.

Chen et al. (2013) employed the GLUE method for three conceptual hydrological models for daily streamflow simulation (for the period of 1960–2011) for a headwater region of the Yellow River, China and obtained ARIL values of 1.34, 1.21, and 0.91 for a threshold value for NSE of 0.6 and 1.13, 0.98, and 0.73 for a threshold value for NSE of 0.7. They found AAD values of 0.33, 0.43, and 0.53 for a threshold value for NSE of 0.6 and 0.43, 0.54, and 0.71 for a threshold value for NSE of 0.7. Comparatively, it is clear that the ARIL and AAD values obtained from the GLUE method for the ANN models (Table 3) are promising.

The main reason for obtaining the highest prediction uncertainty resulting from parametric uncertainty for the Mosel sub-basin could stem from the basin dynamics and differences in the characteristics of the training and testing periods. The mean annual hydrograph of the sub-basins could be informative for the description of basin characteristics as well as annual variations in temperature, precipitation, and snowpack (Kunkel and Pierce 2010). The Mosel and Main are rainfall-dominated sub-basins while snowmelt processes are dominant in the East Alpine (Demirel et al. 2013b). The annual hydrograph of the East Alpine (Fig. 4) is a "simple" hydrograph that shows clear rising and falling limbs, whereas the Mosel and Main show "complex" hydrograph features with fluctuations in both rising and falling limbs (Kunkel and Pierce 2010). The corresponding coefficients of variation (defined as the standard deviation divided by the mean) for the training period are 0.42, 0.72, and 1.25 and for the testing period are 0.30, 0.69, and 0.83 for the East Alpine, Main, and Mosel, respectively. Also, the highest interquartile ranges were obtained for the Mosel sub-basin as 273.5 and 210.6 $m^3\ s^{-1}$ in the training and testing periods relatively to 168 and 195 $m^3\ s^{-1}$, 112 and 141.3 $m^3\ s^{-1}$ in the training and testing periods for the East Alpine and Main sub-basins, respectively.

These characteristics show that the uncertainty is proportional to the variability of the flow similarly to Kasiviswanathan and Sudheer (2013). As noted by Tian et al. (2014), this is mainly due to the fact that less similar observations are found as the discharge becomes more extreme, and lack of information leads to larger uncertainty ranges. Furthermore, the flow regime of the Rhine River is influenced by snowmelt and precipitation runoff from the Alps in the summer months and by precipitation runoff from the uplands in winter that leads to the average discharge maxima shift from summer to winter when moving downstream the Rhine (Disse and Engel 2001). Further away from the Alps, the influences of the oceanic climate and less anticipated precipitation show themselves as stochastic variations in monthly hydrographs (Uehlinger et al. 2009). Thus, more irregular flows can be observed in the Main and Mosel sub-basins that are influenced by both rainfall and snowmelt. Additionally, losses, attenuation, ponding, surface and subsurface storages, and travel time

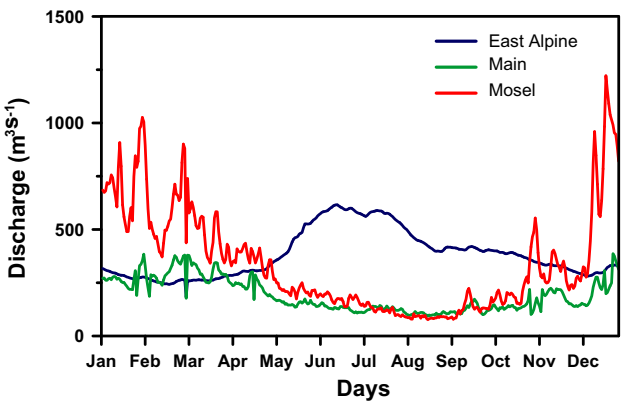


Fig. 4 Long-term mean annual hydrographs for the East Alpine, Main, and Mosel sub-basins for the period of 01.01.1993–31.12.1999

can be influencing factors for the uncertainties in the outputs of the ANN models. Further, socio-economic development and regulation along the Rhine River have contributed to stochastic variations in monthly hydrographs by deteriorating riverine habitats, and by changing the morphological dynamics, i.e. the Main and Mosel sub-basins are more regulated and industrialized than the East Alpine sub-basin. Finally, the Wilcoxon–Mann–Whitney test (Wilcoxon 1945) can be used to identify whether two sub-sets of a time series are taken from the same distribution. The results (not shown) indicated that there are significant differences between training and testing periods of the East Alpine and Mosel sub-basins. Thus, having the highest coefficient of variation and showing the most complex-hydrograph features, the highest prediction uncertainty resulting from parametric uncertainty was obtained for the Mosel sub-basin.

As noted by Chen et al. (2013), decreasing ARIL, POC, and increasing AAD as a function of NSE, could be a clue for the sensitivity of GLUE estimates to the selection of the threshold value of the NSE. To confirm this, we applied the Wilcoxon–Mann–Whitney test (Wilcoxon 1945) for the lower and upper prediction intervals for the threshold values 0.6 and 0.65 at a 5% significance level. In this test, the null hypothesis is that the lower and upper percentiles for different threshold values come from identical populations, and there is not any statistically meaningful difference between them at a 5% significance level. All the obtained p values except for the lower prediction intervals of the Main basin are lower than 0.05 (not shown here), and the null hypothesis is rejected for all rivers. This indicates that the GLUE uncertainty bands are sensitive to the selection of the threshold value for the studied rivers similarly to Chen et al. (2013) and Jin et al. (2010).

Finally, it should be noted that parametric uncertainty is directly related to the model structure. In this study, we have shown that the stepwise regression method could be employed to determine the optimal input structure among the input variables. Some studies, such as Wang et al. (2013) and Tiwari and Chatterjee (2011), selected the input structure of ANN models only based on the partial autocorrelation function. We intend to investigate the model structure uncertainty resulting from well-recognized input determination methods, such as the partial mutual information function (May et al. 2008), and the gamma test (Stefánsson et al. 1997), in future work.

5 Conclusions

Quantification of uncertainty resulting from ANN model parameters is helpful for modelers and policy makers while improving the reliability of forecasts. The uncertainty in the parameters that govern the functional relationship between input and output data is a significant constituent of the prediction uncertainty of ANNs' outputs (Kingston et al. 2005). This study employs the Generalized Likelihood Uncertainty Estimation (GLUE) approach to quantify the parametric uncertainty of ANN models for three sub-basins of the Rhine Basin, namely the East Alpine, Main, and Mosel. The input structures of ANN models were determined with the stepwise regression method considering precipitation, evapotranspiration, temperature, and discharge variables. The final ANN structures were found to be 4-4-1 (input-hidden-output neurons), 4-5-1, and 5-7-1 for the East Alpine, Main, and Mosel, respectively. The prior distributions for the weight and bias parameters of ANN models were determined in the calibration period of 01.01.1993–31.12.1999 with the calibrated 5000 ANN models and using the Bayesian Information Criterion. One of the largest parameter sets for ANN models in the literature—in total 125,000 samples for each model—were drawn from these estimated probability distributions to quantify the uncertainty intervals. The influence of the threshold level to distinguish behavioural and non-behavioural parameter sets in GLUE was tested by selecting two threshold values of 0.6 and 0.65 for the Nash–Sutcliffe criterion. The results of the GLUE method were compared with the direct ensemble and the bootstrap techniques. The results showed that the GLUE method successfully captured the observed discharges with the generated prediction intervals, especially the peak flows. By employing the Wilcoxon–Mann–Whitney test, it was found that uncertainty bands are sensitive to the selection of the threshold value for the studied rivers. Uncertainty measures such as the percentage of coverage, average relative length, and average asymmetry degree were used to present the behaviour of parametric uncertainty for three sub-basins with different hydro-climatic characteristics. The highest prediction uncertainty due to parametric uncertainty was obtained for the Mosel and the smallest one for the East Alpine sub-basin. This is mainly because the Mosel sub-basin is a rainfall-dominated sub-basin with the highest coefficient of variation for streamflow and showing complex-hydrograph features with fluctuations in rising and falling limbs, whereas the East Alpine sub-basin has the smallest coefficient of variation and shows relatively simple hydrograph features with clear rising and falling limbs. Overall, the novelty of this paper is that it addresses the "pure" parametric uncertainty of ANN models through the modified GLUE technique without complex calculations and without altering existing codes of ANN models. Furthermore, the method utilizes correlated parameter sets defined from prior distributions to generate uncertainty bands, contrary to Jin et al. (2010) who argued that the GLUE methodology is not appropriate for parameter sets

that have correlated parameters. Additionally, we compared the obtained uncertainty bands from the GLUE method with the direct ensemble and the bootstrap method in three river basins that have distinct hydro-climatic features to the best of our knowledge for the first time. Some limitations and future directions related to the proposed methodology are:

1. The model structure uncertainty was not addressed in this study, and it would be worthwhile to quantify both parametric and model structure uncertainties besides the effect of different training algorithms on ANN outputs with the GLUE methodology.
2. A relatively small number of behavioural parameters was obtained for the Main and Mosel basins. The main reason for this could be that relatively high ranges of the parameters and a high number of outliers were obtained in the training phase. One possible solution could be employing outlier analysis (Ng et al. 2007) to remove outliers that complicate the analyses.
3. We only considered one-day-ahead forecasts of the discharge. Confidence intervals for multi-day-ahead (different lead times) forecasts could be investigated in a future study employing the proposed methodology.
4. Finally, having gained more interest in recent years, copulas (e.g., Klein et al. 2016; Zhang et al. 2016) could be integrated into the GLUE framework to generate multivariate joint probability distributions for generating parameter sets.

Acknowledgement The authors thank two anonymous reviewers for their constructive comments that improved the paper considerably.

Appendix 1

Obtained distributions and parameters of the ANN models for the East Alpine, Main, and Mosel sub-basins from the 5000 ANNs during the training period can be found below Tables 5, 6, 7.

Table 5 Characteristics of probability distribution functions of ANN parameters for the East Alpine sub-basin

Parameter	Min.	Max.	Std.	Distribution type	μ-location	σ-scale	v-shape
BH_1	−1349.631	319.232	25.234	t-location scale	0.004	2.208	1.548
BH_2	−211.046	1374.895	22.219	Logistic	0.148	0.490	–
BH_3	−1182.187	455.573	20.104	t-location scale	−0.012	0.453	0.916
BH_4	−513.615	522.634	15.004	t-location scale	0.022	1.512	1.328
WI_1H_1	−1320.070	392.985	27.045	t-location scale	0.008	0.759	0.798
WI_1H_2	−1101.213	757.575	21.483	t-location scale	0	0.072	0.350
WI_1H_3	−512.236	1073.786	19.932	Logistic	0.097	0.281	–
WI_1H_4	−472.254	427.343	13.091	t-location scale	0.004	0.229	0.550
WI_2H_1	−1491.685	580.553	26.430	t-location scale	−0.003	0.515	0.676
WI_2H_2	−251.849	1469.365	23.615	Logistic	0.140	0.408	–
WI_2H_3	−965.518	224.522	18.208	t-location scale	0	0.028	0.319
WI_2H_4	−109.449	516.139	12.598	t-location scale	0	0.127	0.436
WI_3H_1	−926.510	248.579	23.453	t-location scale	0.016	1.366	0.973
WI_3H_2	−455.008	735.731	22.739	t-location scale	0.017	0.868	0.871
WI_3H_3	−1752.432	759.109	29.723	t-location scale	−0.005	0.728	0.857
WI_3H_4	−588.107	424.568	21.086	t-location scale	−0.013	1.080	0.957
WI_4H_1	−2802.096	662.573	44.141	t-location scale	−0.010	1.471	1.065
WI_4H_2	−551.249	749.541	20.278	t-location scale	−0.062	1.261	1.163
WI_4H_3	−744.712	516.253	17.227	t-location scale	−0.039	1.212	1.171
WI_4H_4	−420.231	586.536	18.178	t-location scale	0.003	1.388	1.140
BO_1	−151.337	144.314	6.539	t-location scale	−0.291	0.296	0.955
WH_1O_1	−144.686	267.398	6.683	t-location scale	0	0.027	0.331
WH_2O_1	−85.534	74.219	2.778	t-location scale	0	0.026	0.325
WH_3O_1	−20.937	18.049	1.920	t-location scale	0	0.023	0.301
WH_4O_1	−224.791	116.138	5.125	t-location scale	0	0.020	0.314

Table 6 Characteristics of probability distribution functions of ANN parameters for the Main sub-basin

Parameter	Min.	Max.	Std.	Distribution type	μ-location	σ-scale	ν-shape
BH_1	−683.212	589.537	24.127	*t*-location scale	−0.008	2.629	1.375
BH_2	−320.791	1699.428	34.644	*t*-location scale	0.019	1.542	1.270
BH_3	−349.429	397.705	13.279	*t*-location scale	0.017	0.989	1.250
BH_4	−619.337	982.353	20.652	*t*-location scale	−0.017	1.195	1.302
BH_5	−682.059	1030.992	27.118	*t*-location scale	−0.032	1.948	1.423
WI_1H_1	−924.755	432.781	23.427	*t*-location scale	0.006	1.368	1.002
WI_1H_2	−458.918	1338.989	27.794	*t*-location scale	0.012	1.159	1.098
WI_1H_3	−504.678	627.903	16.825	*t*-location scale	0.006	0.964	1.151
WI_1H_4	−640.230	202.965	13.568	*t*-location scale	−0.005	0.927	1.148
WI_1H_5	−474.062	432.346	18.661	*t*-location scale	0.017	0.971	1.036
WI_2H_1	−658.149	758.559	20.384	*t*-location scale	−0.002	0.727	0.765
WI_2H_2	−215.007	611.492	15.141	*t*-location scale	−0.005	0.453	0.725
WI_2H_3	−410.246	294.656	10.761	*t*-location scale	−0.002	0.304	0.695
WI_2H_4	−891.969	435.382	16.073	*t*-location scale	0.005	0.298	0.685
WI_2H_5	−471.137	500.351	15.746	*t*-location scale	−0.002	0.403	0.728
WI_3H_1	−543.759	830.040	26.043	*t*-location scale	0.004	1.624	1.035
WI_3H_2	−739.435	1158.707	28.521	*t*-location scale	−0.008	1.493	1.264
WI_3H_3	−787.871	321.552	16.186	*t*-location scale	−0.061	1.215	1.286
WI_3H_4	−880.197	613.996	22.631	*t*-location scale	0.025	1.188	1.317
WI_3H_5	−217.318	1654.076	31.115	*t*-location scale	0.054	1.324	1.306
WI_4H_1	−682.527	715.107	28.028	*t*-location scale	0.012	1.955	1.050
WI_4H_2	−275.794	1358.049	27.385	*t*-location scale	0.042	1.923	1.306
WI_4H_3	−678.348	1162.095	24.345	*t*-location scale	0.044	1.828	1.538
WI_4H_4	−1393.968	1767.874	35.365	*t*-location scale	−0.091	1.867	1.623
WI_4H_5	−779.642	1818.805	34.131	*t*-location scale	−0.029	1.887	1.456
BO_1	−79.253	156.389	6.694	*t*-location scale	0.090	0.457	1.337
WH_1O_1	−125.346	156.567	6.434	*t*-location scale	−0.006	0.227	0.894
WH_2O_1	−51.990	83.112	2.578	*t*-location scale	0.006	0.271	1.273
WH_3O_1	−16.814	40.455	1.286	*t*-location scale	−0.012	0.332	1.705
WH_4O_1	−28.173	20.683	1.085	*t*-location scale	0	0.315	1.743
WH_5O_1	−43.071	60.129	1.677	*t*-location scale	0.002	0.244	1.235

Table 7 Characteristics of probability distribution functions of ANN parameters for the Mosel sub-basin

Parameter	Min.	Max.	Std.	Distribution type	μ-location	σ-scale	ν-shape
BH_1	−2886.613	856.977	58.737	*t*-location scale	0.009	2.640	1.185
BH_2	−5071.765	874.094	88.089	*t*-location scale	−0.029	1.768	1.172
BH_3	−2353.503	2907.875	64.321	*t*-location scale	−0.009	0.995	1.026
BH_4	−1799.962	544.568	31.301	*t*-location scale	0.006	0.630	0.864
BH_5	−1349.263	1170.563	36.905	*t*-location scale	−0.013	0.853	1.006
BH_6	−1689.360	4806.102	88.984	Logistic	1.054	1.578	–
BH_7	−767.212	3135.727	75.960	*t*-location scale	0.046	2.304	1.450
WI_1H_1	−7258.186	851.701	107.736	*t*-location scale	−0.002	1.315	0.974
WI_1H_2	−992.160	579.006	30.938	*t*-location scale	0.006	1.140	1.060
WI_1H_3	−1045.615	11,631.250	167.692	Logistic	−0.044	4.803	–
WI_1H_4	−606.228	1371.066	25.852	*t*-location scale	0.019	0.988	1.133
WI_1H_5	−2832.963	1683.853	57.462	Logistic	0.067	0.832	–
WI_1H_6	−5960.588	2806.027	100.869	*t*-location scale	0.031	0.939	1.045
WI_1H_7	−1967.980	4524.261	75.098	*t*-location scale	0.037	0.949	1.029
WI_2H_1	−3232.541	3986.939	84.026	*t*-location scale	0.013	0.770	0.751

Table 7 continued

Parameter	Min.	Max.	Std.	Distribution type	μ-location	σ-scale	v-shape
WI_2H_2	−1259.663	1195.962	47.691	t-location scale	−0.009	0.588	0.750
WI_2H_3	−386.260	28,733.715	407.987	t-location scale	−0.013	0.516	0.801
WI_2H_4	−1073.225	419.939	20.625	t-location scale	−0.003	0.471	0.784
WI_2H_5	−1455.134	1232.454	31.261	t-location scale	0.004	0.451	0.798
WI_2H_6	−4178.986	977.456	66.826	t-location scale	0.001	0.549	0.855
WI_2H_7	−381.879	12,376.559	188.037	Logistic	−0.660	4.953	–
WI_3H_1	−3119.147	1031.327	63.249	t-location scale	0.010	1.368	0.831
WI_3H_2	−3628.393	906.327	81.645	t-location scale	−0.018	1.177	0.880
WI_3H_3	−1574.110	37,213.235	527.875	Logistic	5.782	5.878	–
WI_3H_4	−1481.417	827.796	33.485	t-location scale	−0.005	0.956	0.941
WI_3H_5	−2433.331	562.668	45.677	t-location scale	0.031	0.944	0.977
WI_3H_6	−2857.293	1324.393	59.992	t-location scale	−0.025	1.016	0.974
WI_3H_7	−3860.969	1588.180	67.511	t-location scale	0.032	1.092	0.993
WI_4H_1	−2256.734	4728.910	112.160	t-location scale	0.008	1.754	0.858
WI_4H_2	−600.348	2156.845	58.668	Logistic	−0.273	1.246	–
WI_4H_3	−90,549.263	2010.136	1281.771	t-location scale	−0.008	1.607	1.010
WI_4H_4	−1057.643	2387.116	47.866	t-location scale	0.041	1.598	1.028
WI_4H_5	−1039.639	2136.490	48.305	t-location scale	−0.0053	1.582	1.069
WI_4H_6	−5962.084	4528.906	122.048	t-location scale	0.021	1.576	1.032
WI_4H_7	−834.956	3763.819	64.305	t-location scale	−0.014	1.620	1.020
WI_5H_1	−2020.975	2196.225	68.694	t-location scale	−0.083	2.067	0.967
WI_5H_2	−3575.283	1395.080	69.420	t-location scale	−0.006	2.049	1.073
WI_5H_3	−15,562.983	1102.789	227.893	t-location scale	−0.024	1.947	1.200
WI_5H_4	−2138.042	822.239	42.914	t-location scale	0.009	1.839	1.212
WI_5H_5	−787.467	2316.485	62.468	Logistic	0.784	0.867	–
WI_5H_6	−3125.135	3220.543	89.209	t-location scale	−0.040	1.895	1.197
WI_5H_7	−3371.017	2265.586	69.618	t-location scale	0.041	2.058	1.206
BO_1	−1445.138	153.180	31.591	t-location scale	−0.205	0.176	0.702
WH_1O_1	−1445.016	537.635	26.993	t-location scale	0.001	0.095	0.626
WH_2O_1	−1017.548	314.163	16.068	t-location scale	−0.001	0.106	0.736
WH_3O_1	−12.908	71.698	1.515	t-location scale	−0.002	0.109	0.803
WH_4O_1	−30.426	64.610	1.230	t-location scale	0.001	0.091	0.699
WH_5O_1	−8.934	39.661	0.877	t-location scale	0.001	0.084	0.675
WH_6O_1	−35.817	41.999	1.435	t-location scale	0.001	0.080	0.682
WH_7O_1	−104.115	113.796	4.616	t-location scale	0.001	0.087	0.684

Appendix 2

The formulas of the indices for assessing the prediction bounds can be found below Table 8:

where n is the number of time steps used for constructing the prediction bands and c_i denotes whether the corresponding observation is covered by the prediction band being 1 or 0 if the observed value is contained in the prediction band or not, L_t^{upper} and L_t^{lower} are the upper and lower prediction boundary values of the 95% confidence interval, and Q_t is the observed value at time t.

Table 8 Indices for assessing the prediction bounds

Percentage of coverage (POC)	$POC = \left(\frac{1}{n}\sum_{i=1}^{n} c_i\right) \times 100$
Average relative length (ARIL)	$ARIL = \frac{1}{n}\sum_{t=1}^{n} \frac{L_t^{upper} - L_t^{lower}}{Q_t}$
Average asymmetry degree (AAD)	$AAD = \frac{1}{n}\sum_{t=1}^{n} \left\lvert \frac{L_t^{upper} - Q_t}{L_t^{upper} - L_t^{lower}} - 0.5 \right\rvert$

References

Aqil M, Kita I, Yano A, Nishiyama S (2007) Neural networks for real time catchment flow modeling and prediction. Water Resour Manag 21(10):1781–1796. doi:10.1007/s11269-006-9127-y

Badrzadeh H, Sarukkalige R, Jayawardena AW (2013) Impact of multi-resolution analysis of artificial intelligence models inputs on multi-step ahead river flow forecasting. J Hydrol 507:75–85. doi:10.1016/j.jhydrol.2013.10.017

Badrzadeh H, Sarukkalige R, Jayawardena A (2016) Improving ann-based short-term and long-term seasonal river flow forecasting with signal processing techniques. River Res Appl 32(3):245–256

Beven K, Binley A (1992) The future of distributed models: model calibration and uncertainty prediction. Hydrol Process 6(3):279–298. doi:10.1002/hyp.3360060305

Beven K, Freer J (2001) Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J Hydrol 249(1):11–29

Boucher M-A, Laliberté J-P, Anctil F (2010) An experiment on the evolution of an ensemble of neural networks for streamflow forecasting. Hydrol Earth Syst Sci 14(3):603–612

Breinholt A, Grum M, Madsen H, Örn Thordarson F, Mikkelsen PS (2013) Informal uncertainty analysis (GLUE) of continuous flow simulation in a hybrid sewer system with infiltration inflow–consistency of containment ratios in calibration and validation? Hydrol Earth Syst Sci 17(10):4159–4176. doi:10.5194/hess-17-4159-2013

Chang F-J, Chiang Y-M, Chang L-C (2007) Multi-step-ahead neural networks for flood forecasting. Hydrol Sci J 52(1):114–130

Chen X, Yang T, Wang X, Xu C-Y, Yu Z (2013) Uncertainty intercomparison of different hydrological models in simulating extreme flows. Water Resour Manag 27(5):1393–1409

Chiang Y-M, Chang L-C, Chang F-J (2004) Comparison of static-feedforward and dynamic-feedback neural networks for rainfall–runoff modeling. J Hydrol 290(3):297–311

De Villiers J, Barnard E (1993) Backpropagation neural nets with one and two hidden layers. IEEE Trans Neural Netw 4(1):136–141

Demirel MC, Booij MJ, Hoekstra AY (2013a) Effect of different uncertainty sources on the skill of 10 day ensemble low flow forecasts for two hydrological models. Water Resour Res 49(7):4035–4053

Demirel MC, Booij MJ, Hoekstra AY (2013b) Impacts of climate change on the seasonality of low flows in 134 catchments in the River Rhine basin using an ensemble of bias-corrected regional climate simulations. Hydrol Earth Syst Sci 17(10):4241–4257. doi:10.5194/hess-17-4241-2013

Disse M, Engel H (2001) Flood events in the Rhine basin: genesis, influences and mitigation. Nat Hazards 23(2–3):271–290

Elipot S, Lumpkin R, Perez RC, Lilly JM, Early JJ, Sykulski AM (2016) A global surface drifter data set at hourly resolution. J Geophys Res Oceans 121(5):2937–2966. doi:10.1002/2016JC011716

Ghavidel SZZ, Montaseri M (2014) Application of different data-driven methods for the prediction of total dissolved solids in the Zarinehroud basin. Stoch Env Res Risk Assess 28(8):2101–2118

Gong Y, Shen Z, Hong Q, Liu R, Liao Q (2011) Parameter uncertainty analysis in watershed total phosphorus modeling using the GLUE methodology. Agric Ecosyst Environ 142(3):246–255

Hagan MT, Menhaj MB (1994) Training feedforward networks with the Marquardt algorithm. IEEE Trans Neural Netw 5(6):989–993

Hornberger GM, Spear R (1981) Approach to the preliminary analysis of environmental systems. J Environ Manag 12(1):7–18

Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. Neural Netw 2(5):359–366

Hurkmans R, Troch P, Uijlenhoet R, Moors E (2007) Simulating Rhine River discharges using a land surface model. In: CAIWA conference vol. 1215

Jain A, Sudheer K, Srinivasulu S (2004) Identification of physical processes inherent in artificial neural network rainfall runoff models. Hydrol Process 18(3):571–581

Jin X, Xu C-Y, Zhang Q, Singh V (2010) Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. J Hydrol 383(3):147–155

Jothiprakash V, Magar RB (2012) Multi-time-step ahead daily and hourly intermittent reservoir inflow prediction by artificial intelligent techniques using lumped and distributed data. J Hydrol 450–451:293–307. doi:10.1016/j.jhydrol.2012.04.045

Jung I-W, Moradkhani H, Chang H (2012) Uncertainty assessment of climate change impacts for hydrologically distinct river basins. J Hydrol 466–467:73–87

Kasiviswanathan KS, Sudheer KP (2013) Quantification of the predictive uncertainty of artificial neural network based river flow forecast models. Stoch Env Res Risk Assess 27(1):137–146. doi:10.1007/s00477-012-0600-2

Kasiviswanathan K, Sudheer K (2016) Comparison of methods used for quantifying prediction interval in artificial neural network hydrologic models. Model Earth Syst Environ 2(1):1–11

Kasiviswanathan KS, Sudheer KP, He J (2016) Quantification of prediction uncertainty in artificial neural network models. In: Shanmuganathan S, Samarasinghe S (eds) Artificial neural network modelling. Springer, Cham, pp 145–159

Khan MS, Coulibaly P (2006) Bayesian neural network for rainfall-runoff modeling. Water Resour Res. doi:10.1029/2005WR003971

Kingston GB, Lambert MF, Maier HR (2005) Bayesian training of artificial neural networks used for water resources modeling. Water Resour Res. doi:10.1029/2005WR004152

Klein B, Meissner D, Kobialka H-U, Reggiani P (2016) Predictive uncertainty estimation of hydrological multi-model ensembles using pair-copula construction. Water 8(4):125

Kumar S, Tiwari MK, Chatterjee C, Mishra A (2015) Reservoir inflow forecasting using ensemble models based on neural networks, wavelet analysis and bootstrap method. Water Resour Manag 29(13):4863–4883

Kunkel ML, Pierce JL (2010) Reconstructing snowmelt in Idaho's watershed using historic streamflow records. Clim Change 98(1–2):155–176

Lee D-H, Kang D-S (2016) The application of the artificial neural network ensemble model for simulating streamflow. Procedia Eng 154:1217–1224. doi:10.1016/j.proeng.2016.07.434

Liang F (2005) Bayesian neural networks for nonlinear time series forecasting. Stat Comput 15(1):13–29

Lohani AK, Goel N, Bhatia K (2011) Comparative study of neural network, fuzzy logic and linear transfer function techniques in daily rainfall-runoff modelling under different input domains. Hydrol Process 25(2):175–193

Maier HR, Dandy GC (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ Model Softw 15(1):101–124. doi:10.1016/S1364-8152(99)00007-9

May RJ, Maier HR, Dandy GC, Fernando TG (2008) Non-linear variable selection for artificial neural networks using partial mutual information. Environ Model Softw 23(10):1312–1326

Melesse AM, Ahmad S, McClain ME, Wang X, Lim YH (2011) Suspended sediment load prediction of river systems: an artificial neural network approach. Agric Water Manag 98(5):855–866. doi:10.1016/j.agwat.2010.12.012

Middelkoop H, Daamen K, Gellens D, Grabs W, Kwadijk JCJ, Lang H, Parmet BWAH, Schädler B, Schulla J, Wilke K (2001) Impact of climate change on hydrological regimes and water resources management in the Rhine basin. Clim Change 49(1):105–128. doi:10.1023/a:1010784727448

Mirzaei M, Huang YF, El-Shafie A, Shatirah A (2015) Application of the generalized likelihood uncertainty estimation (GLUE) approach for assessing uncertainty in hydrological models: a review. Stoch Env Res Risk Assess 29(5):1265–1273

Ng WW, Panu US, Lennox WC (2007) Chaos based analytical techniques for daily extreme hydrological observations. J Hydrol 342:17–41

Noori R, Hoshyaripour G, Ashrafi K, Araabi BN (2010) Uncertainty analysis of developed ANN and ANFIS models in prediction of carbon monoxide daily concentration. Atmos Environ 44(4):476–482

Papadopoulos G, Edwards PJ, Murray AF (2001) Confidence estimation methods for neural networks: a practical comparison. IEEE Trans Neural Netw 12(6):1278–1287

Pramanik N, Panda RK (2009) Application of neural network and adaptive neuro-fuzzy inference systems for river flow prediction. Hydrol Sci J 54(2):247–260. doi:10.1623/hysj.54.2.247

Ranjithan S, Eheart J, Garrett J (1993) Neural network-based screening for groundwater reclamation under uncertainty. Water Resour Res 29(3):563–574

Rogiers B, Mallants D, Batelaan O, Gedeon M, Huysmans M, Dassargues A (2012) Estimation of hydraulic conductivity and its uncertainty from grain-size data using GLUE and artificial neural networks. Math Geosci 44(6):739–763

Selle B, Hannah M (2010) A bootstrap approach to assess parameter uncertainty in simple catchment models. Environ Model Softw 25(8):919–926

Singh KP, Basant A, Malik A, Jain G (2009) Artificial neural network modeling of the river water quality—a case study. Ecol Model 220(6):888–895

Spear R, Hornberger G (1980) Eutrophication in Peel Inlet—II. Identification of critical uncertainties via generalized sensitivity analysis. Water Res 14(1):43–49

Srivastav RK, Sudheer KP, Chaubey I (2007) A simplified approach to quantifying predictive and parametric uncertainty in artificial neural network hydrologic models. Water Resour Res 43(10): W10407. doi:10.1029/2006WR005352

Stefánsson A, Končar N, Jones AJ (1997) A note on the gamma test. Neural Comput Appl 5(3):131–133

Talebizadeh M, Morid S, Ayyoubzadeh SA, Ghasemzadeh M (2010) Uncertainty analysis in sediment load modeling using ANN and SWAT model. Water Resour Manag 24(9):1747–1761

Tian Y, Booij M, Xu Y-P (2014) Uncertainty in high and low flows due to model structure and parameter errors. Stoch Env Res Risk Assess 28(2):319–332. doi:10.1007/s00477-013-0751-9

Tiwari MK, Chatterjee C (2010) Uncertainty assessment and ensemble flood forecasting using bootstrap based artificial neural networks (BANNs). J Hydrol 382(1):20–33

Tiwari MK, Chatterjee C (2011) A new wavelet–bootstrap–ANN hybrid model for daily discharge forecasting. J Hydroinform 13(3):500–519

Tongal H, Berndtsson R (2016) Impact of complexity on daily and multi-step forecasting of streamflow with chaotic, stochastic, and black-box models. Stoch Env Res Risk Assess. doi:10.1007/s00477-016-1236-4

Tongal H, Demirel MC, Booij MJ (2013) Seasonality of low flows and dominant processes in the Rhine river. Stoch Env Res Risk Assess 27(2):489–503. doi:10.1007/s00477-012-0594-9

Toth E, Brath A (2007) Multistep ahead streamflow forecasting: role of calibration data in conceptual and neural network modeling. Water Resour Res 43(11):W11405. doi:10.1029/2006WR005383

Uehlinger U, Arndt H, Wantzen KM, Leuven RSEW (2009) The Rhine river basin. Rivers of Europe, chap 6. Academic Press, London, pp 199–245

Uniyal B, Jha MK, Verma AK (2015) Parameter identification and uncertainty analysis for simulating streamflow in a river basin of Eastern India. Hydrol Process 29(17):3744–3766

Walker WE, Harremoës P, Rotmans J, van der Sluijs JP, van Asselt MBA, Janssen P, Krayer von Krauss MP (2003) Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. Integr Assess 4(1):5–17. doi:10.1076/iaij.4.1.5.16466

Wang W-C, Chau K-W, Cheng C-T, Qiu L (2009a) A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. J Hydrol 374(3–4):294–306. doi:10.1016/j.jhydrol.2009.06.019

Wang W, Jin J, Li Y (2009b) Prediction of inflow at three gorges dam in yangtze river with wavelet network model. Water Resour Manag 23(13):2791–2803. doi:10.1007/s11269-009-9409-2

Wang Y, Zheng T, Zhao Y, Jiang J, Wang Y, Guo L, Wang P (2013) Monthly water quality forecasting and uncertainty assessment via bootstrapped wavelet neural networks under missing data for Harbin, China. Environ Sci Pollut Res 20(12):8909–8923. doi:10.1007/s11356-013-1874-8

Warmink JJ, Booij MJ (2015) Uncertainty analysis in river modelling. In: Rowiński P, Radecki-Pawlik A (eds) Rivers—physical, fluvial and environmental processes. Springer, Cham, pp 255–277

Wilcoxon F (1945) Individual comparisons by ranking methods. Biom Bull 1(6):80–83

Xiong L, Wan M, Wei X, O'connor KM (2009) Indices for assessing the prediction bounds of hydrological models and application by generalised likelihood uncertainty estimation/Indices pour évaluer les bornes de prévision de modèles hydrologiques et mise en œuvre pour une estimation d'incertitude par vraisemblance généralisée. Hydrol Sci J 54(5):852–871

Yu J, Qin X, Larsen O (2015) Uncertainty analysis of flood inundation modelling using GLUE with surrogate models in stochastic sampling. Hydrol Process 29(6):1267–1279

Zeroual A, Meddi M, Assani AA (2016) Artificial neural network rainfall-discharge model assessment under rating curve uncertainty and monthly discharge volume predictions. Water Resour Manag. doi:10.1007/s11269-016-1340-8

Zhang W, Li T (2015) The influence of objective function and acceptability threshold on uncertainty assessment of an urban drainage hydraulic model with generalized likelihood uncertainty estimation methodology. Water Resour Manag 29(6):2059–2072

Zhang X, Liang F, Srinivasan R, Van Liew M (2009) Estimating uncertainty of streamflow simulation using Bayesian neural networks. Water Resour Res 45(2):1–16

Zhang H, Zhou J, Ye L, Zeng X, Chen Y (2015) Lower upper bound estimation method considering symmetry for construction of prediction intervals in flood forecasting. Water Resour Manag 29(15):5505–5519

Zhang J, Lin X, Guo B (2016) Multivariate copula-based joint probability distribution of water supply and demand in irrigation district. Water Resour Manag 30(7):2361–2375. doi:10.1007/s11269-016-1293-y