CrossMark

Lower Upper Bound Estimation Method Considering Symmetry for Construction of Prediction Intervals in Flood Forecasting

Hairong Zhang[1,2] · Jianzhong Zhou[1,2] · Lei Ye[1,2] ·
Xiaofan Zeng[1,2] · Yufan Chen[3]

Abstract It is widely accepted that Prediction Interval (PI) can provide more accurate and precise information than deterministic forecast when the uncertainty level increases in flood forecasting. Coverage probability and PI width are two main criteria used to assess the constructed PI, rarely has there been an index to quantify the symmetry between target value and PI. This study extends a newly proposed PI estimation method called Lower Upper Bound Estimation (LUBE) method, which adopts an Artificial Neural Network (ANN) with two outputs to directly generate the upper and lower bounds of PI without making any assumption about the data distribution. A new Prediction Interval Symmetry (PIS) index is introduced and a new objective function is developed for the comprehensive evaluation of PI considering their coverage probability, width and symmetry. Furthermore, Shuffled Complex Evolution algorithm (SCE-UA) is used to minimize the objective function and optimize ANN parameters in the LUBE method. The proposed method is applied to a real world flood forecasting case study of the upper Yangtze River Watershed. The result shows that the SCE-UA based LUBE method with new objective function is very efficient, meanwhile, the midpoint forecasting of the PI obtains excellent performance by evidently improving the symmetry of PI.

Keywords Prediction interval · Symmetry · Artificial neural networks · Uncertainty · Flood forecasting · Shuffled complex evolution

✉ Jianzhong Zhou
jz.zhou@hust.edu.cn

Hairong Zhang
zhr@hust.edu.cn

[1] School of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

[2] Hubei Key Laboratory of Digital Valley Science and Technology, Wuhan 430074, China

[3] China Yangtze Power Company Limited, Yichang 443002, China

🙲 Springer

1 Introduction

Prediction interval (PI) has received much attention in recent years due to its characteristics of quantifying and representing uncertainty, which offers important benefits in hydrologic forecasting. A PI consists of upper and lower bounds between which a future unknown value (e.g. an observed flow value) is expected to lie within a prescribed probability $((1-\alpha)\%)$ called the confidence level. While deterministic forecast provides only one predicted value for the target variable, PI not only provides a range that observed flow is most likely to be covered, it also has an indication of its accuracy called the coverage probability, which consequently helps to enhance the reliability and credibility of the hydrologic outputs.

Generally, the construction of PI is conducted through four methods in previous literatures: delta, Bayesian, generalized likelihood uncertainty estimation (GLUE) and bootstrap. Unfortunately, they are far from optimal. The delta technique assumes that the data error in the modeling system is homogeneous and normally distributed. As heterogeneous as these data error noises could be in many real world applications, the constructed PIs can be questioned (Chryssolouris et al. 1996; Ding and He 2003). The Bayesian technique requires prior knowledge of different uncertainty sources, and needs to be calibrated using Markov Chain Monte Carlo (MCMC) method with millions of runs of model (Sha et al. 2014; Su et al. 2013; Ticlavilca and McKee 2011; Ye et al. 2014; Zhang and Zhao 2012). Regardless of the strength of the supporting Bayesian theory, the methodology has not been widely adopted due to the massive computational burden and the calculation of the Hessian matrix of the objective function. The GLUE method is based upon making a large number of runs of a given model with different parameter values, chosen randomly from specified parameter distributions based on modeler's knowledge of the system (Chen et al. 2013b; Vazquez et al. 2009; Zhang and Li 2015). The GLUE method requires an intimate knowledge of the data and an appropriate selection of threshold value, it is also far from optimal for it can be time-consuming sometimes. The bootstrap method generates different realizations of a dataset to create bootstrap samples and their estimates can provide average and variability of the estimates (Kumar et al. 2015; Sehgal et al. 2014; Sreekanth and Datta 2014; Thomas and Famiglietti 2015; Tiwari and Chatterjee 2010). As this method is conducted on the condition of ensemble models, it may decrease the biased estimate of the true regression of the targets. On the whole, the low feasibility, special assumption about the data distribution, and computational burden are the factors that put a limitation to the popularity of these methods.

In order to overcome these problems, Khosravi et al. (2011) proposed a new method to construct PI called Lower Upper Bound Estimation (LUBE) method for the first time. The LUBE method develops an Artificial Neural Network (ANN) model with two outputs to directly construct the upper and lower bounds of PI. Meanwhile, a coverage width-based criterion (CWC), which considers both coverage probability and PI width, is defined to quantify the constructed PIs' properties. As the LUBE method directly estimates upper and lower bounds, therefore, it needs no data/error distribution or severe computation burden which could greatly save computation time than traditional techniques (see Quan et al. 2014a; Khosravi et al. 2011 for detailed computation time comparisons between these methods).

However, despite the fact that LUBE method is superior in constructing PI, it is still not perfect. The coverage probability and PI width were the two main criteria taken into account when constructing PI using LUBE method in previous investigations, little attention has been paid to the symmetry of the constructed PI. For instance, if the observed values are simply the upper or the lower values of PI, the present criteria used in LUBE method may get a good

result, but this is not the instructive PI we want to pursue practically. Desirable PI should possess high degree of symmetry with respect to the observed values, which wouldn't be easily obtained due to the nonlinearity and complexity of the hydrological processes. Therefore, it is of great value to take the symmetry degree of the PI with respect to observed values into consideration. This paper presents a new objective function considering coverage probability, width and symmetry of the PI in LUBE method, and then use the Shuffled Complex Evolution algorithm (SCE-UA) to calibrate ANN model. In addition, on the basis of the high symmetric PI, this paper gets excellent deterministic forecasting by evaluating the mean value of the upper and the lower bounds of PI.

The rest of this paper is organized as follows: Section 2 gives a brief description of the theoretical background. The methodology to construct PI based on the LUBE method using SCE-UA algorithm is introduced in section 3. Section 4 describes the study area and data preparation. In section 5, a real world case study with results and discussions is revealed. Section 6 summaries the conclusions of this study.

2 Theoretical Background

2.1 Prediction Interval Evaluation Indices

Coverage probability and PI width are the most frequently used PI indices in previous investigations. In this paper, a PI symmetry index is also introduced into the construction of PI.

2.1.1 Prediction Interval Coverage Probability (PICP)

Prediction interval coverage probability (PICP) is the fundamental criterion of PI, which represents the percentage of target values covered by PI (Khosravi et al. 2010). PICP is defined as follows:

$$\mathrm{PICP} = \left(\frac{1}{n}\sum_{i=1}^{n} c_i\right)*100\% \tag{1}$$

$$c_i = \begin{cases} 1 & L_i \leq y_i \leq U_i \\ 0 & otherwise \end{cases} \tag{2}$$

where n is the total number of samples; For sample i, U_i and L_i are the upper and lower bound estimations, whose observed value is y_i. c_i equals 1 if the observed value of the target y_i falls in the range between U_i and L_i, otherwise c_i=0. The ideal case for PICP is when it equals 100 %, which means the observed values of target lie within the prediction band totally.

2.1.2 Prediction Interval Average Relative Width (PIARW)

The ideal case that PICP equals 100 % can be easily obtained if the PI width is broad enough, which can guarantee the bounds cover all the targets. However, these too wide PIs conveying little information about the targets are of no use in practice. Therefore, PI width should be

considered into the evaluation criteria. In the original LUBE method, Prediction Interval Normalized Average Width (PINAW) was used to evaluate the width of PI by Khosravi et al. (2011). Then Quan et al. (2014b) further investigated the Prediction Interval Normalized Root-mean-square Width (PINRW) to evaluate the PI width because the PINAW gave equal weights to all forecasting errors, which may magnify error terms so as to get a better performance. We proposed the Prediction Interval Average Relative Width (PIARW) in which the target variable at all the time steps was effectively utilized. They are defined as follows:

$$PINAW = \frac{1}{nR} \sum_{i=1}^{n} (U_i - L_i) * 100\% \tag{3}$$

$$PINRW = \frac{1}{R} \sqrt{\frac{1}{n} \sum_{i=1}^{n} (U_i - L_i)^2} * 100\% \tag{4}$$

$$PIARW = \frac{1}{n} \sum_{i=1}^{n} \frac{U_i - L_i}{y_i} * 100\% \tag{5}$$

where U_i, L_i, y_i and n are the same as in PICP, R is the range of target variable. They are all dimensionless measures representing the average width of PI. In PINAW and PINRW, only the overall range of target variable is used, the effective information incorporated into the evaluation indices is limited, therefore, PIARW is adapted in this paper.

2.1.3 Prediction Interval Symmetry (PIS)

The criteria discussed above have been extensively used, separately or together to assess the PI (Montanari 2005; Olsson and Lindstrom 2008). However, little attention has been paid to the geometric structure of the band formed by the lower and upper prediction bound trajectories. If the target values lie just on the bounds of PI, the criteria discussed above could get superb results, but the constructed PI makes little contribution to decision making. Therefore Prediction Interval Symmetry (PIS) is introduced herein to quantify the symmetry between the target value and PI (Xiong et al. 2009).

$$PIS = \frac{1}{n} \sum_{i=1}^{n} \frac{\left| y_i - (U_i + L_i) \big/ 2 \right|}{U_i - L_i} * 100\% \tag{6}$$

where U_i, L_i, y_i and n are the same as in PICP. PIS is also a dimensionless measurement demonstrating the symmetry degree of PI, which is defined as the ratio of the difference between the target value and the mean value of upper and lower value of PI to the actual band-

width. As to the location of the target value relative to PI, there are three different scenarios defined below.

In case (a), as shown in Fig. 1a, we can see the target value y_i lies between the upper bound U_i and the lower bound L_i, which confines the proposed PIS to be values between 0 and 0.5. If the PI is completely symmetrical around y_i, which is the ideal case we pursue, PIS will be 0 as $y_i = (U_i + L_i)/2$. Hence, the closer the mean values of PI $((U_i + L_i)/2)$ are to their target values y_i, the closer to 0 the PIS values will be, indicating an almost perfect symmetry of PI about the discharge hydrograph.
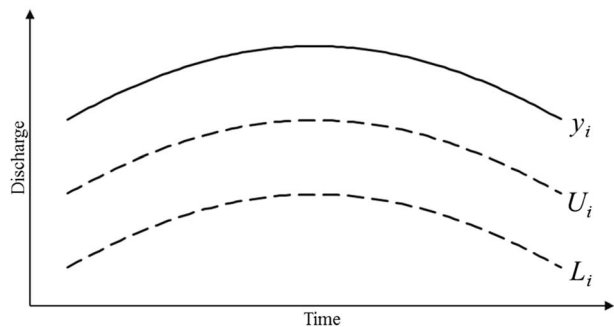

Fig. 1 Three possible scenarios for the location of the observed discharges with respect to the PI
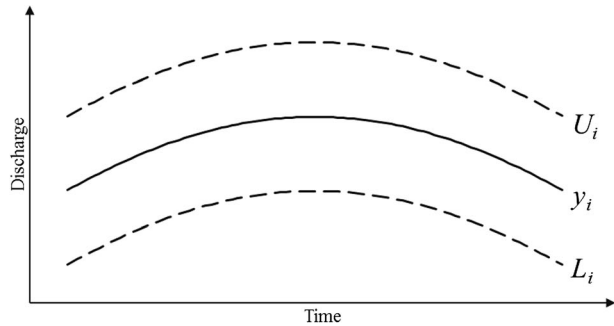
(a) $L_i \leq y_i \leq U_i$, $0 \leq \mathrm{PIS} \leq 0.5$
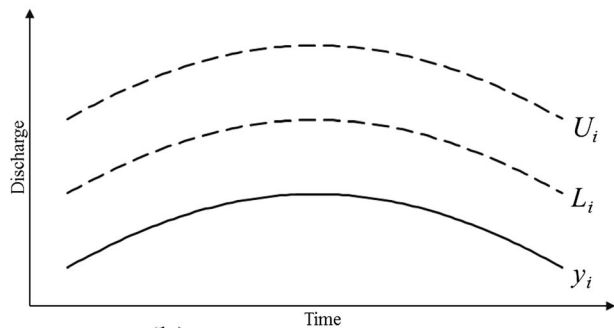
(b) $y_i < L_i \leq U_i$, $\mathrm{PIS} > 0.5$

(c) $L_i \leq U_i < y_i$, $\mathrm{PIS} > 0.5$

In case (b), as shown in Fig. 1b, where the target value y_i lies below the lower bound L_i, thus resulting in a PIS greater than 0.5. Meanwhile, the more the PIS values exceed 0.5 reflects greater asymmetry of the PI about the discharge hydrograph.

In case (c), as shown in Fig. 1c, where the target value y_i lies above the upper bound U_i, each PIS will have a value larger than 0.5, and as in case (b), a larger PIS reflects a greater asymmetry of the PI about the target value.

In conclusion, an average symmetry index PIS<0.5 would indicate that the observed values lie within PI on average. The larger the value of PIS is over 0.5, the more asymmetrical the PI is around the observed values, correspondingly, a zero PIS means a completely symmetrical scenario.

2.2 Brief Description of SCE-UA

The SCE-UA algorithm, developed by Duan et al. (1994) with parameter optimization of conceptual rainfall-runoff models, is based on the synthesis of four concepts: (1) combination of deterministic and probabilistic approaches; (2) systematic evolution of a 'complex' of points spanning the parameter space, in the direction of global improvement; (3) competitive evolution; (4) complex shuffling (Duan et al. 1994). The algorithm is designed specifically to deal with the peculiar problems encountered in conceptual watershed model calibration, and it has the advantages of handling high parameter dimensionality and not relying on the availability of an explicit expression for the objective function or the derivatives. It has been widely applied to various field of optimization due to these capacities (Guo et al. 2013; He et al. 2007; Kang and Lee 2014). The SCE-UA algorithm proceeds as follows:

(1) Initialize parameters: the parameters of SCE-UA algorithm mainly include the number of complexes q; the number of points in each complex m; the population size $s = qm$; the number of evolution steps allowed for each complex before complex shuffling ss; the maximum number of function evaluations max; the number of shuffling loops in which the criterion value must change by pc before optimization is terminated ks; the percentage by which the criterion value must change in ks shuffling loops pc; and the minimum normalized geometric range of parameter ranges $peps$.

(2) Generate sample: generate the population with size s initial points in the feasible space and compute the criterion value at each point.

(3) Rank points: Sort the s points by their criterion value in ascending order and store them in a set $P=\{P_1,P_2,\cdots,P_s\}$, so that the first point represents the point with the smallest value.

(4) Partition into complexes: Partition the set P into q complexes $C^1,C^2,\cdots,C^q$, each complex contains m points, such that the first complex contains every $q(j-1) + 1$ ranked point, the second complex contains every $q(j-1) + 2$ ranked point of P, and so forth, where $j=1,2,\ldots,m$.

(5) Evolve each complex: Evolve each complex independently by taking ss evolution steps according to the competitive complex evolution (CCE). In order to save the length of this paper, CCE is not introduced here. Readers may refer to Duan et al. (1994) for a detailed description of CCE.

(6) Shuffle complexes: Shuffle the q complexes into a single sample population and sort the sample population in ascending order by their criterion value.

(7) Check convergence: Repeat step 5 to 7 until termination criteria are met. The termination criteria include that the maximum number of function evaluations max is reached; no

obvious improvement greater than *pc* is achieved for *ks* shuffling loops; a normalized geometric range of parameter lower than *peps* is reached.

3 Methodology

3.1 Artificial Neural Network

Artificial Neural Network (ANN), which was first introduced in 1943 (McCulloch and Pitts 1943), has been successfully applied to different areas (Hippert et al. 2001; Latt et al. 2015; Mohanty et al. 2010; Parmar and Bhardwaj 2015; Parsaie and Haghiabi 2015) for its excellent ability of approximation and learning. The LUBE method takes the two outputs of ANN model to directly construct the upper and lower bounds of PI. Figure 2 shows a symbolic ANN with two outputs for the LUBE method. A more flexible ANN can be modeled with different inputs and hidden layers. For a typical three-layered ANN, the mathematical mapping between the inputs and the outputs is shown in Formula (7) (Chen et al. 2008).

$$y_i = f_1\left(\sum_{j=1}^{N_h}\left(w_{ij}f_2\left(\sum_{k=1}^{N_i}v_{jk}x_k + b_{vj}\right) + b_{wi}\right)\right) \tag{7}$$

where y_i is the output of the *i*th node on the output layer; x_k is the input of the *k*th node in the input layer; w_{ij} is the connection weight between nodes in the hidden and output layers; v_{jk} is the connection weight between nodes in the input and hidden layers; and b_{wi} and b_{vj} are bias terms that represent the threshold of transfer function f_1 and f_2 respectively. The numbers of nodes in the input, hidden and output layers are N_i, N_h and N_o. In the LUBE method, N_o equals 2, and the bigger one is used as the upper bound and the smaller one is used as the lower bound at each time step.
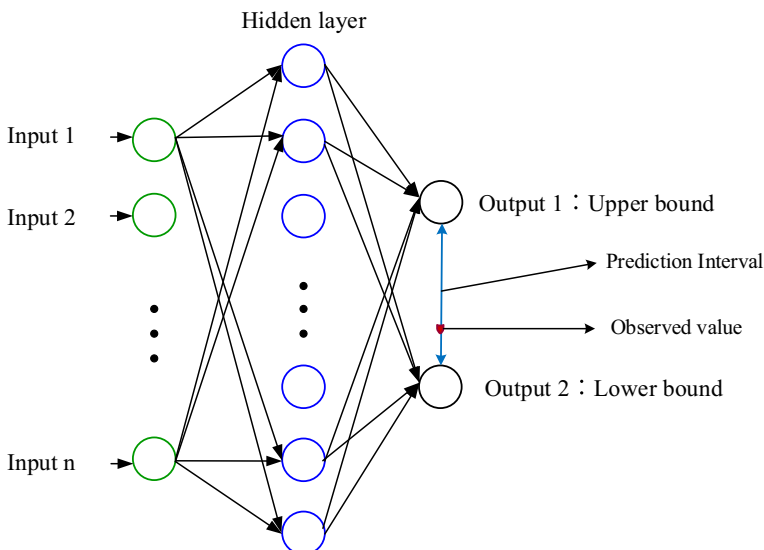
Fig. 2 ANN model structure for constructing upper and lower bounds of PI in the LUBE method

3.2 Coverage Width Symmetry-Based Criterion (CWSC)

PICP and PIARW are two conflicting criterion when they are applied to construct PI. A higher PICP will more likely result in a wider PIARW, and a narrower PIARW will always lead to a lower PICP. To construct a PI with higher PICP and narrower PIARW, an appropriate function is needed to qualify PI thoroughly, which is why Coverage Width-based Criterion (CWC) is brought out. Khosravi et al. (2011) defined a CWC function first, and then Quan et al. (2014b) improved it by replacing PINAW with PINRW. They are defined as follows:

$$\text{CWC}_{Original} = \text{PINAW}\left(1 + \gamma(\text{PICP})e^{-\eta(\text{PICP}-\mu)}\right) \tag{8}$$

$$\text{CWC}_{Quan} = \text{PINRW}\left(1 + \gamma(\text{PICP})e^{-\eta(\text{PICP}-\mu)}\right) \tag{9}$$

$\text{CWC}_{Original}$ and CWC_{Quan} are the CWCs used by Khosravi et al. and Quan et al., respectively. μ and η are two constant hyper-parameters that determine how much the penalty is in case of unsatisfying the required condition. μ is the nominal confidence level associated with PI and can be set to $((1-\alpha)\%)$. η exponentially magnifies the difference between the PICP and μ. During the calibration period, $\gamma(\text{PICP})$ is constantly equal to 1. While in the evaluation period, $\gamma(\text{PICP})$ is a step function which entirely depends on the acquired PICP

$$\gamma(\text{PICP}) = \begin{cases} 0 & \text{PICP} \geq \mu \\ 1 & \text{PICP} < \mu \end{cases} \tag{10}$$

Although these CWCs are well behaved in previous studies, they still have their shortcomings. Considering that the two CWCs mentioned above only measure the coverage probability and PI width, some improvement can be made by taking PI symmetry into account, so we build a new Coverage Width Symmetry-based Criterion (CWSC).

$$\text{CWSC}_{Proposed} = \gamma(\text{PIS})e^{\eta_3(\text{PIS}-\mu_2)} + \eta_2\text{PIARW} + \gamma(\text{PICP})e^{-\eta_1(\text{PICP}-\mu_1)} \tag{11}$$

where PIARW and PIS are used here to quantify width and symmetry of PI; the three hyper-parameters η_1, η_2, μ_1 and the $\gamma(\text{PICP})$ are the same as in previous CWCs. Another hyper-parameters η_3, μ_2 and the function $\gamma(\text{PIS})$ are added herein to control the PIS. During the calibration period, $\gamma(\text{PIS})$ is equal to constant 1. While in the evaluation period, $\gamma(\text{PIS})$ is a step function just as $\gamma(\text{PICP})$.

Unlike PICP is better to be higher than μ_1, a PIS lower than μ_2 is preferred because of its definition. As PICP is theoretically the fundamental feature of the PIs, the η_1 should be selected to have a large value so as to put more weight on the variation of PICP. The rationale of using exponents for PICP and PIS is that a PICP lower than μ_1 or a PIS higher than μ_2 will give a misleading optimistic PI, which should be penalized more. A PICP higher than μ_1 and a PIS lower than μ_2 will be rewarded, which should be penalized less so the PIARW could play a leading role in later calibration. The reason why the proposed CWSC should be the sum rather than the product of PICP, PIARW and PIS is that the sum of them is easier to control. Meanwhile, the CWSC function in additive form can also void the undesirable situation where a minimum CWSC value of zero can occur by finding a zero PI width.

At the beginning of calibration where PICP is smaller than μ_1, a large penalization is paid to PICP. While the PICP increases to its nominal confidence level gradually, the penalization for PICP will become weaker, and the PIS would decrease to its own nominal confidence level at the same time. Later on these three criteria would transfer into a conflicting status, and result in a comprehensive PI considering all indices eventually.

3.3 LUBE Method

The basic concept of LUBE method is to develop a simple ANN with two outputs to generate the upper and lower bounds of PIs. Traditional methods for constructing PI mainly have two steps: firstly, conduct deterministic forecasts; secondly, evaluate the mean and variance values of deterministic forecasts for PI construction. By contrast, the LUBE method directly construct upper and lower bounds of PI in one step.

The flow chart of the LUBE method based on SCE-UA is shown in Fig. 3. As an evolutionary-based procedure, it simultaneously evolves a group of solutions (parameter sets) towards better solutions in the search space, trying to cover the global optimum of the objective function (Blasone et al. 2007). A general description of the optimization procedure is outlined as follows.

(1) Split the whole data into calibration and validation.
(2) Determine the input variables and architecture of the ANN model.
(3) Define feasible ranges for the ANN model parameters and sample s points randomly in the feasible parameter space.
(4) Construct PI for each point and the performance of each point is evaluated in terms of $CWSC_{Proposed}$ value.
(5) Conduct the optimization search using the SCE-UA algorithm to generate an optimal set of solutions in the feasible parameter space.
(6) Construct PI for the evaluation period using the optimal parameters obtained from above.

4 Study Area and Data Preparation

We take the Yangtze River (Chang Jiang) as the study area of flood forecasting in this paper. The Yangtze River, which originates from the Qinghai-Tibet Plateau with an elevation of 6600 m and flows about 6300 km eastward into the East China Sea, is the longest river in China, and it plays a vital role in the economic development and ecological environmental conservation of China. This watershed is located between 91°E–122°E and 25°N–35°N, and covers a total area of about 1,808,500 km^2, which is nearly 18 % of China's area. With the completion of the Three Gorges Hydropower Dam, the Yangtze River hydropower resources is effectively developed and utilized. Thus the Three Gorges Dam takes the responsibility of flood control, agricultural hydroelectric power generation, and municipal and industrial water supply. Discharge forecasting is significant to the optimal operation of the Three Gorges Dam.

This study mainly focuses on the upper Yangtze River above Yichang, where the Three Gorges Dam is located. There are a complex of tributaries joining the upper Yangtze River, such as Jinsha, Min, Tuo, Jialing and Wu Rivers. A schematic of the regional major tributary rivers and gauging stations is given in Fig. 4. As Yalong River joins Jinsha River,
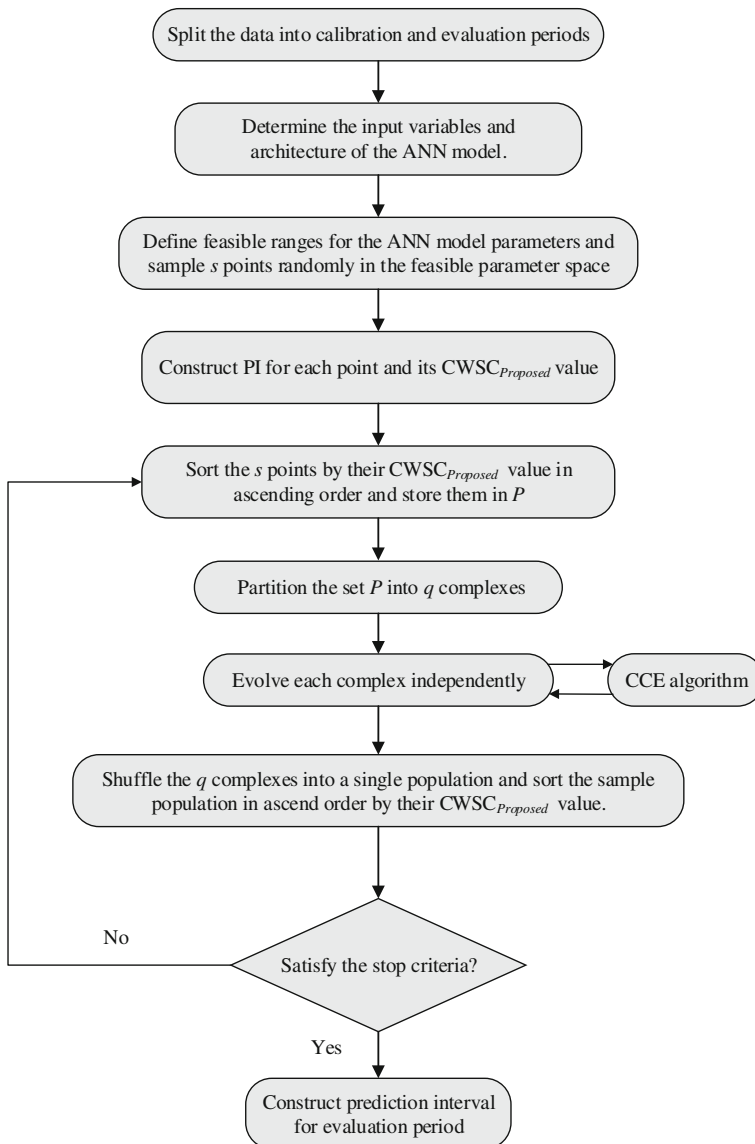
Fig. 3 The flow chart describing the SCE-UA based LUBE method for generating PI

it can be viewed as a tributary of Jinsha River as well. Six hydrological controlling stations of the five tributaries and mainstream are taken into consideration. They are Pingshan, Gaochang, Lijiawan, Beibei, Wulong, Yichang station from upstream to downstream.

The daily discharge data measured between 1953 and 2007 of the six hydrological controlling stations during flood season (June – September) is adopted in this research, representing a wide variety of hydrological conditions. During model calibration period, the first 35 years (1953 to 1987) were taken into account. During model evaluation period, the rest 10 years (1988 to 2007) were used. We naturalized the discharge at the
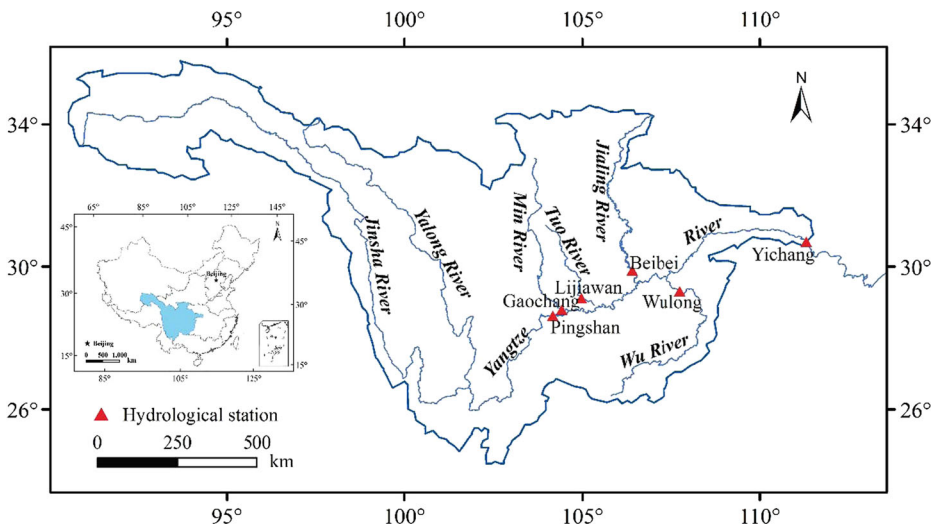
Fig. 4 Locations of regional tributary rivers and control hydrological stations for the upper Yangtze River

beginning to remove the storage effects of the upper hydropower stations. Further, Zhang et al. (2007) showed that the discharge over the 45 years only has a very slight downward tendency which is not remarkable. Therefore, the consistency of the calibration and validation is guaranteed.

5 Case Study

5.1 Architecture of ANN Model

According to the copula entropy input determination results in Chen et al. (2013a), the best input set for the ANN model to forecast $Q_{yc,t}$ consisted of $Q_{yc,t-1}$, $Q_{yc,t-2}$, $Q_{ps,t-1}$, $Q_{gc,t-3}$, $Q_{bb,t-2}$, $Q_{wl,t-2}$, $Q_{ljw,t-3}$, in which subscript yc, ps, gc, bb, wl, ljw means Yichang, Pingshan, Gaochang, Beibei, Wulong, Lijiawan hydrological station, respectively, t represents the time step. In LUBE method, the outputs of the ANN model are the upper and lower bounds of $Q_{yc,t}$. Hence the number of input and output neurons is seven and two, respectively. The optimal number of hidden neurons should be 3 after trial and error, thus the number of the ANN model parameters needs to be optimized is 32.

5.2 Calibration Process

The parameter settings of SCE-UA were based on the recommendations in Duan et al. (1994), the value of p, m, s, ss, max, pc, ks and $peps$ in SCE-UA was set to 4, 65, 260, 65, 100,000, 0.0001, 30 and 0.001 respectively. For the CWSC$_{Proposed}$, the prescribed level of confidence of PI was 90 %, the value of μ_1, μ_2, η_1, η_2 and η_3 was set to 0.9, 0.5, 80, 70 and 50 respectively to balance the trade-off between PICP, PIARW and PIS.

The calibration process was terminated because of the function convergence termination criterion pc was met. The calibration process of CWSC$_{Proposed}$ is illustrated in Fig. 5, showing
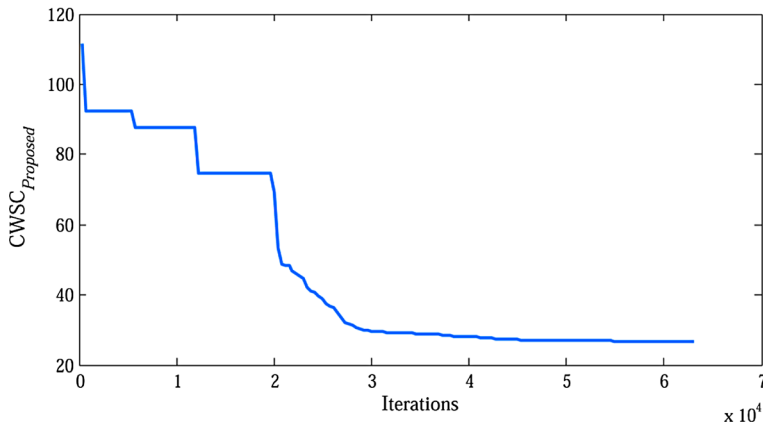
Fig. 5 Calibration process to minimize CWSC$_{Proposed}$ objective function

that CWSC$_{Proposed}$ value of the best point converged to a good result. The CWSC$_{Proposed}$ decreased rapidly at the very beginning of the calibration process, and then the CWSC$_{Proposed}$ stayed stable for a while because the optimization fell into a local optimum. After this, the CWSC$_{Proposed}$ went on decreasing and gradually converged its optimal value. The calibration process clearly demonstrates that the strong searching capacity of the SCE algorithm.

5.3 Results and Discussions

As a benchmark for comparison, we conduct calculation using CWSC$_{Proposed}$ as well as the previous two objective functions. The parameter of SCE-UA and ANN for all the methods are set to be the same for the consideration of the veracity in every objective function's calculation. Meanwhile, it's also significant to ensure the coverage probability, width and symmetry of PI reaches an optimum balance, so several experiments were conducted and we managed to get the optimal parameter values for each objective function: η is set to 35 and 80 in CWC$_{Ori}$ and CWC$_{Quan}$ respectively.

The performances of these three objective functions during the calibration and validation periods are shown in Table 1. We calculated the deterministic forecast for the 3 PIs generated by different objective function, which is the midpoint of the upper and lower

Table 1 Estimation indices for PIs generated using the four different objective functions during calibration and validation period

	Types of PI	PICP(%)	PIRAW(%)	PIS(%)	RMSE
Calibration	PI$_{Original}$	**94.7**	44.5	20.2	3158.7
	PI$_{Quan}$	92.9	39.4	24.1	2838.1
	PI$_{Proposed}$	94.6	**34.2**	**17.0**	**2215.0**
Validation	PI$_{Original}$	92.7	45.2	20.7	3046.5
	PI$_{Quan}$	91.5	40.5	24.1	2812.6
	PI$_{Proposed}$	**93.8**	**34.5**	**17.4**	**2193.5**

The optimal indicators for these indices are marked in bold

bounds of PI. Root Mean Square Errors (RMSE) of the deterministic forecast are listed in the tables as well, with the optimal indicators marked in bold. As demonstrated in Table 1, PICP for every PI is greater than 90 %, which indicates that the objective functions used above are all suitable and LUBE is an effective method to construct PI.

By observing the evaluation indices of PI during calibration period in Table 1, it's obvious that the highest PICP was generated from PI_{Ori}. However, its corresponding width index PIRAW reaches 44.5 %, which makes the PI_{Ori} a less informative way to use practically. $PI_{Proposed}$ has a relatively balanced control on the coverage probability and PI width for it has the best width index while keeping a good performance in PICP. As for the PIS index, the $CWSC_{Proposed}$ we brought up in this paper is the optimal choice since it's the only objective function that takes symmetry into account. Its according RMSE generated from deterministic forecast of PI is also the optimal one in the three functions, which implies that the improvement in symmetry would evidently improve the accuracy of midpoint forecasting and it's of great significance for practical usage. The performance of the three PIs during validation period is consistent to that during calibration period. It can be figured from Table 1 that the $CWSC_{Proposed}$ we brought up has a perfect balance between PICP, PIRAW and PIS, and the according RMSE of deterministic forecast is optimal as well.

Figure 6 displays the PI, observed values and deterministic forecast of the flow diagram generated by the $CWSC_{Proposed}$. Because the evaluation period is quite long, we plot the flow diagram for four representative years 1989, 1998, 2000 and 2004, in which the entire Yangtze River basin suffered from severe flooding. Intuitively shown in the figures, the PI generated by taking $CWSC_{Proposed}$ as objective function in LUBE method is excellent: they demonstrates high coverage probability (expect for some flood peak flow), proper control on width and keeps the symmetry in the meantime. The generated midpoint forecasting have a great matching with the observed values, which is superior to the outcome of deterministic forecast generated using ANN with only one output by Chen et al. (2013a).
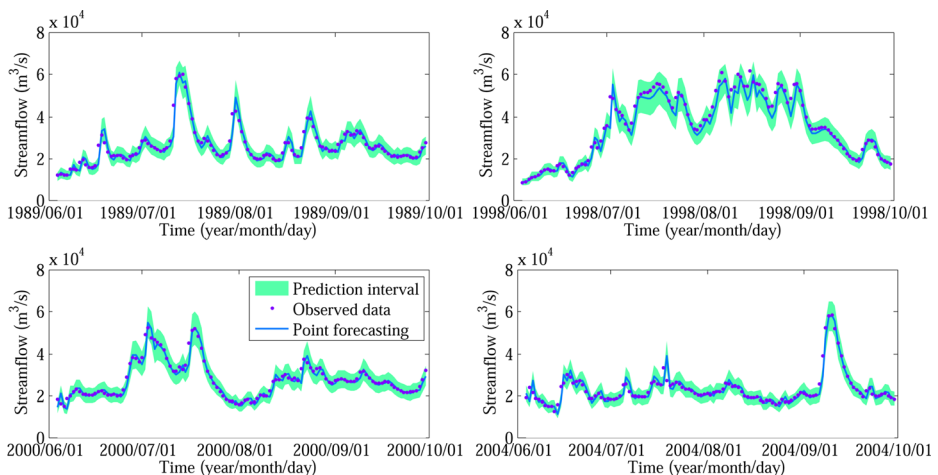
Fig. 6 Prediction interval generated by the improved LUBE along with observed flows

6 Conclusions

Improvements are made to the recently developed LUBE method in this paper, and the improved method is applied to construct prediction intervals in flood forecasting. We firstly add symmetry as one of the estimation criteria in the LUBE method. A new objective function that concerns coverage probability, width and symmetry of PI simultaneously is proposed when conducting parameter optimization on ANN model. The minimization of $CWSC_{Proposed}$ function value is realized through SCE-UA algorithm, so as to reach the purpose of increase coverage probability, decrease width, and improve symmetry, thus constructing the optimal PI. In addition, we also get excellent deterministic forecasting by using the average of the lower and upper bounds of PI.

By applying the proposed CWSC function as well as the former two CWC functions to the flood forecasting of upper Yangtze River, the comparison results show that the proposed CWSC function achieved good performance in PICP, PIRAW and PIS simultaneously, and showed great performance in the midpoint forecasting by introducing symmetry, which is even superior to the outcome of deterministic forecasting generated using one output ANN.

To the best of our knowledge, this is the first application of the LUBE method in hydrologic field. The LUBE is a special ANN method which needs no data/error distribution or severe computation burden, providing the possibility of applying it to other practical examples such as flash flood or seasonal flood as long as it can be modeled using ANN model. Further studies should extend the usage of LUBE method to other fields and test the applicability of the method.

References

Blasone R, Madsen H, Rosbjerg D (2007) Parameter estimation in distributed hydrological modelling: comparison of global and local optimisation techniques. Nord Hydrol 38(4–5):451–476

Chen C, Wu J, Chen J (2008) Prediction of flutter derivatives by artificial neural networks. J Wind Eng Ind Aerodyn 96(10):1925–1937

Chen L, Ye L, Singh V, Zhou J, Guo S (2013a) Determination of input for artificial neural networks for flood forecasting using the copula entropy method. J Hydrol Eng 4014021

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

Chryssolouris G, Lee M, Ramsey A (1996) Confidence interval prediction for neural network models. IEEE T Neural Networ 7(1):229–232

Ding AA, He X (2003) Backpropagation of pseudo-errors: neural networks that are adaptive to heterogeneous noise. IEEE T Neural Networ 14(2):253–262

Duan Q, Sorooshian S, Gupta VK (1994) Optimal use of the SCE-UA global optimization method for calibrating watershed models. J Hydrol 158(3–4):265–284

Guo J, Zhou J, Zou Q, Liu Y, Song L (2013) A novel multi-objective shuffled complex differential evolution algorithm with application to hydrological model parameter optimization. Water Resour Manag 27(8):2923–2946

He B, Takase K, Wang Y (2007) Regional groundwater prediction model using automatic parameter calibration SCE method for a coastal plain of Seto Inland Sea. Water Resour Manag 21(6):947–959

Hippert HS, Pedreira CE, Souza RC (2001) Neural networks for short-term load forecasting: a review and evaluation. IEEE T Power Syst 16(1):44–55

Kang T, Lee S (2014) Modification of the SCE-UA to include constraints by embedding an adaptive penalty function and application: application approach. Water Resour Manag 28(8):2145–2159

Khosravi A, Nahavandi S, Creighton D (2010) Construction of optimal prediction intervals for load forecasting problems. IEEE T Power Syst 25(3):1496–1503

Khosravi A, Nahavandi S, Creighton D, Atiya AF (2011) Lower upper bound estimation method for construction of neural network-based prediction intervals. IEEE T Neural Networ 22(3):337–346

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 1–21

Latt ZZ, Wittenberg H, Urban B (2015) Clustering hydrological homogeneous regions and neural network based index flood estimation for ungauged catchments: an example of the Chindwin River in Myanmar. Water Resour Manag 29(3):913–928

McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5(4):115–133

Mohanty S, Jha MK, Kumar A, Sudheer KP (2010) Artificial neural network modeling for groundwater level forecasting in a River Island of Eastern India. Water Resour Manag 24(9):1845–1865

Montanari A (2005) Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations. Water Resour Res 41(8)

Olsson J, Lindström G (2008) Evaluation and calibration of operational hydrological ensemble forecasts in Sweden. J Hydrol 350(1):14–24

Parmar KS, Bhardwaj R (2015) River water prediction modeling using neural networks, fuzzy and wavelet coupled model. Water Resour Manag 29(1):17–33

Parsaie A, Haghiabi A (2015) The effect of predicting discharge coefficient by neural network on increasing the numerical modeling accuracy of flow over side weir. Water Resour Manag 29(4):973–985

Quan H, Srinivasan D, Khosravi A (2014a) Uncertainty handling using neural network-based prediction intervals for electrical load forecasting. Energy 73:916–925

Quan H, Srinivasan D, Khosravi A (2014b) Particle swarm optimization for construction of neural network-based prediction intervals. Neurocomputing 127:172–180

Sehgal V, Tiwari MK, Chatterjee C (2014) Wavelet bootstrap multiple linear regression based hybrid modeling for daily river discharge forecasting. Water Resour Manag 28(10):2793–2811

Sha J, Li Z, Swaney DP, Hong B, Wang W, Wang Y (2014) Application of a Bayesian watershed model linking multivariate statistical analysis to support watershed-scale nitrogen management in China. Water Resour Manag 28(11):3681–3695

Sreekanth J, Datta B (2014) Stochastic and robust multi-objective optimal management of pumping from coastal aquifers under parameter uncertainty. Water Resour Manag 28(7):2005–2019

Su VS, Nik J, Molina J, Vamvakeridou-Lyroudia LS, DA Savi C, Kapelan Z (2013) Comparative analysis of system dynamics and object-oriented bayesian networks modelling for water systems management. Water Resour Manag 27(3):819–841

Thomas BF, Famiglietti JS (2015) Sustainable groundwater management in the arid Southwestern US: Coachella Valley, California. Water Resour Manag 29(12):4411–4426

Ticlavilca AM, McKee M (2011) Multivariate Bayesian regression approach to forecast releases from a system of multiple reservoirs. Water Resour Manag 25(2):523–543

Tiwari MK, Chatterjee C (2010) Development of an accurate and reliable hourly flood forecasting model using wavelet–bootstrap–ANN (WBANN) hybrid approach. J Hydrol 394(3):458–470

Vazquez RF, Beven K, Feyen J (2009) GLUE based assessment on the overall predictions of a MIKE SHE application. Water Resour Manag 23(7):1325–1349

Xiong L, Wan M, Wei X et al (2009) Indices for assessing the prediction bounds of hydrological models and application by generalised likelihood uncertainty estimation. Hydrol Sci J 54(5):852–871

Ye L, Zhou J, Zeng X, Guo J, Zhang X (2014) Multi-objective optimization for construction of prediction interval of hydrological models based on ensemble simulations. J Hydrol 519:925–933

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, Zhao K (2012) Bayesian neural networks for uncertainty analysis of hydrologic modeling: a comparison of two schemes. Water Resour Manag 26(8):2365–2382

Zhang J et al (2007) Study of runoff of the six large basins in China over the past 50 years. Adv Water Sci 02: 230–234

(**In Chinese**)