

## Research Article

# Calibration of Conceptual Rainfall-Runoff Models Using Global Optimization

Chao Zhang,<sup>1,2</sup> Ru-bin Wang,<sup>1,2</sup> and Qing-xiang Meng<sup>1,2,3</sup>

<sup>1</sup>Key Laboratory of Ministry of Education for Geomechanics and Embankment Engineering, Hohai University, Nanjing 210098, China

<sup>2</sup>Research Institute of Geotechnical Engineering, Hohai University, Nanjing 210098, China

<sup>3</sup>Department of Civil & Environmental Engineering, University of Waterloo, Waterloo, ON, Canada N2L 3G1

Correspondence should be addressed to Chao Zhang; zchohai@foxmail.com

Received 19 November 2014; Revised 9 February 2015; Accepted 19 February 2015

Academic Editor: Hann-Ming H. Juang

Copyright © 2015 Chao Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Parameter optimization for the conceptual rainfall-runoff (CRR) model has always been the difficult problem in hydrology since watershed hydrological model is high-dimensional and nonlinear with multimodal and nonconvex response surface and its parameters are obviously related and complementary. In the research presented here, the shuffled complex evolution (SCE-UA) global optimization method was used to calibrate the Xinanjiang (XAJ) model. We defined the ideal data and applied the method to observed data. Our results show that, in the case of ideal data, the data length did not affect the parameter optimization for the hydrological model. If the objective function was selected appropriately, the proposed method found the true parameter values. In the case of observed data, we applied the technique to different lengths of data (1, 2, and 3 years) and compared the results with ideal data. We found that errors in the data and model structure lead to significant uncertainties in the parameter optimization.

## 1. Introduction

Over the past few decades, a wide range of conceptual rainfall-runoff (CRR) models have been developed. CRR models are often preferable to other types of watershed hydrological models (e.g., physically based models). This is because they have a reasonable accuracy and are simpler to compute in many practical cases, because we may only require runoff process estimates from rainfall at the watershed outlet or a given location.

CRR models that simulate watershed hydrological processes using methods from mathematical physics can be expressed in terms of their model structure and parameters. In other words, in addition to the rationality of the model structure, the parameters directly determine the accuracy of the model and its forecasts. In theory, most CRR model parameters could be determined directly or indirectly through measurements or a physical method, because they have physical meaning. The process of model parameter

conditioning to historical system response data is called calibration.

There are generally two types of optimization procedures, manual selection and automatic optimization. The manual selection process relies on certain subjectivity, so the results can vary from person to person. It requires some experience and understanding of the model structure. Additionally, manual selection can be laborious and time consuming, especially for inexperienced hydrology workers. Automatic optimization methods have become increasingly popular because of rapid developments in computing technology. For example, genetic algorithms, the shuffled complex evolution method (SCE-UA), the multiple start simplex, adaptive random search, particle swarm optimization (PSO), multiobjective shuffled complex evolution metropolis (MOSCEM) algorithm, and dynamically dimensioned search (DDS) have been successfully applied to model calibration [1–11]. However, if a model is calibrated by either of these two procedures, we cannot be certain that we will obtain a unique set of optimal parameters for a CRR model.

Despite nearly 20 years of research, little progress has been made on the parameter estimation problem for CRR models since objective function response surface typically contains hundreds, if not thousands, of local optima, nested at several scales [3]. In the research presented here, we used the SCE-UA method to calibrate the XAJ model. The XAJ model is extensively used throughout the world to analyze the stability of parameters from automatic optimization methods. It considers the hydrological data, parameter search interval, objective functions, and other aspects.

## 2. XAJ Model

*2.1. The Model Structure and Its Main Characteristics.* The XAJ [12, 13] model was developed in 1973 by the East China College of Hydraulic Engineering (now Hohai University). Its underlying aim was to forecast flows to the Xinanjiang reservoir. The model has been successfully and widely applied in humid and semihumid regions. It is based on “runoff formation at the natural storage,” which is the distinguishing feature of the XAJ model when compared to other models. The basin is divided into a set of subbasins using a method such as Thiessen polygon modification, considering the uneven distribution of rainfall and the underlying surface. Then, the discharge curve on the outlet section of each subbasin is simulated, and flood routing is determined. Finally, the total discharge is obtained using a simple sum. The XAJ model is composed of four modules: the evaporation module, the runoff production module, the runoff separation module, and the runoff concentration module.

Figure 1 shows the flow chart of the XAJ model. Rainfall ( $P$ ) and water-surface evaporation ( $EM$ ) are the input data, and the discharge curve for the outlet section ( $Q$ ) and the evaporation of the watershed ( $E$ ) are the output results. The state variables are in boxes and the model parameters are outside the boxes (Figure 1). For this research,  $W$  is the tension water storage;  $WU$  is the upper layer tension water storage;  $WL$  is the lower layer tension water storage;  $FR$  is the runoff contributing area factor;  $S$  is the free water storage;  $RS$  is the surface runoff;  $RI$  is the interflow runoff;  $RG$  is the ground water runoff;  $QS$  is the surface flow;  $QI$  is the interflow; and  $QG$  is the ground water flow.

The XAJ model has several characteristics that can be summarized as follows.

- (1) The rainfall-runoff process is divided into two stages: runoff generation and concentration in the watershed. It is thought that, in the runoff yield stage, runoff is produced only after the deficit of the vadose zone is satisfied. A homogeneous vadose zone is subject to ground water flow and excess surface flow infiltration. Interflow will be produced in a vadose zone with a relatively impermeable layer, in addition to ground water flow and saturated overland flow. In the runoff concentration stage, the river network and subbasin concentration can be considered as two types of watershed concentration. The subbasin combined with a river network can fully embody the watershed concentration. Because the XAJ model is

very compatible in the treatment of the watershed concentration, the subbasin and river network concentration are commonly represented by the Sherman unit hydrograph and Muskingum methods for successive routing by subreaches [14], as well as other methods (such as Clark method, etc.)

- (2) A three-layer evaporation model is used. Here, the “layer” takes soil moisture constants such as the field capacity and wilting point as thresholds. In addition to the soil moisture constants, the soil evaporation ability is an important factor in the three-layer evaporation model and has a great effect on the accuracy. It is generally difficult to directly obtain an accurate value using instrumental observations. Therefore, in the XAJ model, the measured water-surface evaporation is revised by a correction factor and improved by the water balance of the watershed. In this way, we can avoid using empirical formulas to calculate the soil evaporation. This method of dealing with evaporation is the only one used in practical applications.
- (3) The XAJ model considers runoff separation, which means that the calculated flood process is more in line with the actual situation. Because the runoff production components have different flow velocities, the runoff concentration results are more accurate if we calculate the runoff separation using the appropriate velocities. The XAJ model uses a “downward” structure for the runoff separation, whereas other CRR models generally use an “upward” structure.
- (4) The XAJ model uses a statistically significant watershed storage capacity curve and a watershed free water capacity curve. Note that these curves are only applicable to the analysis of runoff area variabilities caused by an uneven distribution of the underlying surface, under the condition of a uniform rainfall spatial distribution. Additionally, for a closed watershed, we should use the upper limits of both curves. For an unclosed watershed, the upper limit is infinite. The infiltration capacity area distribution curve is used by Stanford model to account for the influence of the uneven distribution of the underlying surface on the infiltration of the excess surface runoff. This is because the watershed storage and watershed free water capacity curves are set in the model. However, the XAJ model is more advanced than the other CRR models.

*2.2. Model Parameters.* The XAJ model has 17 parameters that must be determined by the user (XAJ Model Parameters section) when computing the flood using the Muskingum method for successive routing by subreaches. We temporarily set the feasible parameter space by fixing the upper and lower parameter bounds (see Table 1). The Muskingum method parameters were predetermined based on the observed hydrograph and were not included in the optimization.

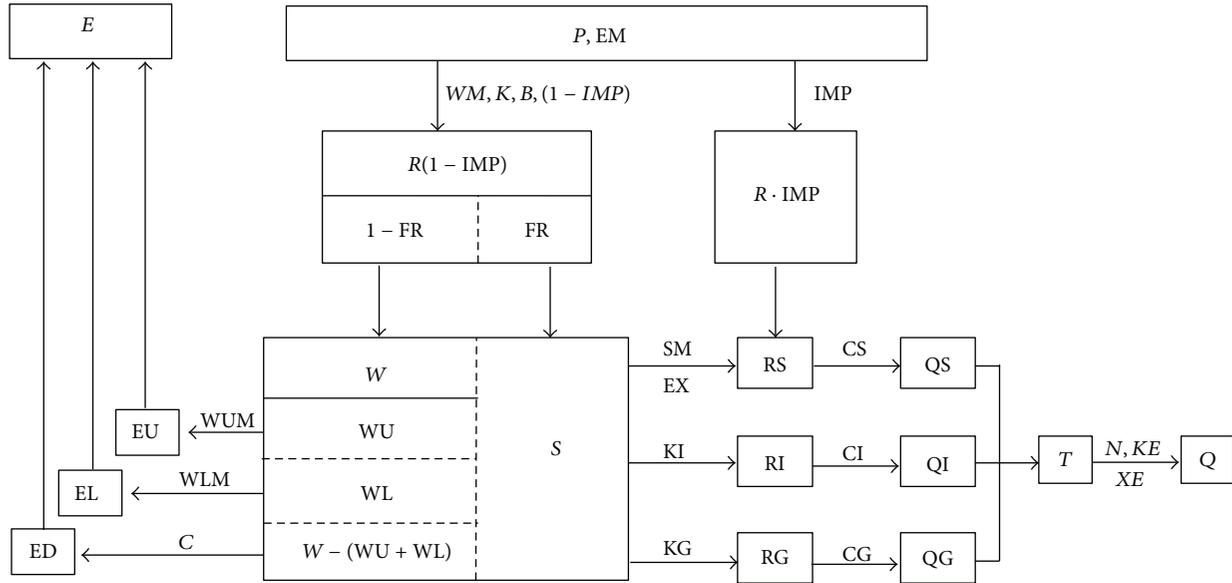


FIGURE 1: Flow chart for the XAJ model.

TABLE 1: Lower and upper bounds on the parameters.

Parameter	$K$	$B$	$C$	$WM$	$WUM$	$WLM$	$SM$	$EX$	$KI$	$CS$	$CI$	$CG$	$IMP$
Lower bound	0.1	0.1	0.1	50	1	50	2	0.1	0.01	0.01	0.5	0.5	0.001
Upper bound	1.5	0.4	0.9	200	50	100	100	2	0.7	0.9	0.999	0.999	0.05

We divided the model parameters for each subbasin into four categories according to the characteristics of the model. Their physical significances are explained in the following.

- (1) Evaporation parameters are  $K$ ,  $WUM$ ,  $WLM$ , and  $C$ .  $K$  is the reduction coefficient of the evaporation, which is equal to the ratio of the potential evapotranspiration to the pan evaporation.  $WUM$  and  $WLM$  are areal mean water capacity tensions for the upper and lower layers of the watershed and are in the ranges of 5–20 mm and 60–90 mm, respectively [12, 13]. Their values depend on the condition of the soil and vegetation; rich soil and vegetation result in larger values.  $C$  depends on the proportion of the area that is covered by vegetation with deep roots; a larger value means less evaporation. It takes values in 0.09–0.12 in semihumid and semiarid regions and in 0.15–0.20 in humid regions [12, 13].
- (2) Runoff production parameters are  $IMP$ ,  $B$ , and  $WM$ .  $IMP$  is the ratio of impervious areas (including saturated areas) to the total area of the basin.  $B$  is the inhomogeneity distribution of the water deficit in the vadose zone and is proportional to the inhomogeneity of the water deficit in the watershed.  $WM$  represents the severity of a drought and is the sum of  $WUM$ ,  $WLM$ , and  $WDM$  (in the deepest layer). It takes values in the range 120–180 mm. Note that  $B$  is related to  $WM$ ; if  $WM$  increases,  $B$  decreases (and vice versa).

- (3) Runoff separation parameters are  $SM$ ,  $EX$ ,  $KG$ , and  $KI$ .  $SM$  is the areal mean of the free water capacity of the surface soil layer.  $EX$  is the spatial distribution of the watershed’s free water storage capacity.  $KG$  and  $KI$  are the outflow coefficients for the free water storage to groundwater and interflow relationships, respectively. The sum of  $KG$  and  $KI$  is the free water flow velocity. Because it lasts 3 days to fall, we take  $KG + KI = 0.7$  so that the independence of the parameter problem can be solved.
- (4) Runoff concentration parameters are  $CG$ ,  $CI$ , and  $CS$ . These are regression constants for the surface runoff, interflow runoff, and groundwater runoff, respectively.

The above classification of the parameters into four groups corresponds to different optimization characteristics. The first group influences the total runoff production. Their aim is to balance the rainfall, evaporation, and runoff. The second and third groups determine the runoff production and separation. The fourth group is the most sensitive to changes over time and determines the discharge process.

Parameters within the same group tend to be mutually dependent, whereas those in different groups are relatively independent. Based on the above analysis, parameters of the lower numbered groups should be optimized before those of the higher numbered groups, as was suggested by Zhao [12, 13]. The parameters determined for groups X-X can be directly applied to groups X-X. This is the method used in this paper.

TABLE 2: Yandu River watershed.

Catchment area (km <sup>2</sup> )	Number of rainfall stations	Average annual precipitation	Average annual evaporation	Average annual runoff depth
601	5	1657	741	1281

### 3. Materials and Methods

*3.1. Study Area.* Yandu River watershed is located in the Three Gorges Region of Yangtze River and has a catchment area of 601 km<sup>2</sup>. It has a large watershed slope with an average gradient of 2.87 and an elevation drop of 2,800 m. The watershed is described in Table 2.

There is plentiful precipitation in Yandu River watershed, with an average annual precipitation of 1,337 mm, 68% of which occurs in summer and autumn. The maximum annual rainfall is 2,448.2 mm and the minimum is 808.4 mm. Runoff is mainly due to rain. The flood season is from April to October, and the annual maximum flood peak flow typically occurs between May and July. With high mountains, steep slopes, deep valleys, and narrow rivers, the floods are characterized by rapid confluence and sharp peak flows.

Yandu River watershed belongs to mountain terrain. Affected by human activities since 1990s, obvious change in the climate and underlying surface conditions is happening and confluence process changes accordingly [15]. The measured data does not present the natural condition of watershed; that is, the degree of response of runoff to rainfall is reduced. Historical records of 1981 to 1987 used in this paper could basically meet the demand of the research.

*3.2. Definition of Ideal Data.* Based on the physical significance of the parameters of the hydrological models and the watershed characteristics, we randomly generated a set of parameters within the parameter search interval. We took these as the true values of the model parameters. With this true parameter set and the hydrologic input data (continuous daily rainfall and evaporation data), we generated a sequence of streamflows, which we considered the ideal data (or “observed” streamflows) for the calibration time period. Obviously, by using this ideal data in the parameter optimization, we avoid the influences of errors in the input and output data and the model structure.

*3.3. SCE-UA Method.* The selection of an automatic parameter optimization algorithm in calibration of CCR models has been studied extensively. Virtually a lot of parameter optimization studies have used “local-search” procedures such as the downhill simplex method, the pattern search method, and the rotating directions method. The “optimal” parameters provided by them vary with the choice of starting point. It is now known that the objective function response surface contains hundreds of thousands of local optima. In this study, we used a relatively common algorithm called the SCE-UA method (an abbreviation for the shuffled complex evolution method) developed at The University of Arizona,

by Duan et al. [3, 16, 17]. The SCE-UA strategy combines the strengths of the simplex procedure [18], with controlled random search [19], competitive evolution [20], and complex shuffling. The algorithm with global search has proved to be both effective and relatively efficient and provided superior performance compared with other optimization algorithms, such as the downhill simplex method, the genetic algorithm, and multistart versions of the downhill simplex for parameter identification [3, 4, 16, 17].

The SCE-UA method regards a global search as an evolutionary process. Detailed descriptions and explanations of the method were given by Duan et al. [3, 16] and so will not be repeated here. Briefly, the method begins with a population of points sampled randomly from the feasible parameter space. The population is partitioned into several “complexes,” with each complex containing  $2n + 1$  points, where  $n$  is the dimension of the problem. Each complex “evolves” according to a statistical reproduction process that uses the shape of the simplex to direct the search in an appropriate direction. At periodic stages in the evolution, the entire population is shuffled and points are reassigned to complexes to ensure information sharing. This procedure enhances survivability by sharing search space information that was independently gained by each complex. The processes of competitive evolution and complex shuffling are inherent to the SCE-UA algorithm and help to ensure that the information contained in the sample is efficiently and thoroughly exploited. These properties endow the SCE-UA method with good global convergence properties over a broad range of problems. In other words, given a prespecified number of function evolutions (i.e., a fixed level of efficiency), the SCE-UA method should have a high probability of succeeding in its objective of finding the global optimum.

The SCE-UA method contains some certain and random factors that are controlled by the algorithm’s parameters. These parameters must be carefully chosen so that the method performs optimally. They include the number of points in each complex ( $m$ ), the number of points in each subcomplex ( $q$ ), the number of offspring that can be generated by each subcomplex ( $\alpha$ ), the number of evolution steps for each complex ( $\beta$ ), and the number of complexes ( $p$ ). In theory,  $m$  can take any value greater than 1. However, if  $m$  is too small or there are too few points in a single complex, the search process is the same as the general simplex method. This reduces the possibility of a global search. Conversely, if  $m$  is too big, the method takes too long to compute and is not effective. The SCE-UA algorithm used by Duan et al. (1992, 1994) used the values  $m = 2n + 1$ ,  $q = n + 1$ ,  $\alpha = 1$ , and  $\beta = 2n + 1$ . Hence, the only variable to be specified is the number of complexes ( $p$ ), which depends on the optimization problem. In theory,  $p$  can take any integer greater than or equal to 1. However, if  $p$  is too small for high-dimension problem, the global optima may not be found and if  $p$  is too big, the method takes too long to compute and is not effective. Generally speaking, a more complicated problem needs more complexes to obtain the global optimum; in other words, the value of  $p$  should be increased. The optimization results of  $p$  for the ideal and measured data are shown in Tables 3(a) and 3(b). By fixing some model parameters ( $C = 0.15$ ,  $WM = 120$ ,

WUM = 20, WLM = 70, and IMP = 0.01), we could calculate the other optimum parameters for the ideal data, as shown in Table 4.

As can be seen from Tables 3(a) and 3(b), the algorithm could not find the true parameter values when  $p = 1$  and  $p = 2$ , for both the ideal and measured data. When  $p = 4$  or  $p = 10$ , besides some insensitive parameters, we determined the true values of most parameters for the measured data. When using the ideal data, we required  $p > 4$ .

Table 4 shows that when we fixed some parameters (C, WM, WUM, WLM, and IMP), we could determine the true parameter values if  $p \geq 2$ . This further illustrates that the SCE-UA parameter depends on the dimensions of the optimization problem. In this paper, we used  $p = 4$  for the ideal data and  $p = 10$  for the measured data.

**3.4. Objective Functions.** The selection of an appropriate objective function has been discussed extensively. The selection is full of subjectivity. It would affect the calibration of model if the function lacks considering random factors of data. The most commonly used functions have been simple least squares (SLS), heteroscedastic error maximum likelihood estimation (HMLE), determination coefficient (DY), and multiobjective function. For CRR model, multiobjective function usually is converted into some specific operational function like that which considers the total runoff error and evaluates the water balance or the relative error of the calculated discharge and so on. XAJ model generally could be divided into daily rainfall-runoff simulation and flood simulation. Due to the different task background, there is also a difference of the selection of objective function. The former focuses on the water balance and the latter in addition to the consideration of water balance also focuses on the flood peak simulation. Different functions evaluate different characteristics of the hydrological process. The objective functions directly influence the results of the optimization. The most commonly used functions are as follows. The first considers the total runoff error and evaluates the water balance. It is

$$\text{Obj}_1 = \min \left\{ \left| \frac{\sum_{i=1}^n Q_{\text{obs}}(i) - \sum_{i=1}^n Q_{\text{cal}}(i)}{\sum_{i=1}^n Q_{\text{obs}}(i)} \right| \right\}, \quad (1)$$

where  $Q_{\text{obs}}(i)$  is the measured discharge,  $Q_{\text{cal}}(i)$  is the calculated discharge, and  $n$  is the length of the measured data. The second considers the relative error of the calculated discharge and is defined as

$$\text{Obj}_2 = \min \left\{ \frac{\sum_{i=1}^n |Q_{\text{obs}}(i) - Q_{\text{cal}}(i)|}{\sum_{i=1}^n Q_{\text{obs}}(i)} \right\}. \quad (2)$$

The third is a determination coefficient proposed by [21], which reflects the accuracy of the simulation results. It is defined as

$$\text{DY} = \max \left\{ 1 - \frac{\sum_{i=1}^n [Q_{\text{obs}}(i) - Q_{\text{cal}}(i)]^2}{\sum_{i=1}^n [Q_{\text{obs}}(i) - \bar{Q}_{\text{obs}}]^2} \right\}. \quad (3)$$

A larger DY indicates a better fit between the measured and calculated discharges.

**3.5. Stopping Criterion for the Iterative Process.** Parameter optimization methods are generally iterative. In theory, there is a point in the parameter space that achieves the optimum value of the objective function. However, in practice, we do not know when we have reached this point, and we may not find it at all. We thus require a termination condition so that method finds an optimal parameter value within the optimum running time. The most commonly used stopping criteria have been objective function convergence criterion, parameter convergence, and maximum iteration number.

- (1) The convergence of the objective function:  
that is,

$$\left| \frac{(f_{i-k} - f)}{f_i} \right| \leq \text{TOL}, \quad (4)$$

where  $f_i$  and  $f_{i-k}$  are objective function values for the  $i$ th and the  $(i - k)$ th iterations and TOL is the tolerance. The values of TOL and  $k$  are both determined according to practical conditions. If the function value after  $k$  iterations does not improve the precision, we stop the search. The optimal position is most likely on a relatively flat response surface. The function's convergence is useful when considering inaccurate solutions.

- (2) The convergence of the parameters:  
that is,

$$\left| \frac{(\lambda_{i-k}(j) - \lambda_i(j))}{(\lambda_{\max}(j) - \lambda_{\min}(j))} \right| \leq \text{TOL}_\lambda, \quad (5)$$

where  $\lambda_i(j)$  and  $\lambda_{i-k}(j)$  are the  $j$ th parameter values at the  $i$ th and  $i - k$ th iterations,  $\lambda_{\max}(j)$  and  $\lambda_{\min}(j)$  are the maximum and minimum values of  $j$  for all parameter combinations, and  $\text{TOL}_\lambda$  is a tolerance value. If the parameter does not obviously change over  $k$  iterations, we stop the search.

- (3) The maximum iteration number:

to prevent a dead cycle, we set a maximum number of iterations in advance. If we exceed this, we stop the search.

The parameter convergence criterion is more suitable for parameter optimization, because it terminates when the parameters are not obviously changing [22]. The maximum iteration criterion should also be used to avoid wasting computational time. If the algorithm does not terminate within a reasonable number of iterations, we must check the computational procedure.

We controlled the iterations using the following values:

- (i) for the objective functions,  $k = 10$  and  $\text{TOL} = 10^{-6}$ ;
- (ii) for the parameter convergence,  $k = 10$  and  $\text{TOL}_\lambda = 10^{-4}$ ;
- (iii) for the maximum iteration number,  $k_{\max} = 10^6$ .

TABLE 3: (a) Parameter optimization results using ideal data. (b) Parameter optimization results using measured data.

(a)													
Parameter	$K$	$B$	$C$	WM	WUM	WLM	SM	EX	KI	CS	CI	CG	IMP
True value	<b>1.027</b>	<b>0.204</b>	<b>0.285</b>	<b>162.077</b>	<b>35.262</b>	<b>71.240</b>	<b>11.888</b>	<b>1.436</b>	<b>0.420</b>	<b>0.497</b>	<b>0.608</b>	<b>0.610</b>	<b>0.039</b>
$p = 1$	<b>1.026</b>	<b>0.207</b>	<b>0.102</b>	<b>162.919</b>	<b>32.687</b>	<b>79.639</b>	<b>11.995</b>	<b>1.443</b>	<b>0.681</b>	0.497	<b>0.603</b>	<b>0.767</b>	<b>0.037</b>
$p = 2$	1.027	0.204	<b>0.280</b>	<b>162.167</b>	<b>35.036</b>	<b>72.046</b>	<b>11.889</b>	<b>1.437</b>	<b>0.222</b>	0.497	0.609	0.609	0.039
$p = 4$	1.027	0.204	<b>0.280</b>	162.077	35.263	71.240	11.888	1.436	0.420	0.497	0.609	0.610	0.039
$p = 10$	1.027	0.204	0.285	162.077	35.262	71.240	11.888	1.436	0.420	0.497	0.608	0.610	0.039

(b)													
Parameter	$K$	$B$	$C$	WM	WUM	WLM	SM	EX	KI	CS	CI	CG	IMP
True value	<b>0.923</b>	<b>0.313</b>	<b>0.192</b>	<b>146.840</b>	<b>38.159</b>	<b>81.504</b>	<b>19.759</b>	<b>1.623</b>	<b>0.613</b>	<b>0.323</b>	<b>0.645</b>	<b>0.964</b>	<b>0.038</b>
$p = 1$	<b>0.919</b>	<b>0.309</b>	<b>0.177</b>	<b>154.368</b>	<b>39.999</b>	<b>76.720</b>	<b>19.642</b>	<b>1.640</b>	0.613	0.324	0.645	0.964	<b>0.034</b>
$p = 2$	<b>0.920</b>	<b>0.309</b>	<b>0.164</b>	<b>154.436</b>	<b>37.699</b>	<b>81.651</b>	<b>19.672</b>	<b>1.641</b>	0.612	0.323	0.645	0.964	<b>0.034</b>
$p = 4$	0.923	0.313	<b>0.146</b>	146.840	38.159	81.504	19.759	1.623	0.613	0.323	0.645	0.964	0.038
$p = 10$	0.923	0.313	<b>0.196</b>	146.840	38.159	81.504	19.759	1.623	0.613	0.323	0.645	0.964	0.038

TABLE 4: Optimization results for the remaining parameters after fixing some parameters, using the ideal data.

Parameter	$K$	$B$	SM	EX	KI	CS	CI	CG
True value	<b>1.025</b>	<b>0.192</b>	<b>6.036</b>	<b>1.825</b>	<b>0.511</b>	<b>0.215</b>	<b>0.607</b>	<b>0.894</b>
$p = 1$	1.025	0.192	<b>6.04</b>	1.825	<b>0.51</b>	0.215	<b>0.606</b>	0.894
$p = 2$	1.025	0.192	6.036	1.825	0.511	0.215	0.607	0.894
$p = 4$	1.025	0.192	6.036	1.825	0.511	0.215	0.607	0.894
$p = 10$	1.025	0.192	6.036	1.825	0.511	0.215	0.607	0.894

If any criterion is satisfied, the optimization terminates. The experience with the SCE-UA algorithm indicated that after about a set number of shuffling iterations, the parameter estimates would stabilize in a region where search would subsequently terminate due to parameter convergence or maximum iteration criterion to avoid wasting computational time.

## 4. Results and Discussion

### 4.1. Parameter Optimization Using Ideal Data

4.1.1. *The Effect of the Objective Function.* We considered the effect of the objective function on parameter optimization for daily rainfall-runoff and flood simulations.

For the daily rainfall-runoff simulation, we used  $Obj_1$ ,  $Obj_2$ , DY, and a simple combination of  $Obj_1$  and  $Obj_2$  as the objective functions. We used default values for the algorithm parameters to calibrate the model and ran the algorithm 10 times for each objective function. Table 5 shows the parameter optimization results for the ideal data from 1981–1985 for the Yandu River watershed, as calculated by the SCE-UA method.

From Table 5, we can see that the objective functions can all be applied to the parameter optimization using ideal data, except for  $Obj_1$ , which could not find the true values.

Table 5 shows that the parameter optimization results are consistent with the true values, in addition to  $C$ . We ran the optimization on the ideal data to remove the effect of errors

in the data and model structure. There is a big difference between the calibrated and true values for  $C$ , mainly because it is a coefficient for deep evapotranspiration and depends on the coverage area of deep-rooted plants. Deep evaporation is rare in humid areas, so  $C$  is not sensitive enough for the optimization to obtain its true value.

Less iterations were required when using DY. However,  $Obj_2$  produced objective function values that were closer to the true values. The results obtained using  $Obj_2$  were (0.000010, 0.000010, 0.000010, 0.000010, 0.000009, 0.000009, 0.000010, 0.000010, 0.000010) and using DY were (0.000073, 0.000071, 0.000048, 0.000033, 0.000067, 0.000105, 0.000081, 0.000040, 0.000054, 0.000079). Considering the optimization speed, DY was quicker to achieve the optimal value.

It is difficult to find the true value using only  $Obj_1$ , because it reflects the total error between the measured and calculated total flow and focuses on the overall water balance. Because of the mutual compensation between parameters, more than one set of parameters can satisfy the total water balance. That is, the optimization results are not unique. To obtain a good combination of parameters, we should simultaneously use other functions.

These results show that correlations between parameters are important to parameter optimization. The choice of objective function has an impact on the parameter optimization, and the results vary with different functions. To exclude the influence of errors in the data and model structure we used the ideal data to investigate the daily rainfall-runoff model.

TABLE 5: Characteristics of the optimization results for the daily rainfall-runoff simulation.

Parameter	True value	Obj <sub>1</sub>			Obj <sub>2</sub>			DY		
		Maximum	Average	Variance	Maximum	Average	Variance	Maximum	Average	Variance
<i>K</i>	<b>0.533</b>	0.538	0.533	$9.2E - 06$	0.533	0.533	0	0.533	0.533	0
<i>B</i>	<b>0.306</b>	0.323	0.306	$5.3E - 05$	0.306	0.306	0	0.306	0.306	0
<i>C</i>	<b>0.149</b>	0.253	0.195	$9.5E - 04$	0.3	0.177	$5.4E - 03$	0.286	0.203	$3.0E - 03$
WM	<b>153.246</b>	155.029	153.828	$8.3E - 01$	153.246	153.246	0	153.246	153.246	0
WUM	<b>18.286</b>	20.798	19.499	$3.9E - 01$	18.286	18.286	0	18.286	18.286	0
WLM	<b>89.436</b>	86.524	80.176	$2.6E + 01$	89.436	89.436	0	89.436	89.436	0
SM	<b>14.942</b>	24.203	15.175	$2.3E + 01$	14.942	14.942	0	14.942	14.942	0
EX	<b>0.72</b>	1.908	1.479	$7.4E - 02$	0.72	0.72	0	0.72	0.72	0
KI	<b>0.665</b>	0.693	0.508	$1.4E - 02$	0.665	0.665	0	0.665	0.665	0
CS	<b>0.058</b>	0.408	0.222	$7.1E - 03$	0.058	0.058	0	0.058	0.058	0
CI	<b>0.54</b>	0.656	0.559	$1.8E - 03$	0.54	0.54	0	0.54	0.54	0
CG	<b>0.548</b>	0.854	0.735	$5.8E - 03$	0.548	0.548	0	0.548	0.548	0
IMP	<b>0.03</b>	0.042	0.029	$7.2E - 05$	0.03	0.03	0	0.03	0.03	0
Average iterations			84			522			139	

Note that the average iterations are the average number of evolutions of complexes.

The SCE-UA method found the true parameter values when using an appropriate objective function.

For flood simulation, we used Obj<sub>2</sub>, DY, and a simple combination of Obj<sub>1</sub> and Obj<sub>2</sub> as objective functions. We first fixed the parameters that did not depend on the computing time ( $K = 0.533$ ,  $WUM = 18.286$ ,  $WLM = 89.436$ ,  $C = 0.149$ ,  $WM = 153.246$ ,  $B = 0.306$ ,  $IMP = 0.03$ , and  $EX = 0.72$ ). We then optimized the remaining parameters for flood simulation. The results calculated using these objective functions are listed in Table 6.

Table 6 shows that the global optimum was achieved with these objective functions, because the targets of these objective functions are consistent with the flood simulation.

*4.1.2. The Effect of the Data Length.* We randomly generated a set of parameters within the parameter search space for the XAJ model and used these as the true values (see Table 7). Then, we calculated the ideal data using the true values and the daily areal rainfall for 1981–1985, to study the effect of the data length on parameter optimization. The results of the parameters optimized using Obj<sub>2</sub> for different lengths of ideal data are shown in Table 7.

The results are generally consistent with the true values (except for *C*), regardless of the length of the data (1, 3, or 5 years). If we exclude the impact of measured data and model structure errors, automatic optimization methods can obtain the global optimum when using reasonable algorithm parameters. That is, for ideal data, global optimization methods obtained the global optimum.

Table 7 shows that the results calculated for 1, 3, and 5 years of data are almost equal. So the length of the data did not affect the optimization method.

As previously mentioned, it is difficult to obtain the true value of *C*. However, as mentioned above, avoiding the

influences of errors in the input and output data and the model structure by using ideal data, the result for KI is still different from its true value, which cannot be attributed to data or model structure errors. The large differences in the calibrated and true values for KI are mainly because the water source has been divided by the structural constraint that KI and KG are independent. This demonstrates that the correlations between parameters are important to parameter optimization.

To more clearly reveal the parameter optimization process, Figure 2 shows the optimization results using ideal data from 1981 (1 year), 1981–1983 (3 years), and 1981–1985 (5 years). For illustrative purposes, we normalized all the parameter values using  $(x - x_{\min}) / (x_{\max} - x_{\min}) = x_{\text{new}}$ , so that they are between 0 and 1.

Figure 2 demonstrates that, for ideal data, the parameters ultimately converged to the global optimum, regardless of the data length. However, some parameters had long time fluctuations and were not stable (e.g., WUM and WLM). Although KI was constant for a long period, there were signs of volatility in later iterations. *C* had an irregular shock in the optimization process and did not converge. This illustrates that insensitive parameters and the correlations between parameters have important influences on the results of the optimization.

## 4.2. Parameter Optimization Using Measured Data

*4.2.1. The Effect of Objective Function.* As previously mentioned, if the objective function is selected appropriately, the true parameter values can be obtained by a direct search. We used Obj<sub>1</sub>, Obj<sub>2</sub>, and their combination as objective functions and used observed data from 1981–1985. We ran

TABLE 6: Parameter optimization results for the flood simulation.

Parameter	SM	KI	CS	CI	CG	KE	XE
True value	34.958	0.135	0.284	0.631	0.875	1.137	0.313
Obj <sub>2</sub>	34.958	0.135	0.284	0.631	0.875	1.137	0.313
DY	34.958	0.135	0.284	0.631	0.875	1.137	0.313
Obj <sub>1</sub> + Obj <sub>2</sub>	34.958	0.135	0.284	0.631	0.875	1.137	0.313

TABLE 7: Parameter optimization results.

Parameter	True value	Bound		Optimization results of parameters		
		Lower	Upper	1 year	3 years	5 years
<i>K</i>	0.681	0.2	1.5	0.681	0.681	0.681
<i>B</i>	0.49	0.1	0.9	0.49	0.49	0.49
<i>C</i>	0.254	0.1	0.9	0.632	0.625	0.627
WM	152.334	50	200	152.334	152.334	152.334
WUM	67.516	10	100	67.516	67.516	67.516
WLM	53.000	10	100	53.000	53.000	53.000
SM	74.913	1	100	74.913	74.913	74.913
EX	1.356	0.1	2	1.356	1.356	1.356
KI	0.683	0.01	0.7	0.513	0.187	0.604
CS	0.801	0.01	0.9	0.801	0.801	0.801
CI	0.891	0.5	0.999	0.891	0.891	0.891
CG	0.988	0.5	0.999	0.988	0.988	0.988
IMP	0.033	0.001	0.05	0.033	0.033	0.033

the algorithm ten times for each objective function. The optimization results are listed in Table 8.

Table 8 shows that the results were different for the considered objective functions and that the values of some parameters (*C* and WM) were different when using the same objective function. Using measured data, the parameter optimization results were diverse because of the influences of errors in the data and the model structure.

**4.2.2. The Effect of Data Length.** As previously mentioned, the data length has no effect on the optimization when using ideal data. We selected the combination function as the objective function and applied it to measured data from 1981 to 1987. We ran the algorithm ten times and compared the results. We considered that if the algorithm converged to the same group parameters all ten times, we had found the optimum value. The results are listed in Table 9.

The parameter optimization results using different lengths of measured data were different. In some cases, all instances of the same data length were different (seven situations when using one year of measured data, six times when using two consecutive years, and five times when using three consecutive years).

The cumulative distributions of some parameters using different lengths of measured data are shown in Figure 3. The *y*-coordinates indicate the cumulative distribution functions, and the *x*-coordinates indicate the parameter. The vertical dotted line represents the optimal parameter calculated using the measured data from 1981 to 1987. "One year" indicates

the optimal parameter calculated using one year of measured data and so on.

Figure 3 shows that the cumulative distributions of the parameters changed irregularly as the length of the measured data increased. That is, the parameter optimization did not converge or become stable as the length of the measured data increased.

## 5. Conclusions

Several observations are noted from calibrating XAJ model with the SCE-UA optimization method.

Models represent the core of hydrological forecasting, and parameter identification is key to applying a model. In this paper, we investigated parameter optimization for the XAJ model using the SCE-UA algorithm. On the basis of previous research, we used ideal data to study the relationships between the data length, search range of the parameters, objective functions, and parameter stability. The results indicated that the calculated global optimal parameters did not depend on the data length but did depend on the objective functions. For the daily rainfall-runoff model, the optimum values could not be determined using only the water balance objective function. However, the global optimal values could be found when using the appropriate objective function. The global optimization algorithms did not have the same accuracy. For flood simulation, we fixed the parameters that do not depend on time and found that the choice of objective function did not affect the results.

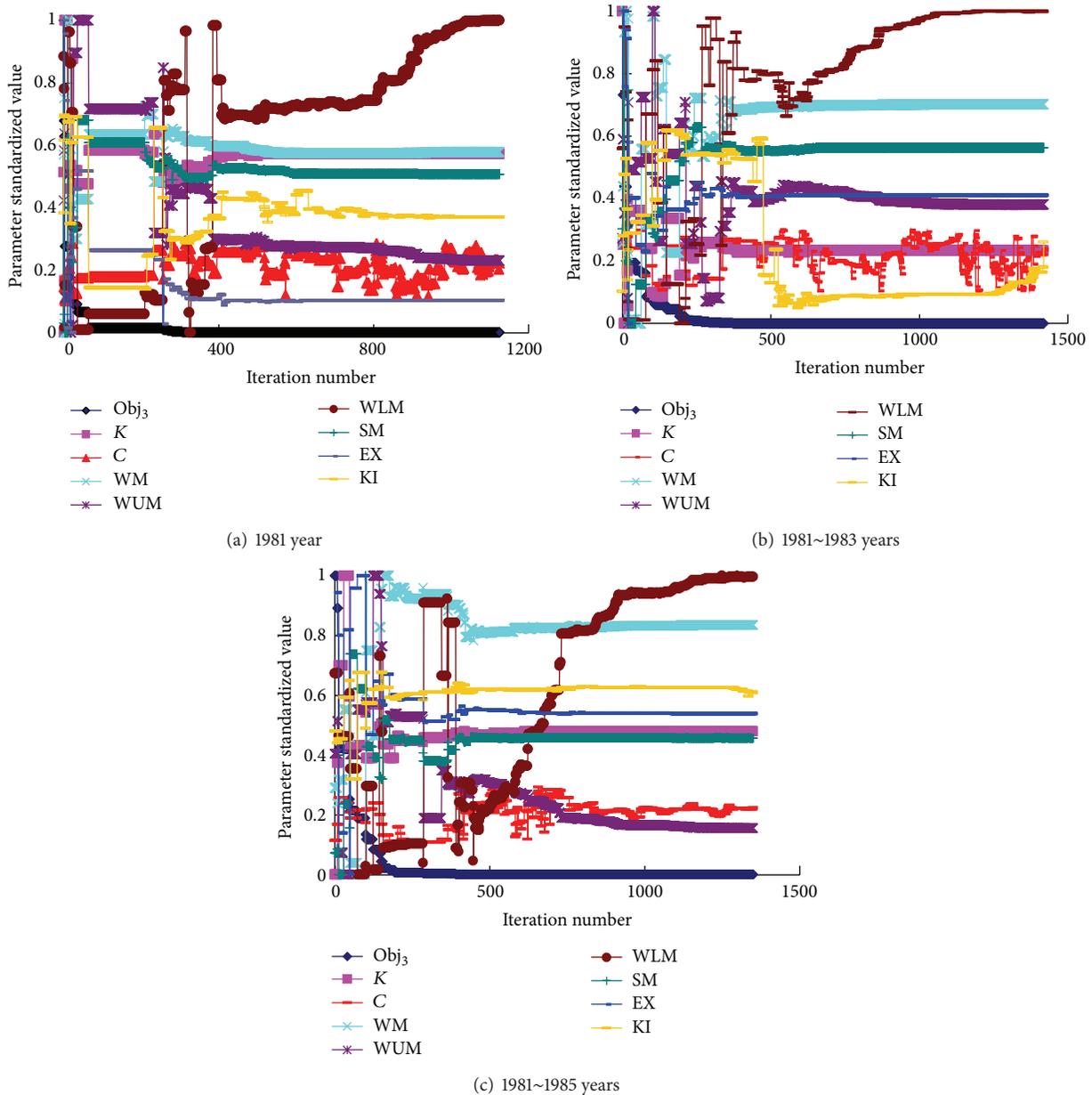


FIGURE 2: Convergence of the optimization using ideal data.

When using observed data, the optimization results were different when using different objective functions and were even different when using the same objective function. The values were significantly different because of the influence of errors in the data and model structure. Although the data length had no effect when using ideal data, when using observed data the parameter optimization results were different. That is, the parameter optimization did not converge to a certain value as the length of the measured data increased. In summary, errors in the data and model structure lead to uncertainties in the parameter optimization.

An approach that has the flexibility of emphasizing different portions of the model residuals according to one's preference is multiobjective optimization. The advantage of a

multiobjective approach is to explicitly focus attention on the model performance. We will focus on this issue in the future.

### XAJ Model Parameters

- K: Ratio of potential evapotranspiration to pan evaporation
- WM: Areal mean tension water capacity
- WUM: Upper layer areal mean tension water capacity
- WLM: Lower layer areal mean tension water capacity
- C: Coefficient of deep evapotranspiration

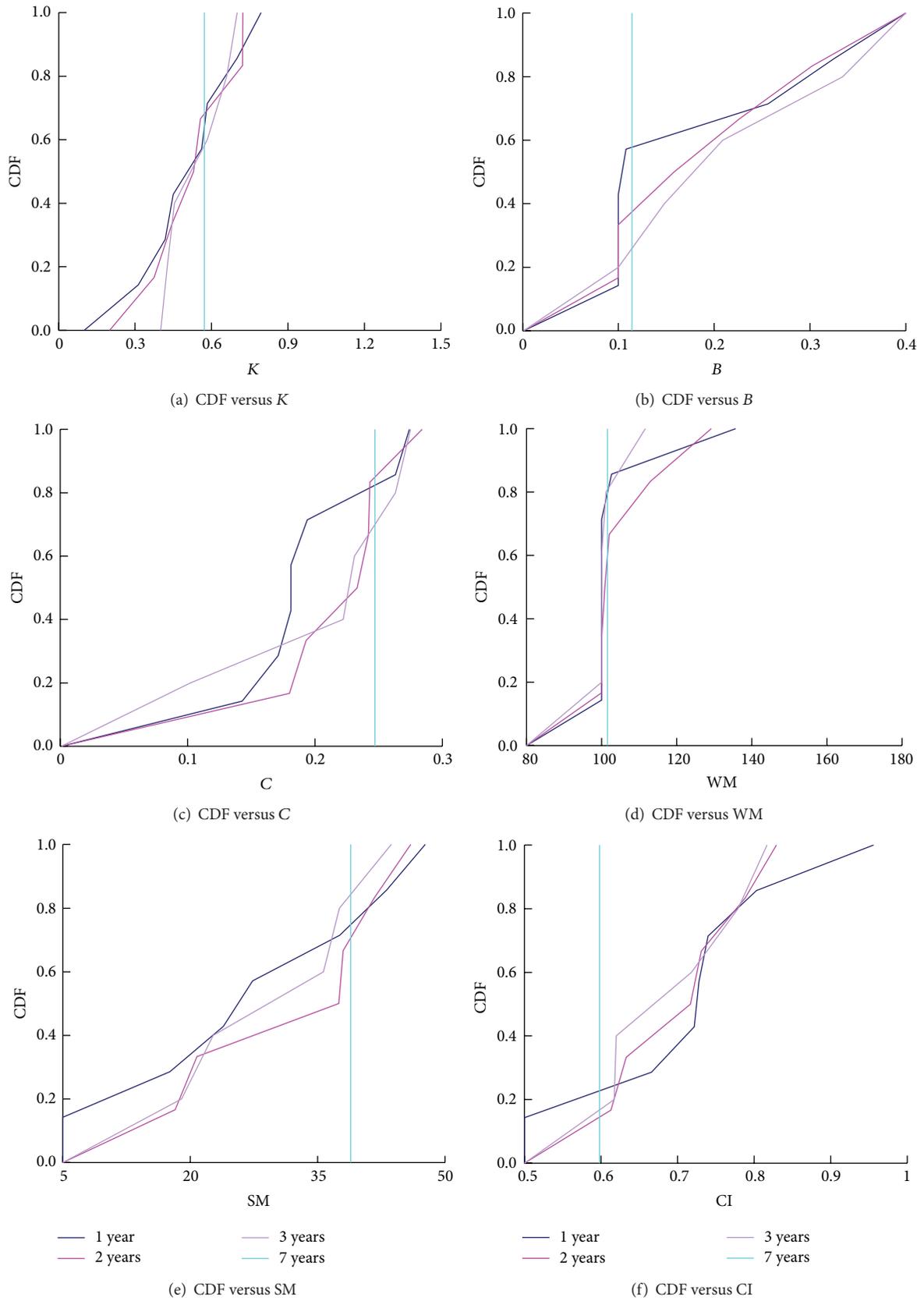


FIGURE 3: Cumulative distribution functions of the parameters.

TABLE 8: Characteristics of the optimization results.

Parameter	<i>K</i>	<i>B</i>	<i>C</i>	WM	WUM	CS	CI
Obj <sub>1</sub>							
Minimum	0.589	0.34	0.117	119.567	11.729	0.82	0.836
Maximum	0.604	0.395	0.264	122.744	13.454	0.82	0.851
Average	<b>0.595</b>	<b>0.377</b>	<b>0.191</b>	<b>121.112</b>	<b>12.394</b>	<b>0.82</b>	<b>0.841</b>
Square deviation	2.50E – 05	2.10E – 04	1.60E – 03	1.30E + 00	2.20E – 01	0	2.50E – 05
Obj <sub>2</sub>							
Minimum	0.556	0.175	0.103	100	10	0.138	0.753
Maximum	0.556	0.175	0.265	100	10	0.166	0.753
Average	<b>0.556</b>	<b>0.175</b>	<b>0.171</b>	<b>100</b>	<b>10</b>	<b>0.155</b>	<b>0.753</b>
Square deviation	0	0	2.30E – 03	0	0	1.00E – 04	0
Obj <sub>1</sub> + Obj <sub>2</sub>							
Minimum	0.537	0.192	0.134	100.574	10.574	0.01	0.598
Maximum	0.537	0.192	0.279	100.574	10.574	0.01	0.598
Average	<b>0.537</b>	<b>0.192</b>	<b>0.207</b>	<b>100.574</b>	<b>10.574</b>	<b>0.01</b>	<b>0.598</b>
Square deviation	0	0	1.80E – 03	0	0	0	0

TABLE 9: Parameter optimization results using measured data.

Parameter	1981–1987	One year			Two consecutive years			Three consecutive years		
		1981	...	1987	1981-1982	...	1986-1987	1981–1983	...	1985–1987
<i>K</i>	0.571	0.449		0.418	0.444		0.529	0.454		0.583
<i>B</i>	0.114	0.108		0.325	0.226		0.1	0.1		0.148
<i>C</i>	0.247	0.194		0.274	0.243		0.284	0.222		0.102
WM	101.539	100.001		100	112.963		100.84	111.657		101.18
WUM	11.539	10		10	22.962		10.839	21.656		11.18
WLM	60	60		60	60		60	60		60
SM	38.901	23.895	∴	43.129	45.968	∴	41.708	37.55	∴	43.673
EX	1.494	0.251		1.311	2		1.458	1.823		1.464
KI	0.447	0.482		0.507	0.504		0.503	0.475		0.543
CS	0.01	0.01		0.081	0.01		0.056	0.01		0.053
CI	0.598	0.666		0.728	0.613		0.717	0.62		0.718
CG	0.957	0.987		0.999	0.976		0.998	0.964		0.998
IMP	0.05	0.05		0.05	0.032		0.045	0.05		0.05

- IMP: Ratio of impervious area to total area of basin
- B*: Exponent of the tension water capacity curve
- SM: Areal mean of the free water capacity of the surface soil layer
- EX: Exponent of free water capacity curve
- KI: Outflow coefficients of free water storage to interflow relationships
- KG: Outflow coefficients of free water storage to groundwater relationships
- CS: Recession constant for routing through the channel system within each subbasin
- CI: Recession constant of the lower interflow storage
- CG: Recession constant of groundwater
- N*: Number of subreaches
- KE: Travel time of flood wave through subreach
- XE: Flow ration of subreaches.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgments

The study is financially supported by the National Natural Science Foundation of China (no. 51409082), the China Postdoctoral Science Foundation (2013M531264), and the Fundamental Research Funds for the Central Universities (no. 2014B04914).

### References

[1] Q. J. Wang, “The genetic algorithms and its application to calibrating conceptual rainfall-runoff models,” *Water Resources Research*, vol. 27, no. 9, pp. 2467–2471, 1991.

- [2] M. Franchini, "Use of a genetic algorithm combined with a local search method for the automatic calibration of conceptual rainfall-runoff models," *Hydrological Sciences Journal*, vol. 41, no. 1, pp. 21–37, 1996.
- [3] S. Sorooshian and V. Gupta, "Effective and efficient global optimization for conceptual rainfall-runoff models," *Water Resources Research*, vol. 28, no. 4, pp. 1015–1031, 1992.
- [4] S. Sorooshian, Q. Duan, and V. K. Gupta, "Calibration of rainfall-runoff models: application of global optimization to the Sacramento soil moisture accounting model," *Water Resources Research*, vol. 29, no. 4, pp. 1185–1194, 1993.
- [5] P. O. Yapo, H. V. Gupta, and S. Sorooshian, "Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data," *Journal of Hydrology*, vol. 181, no. 1–4, pp. 23–48, 1996.
- [6] T. Y. Gan, E. M. Dlamini, and G. F. Biftu, "Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling," *Journal of Hydrology*, vol. 192, no. 1–4, pp. 81–103, 1997.
- [7] V. A. Cooper, V.-T.-V. Nguyen, and J. A. Nicell, "Calibration of conceptual rainfall-runoff models using global optimisation methods with hydrologic process-based parameter constraints," *Journal of Hydrology*, vol. 334, no. 3–4, pp. 455–466, 2007.
- [8] M. K. Gill, Y. H. Kaheil, A. Khalil, M. McKee, and L. Bastidas, "Multiobjective particle swarm optimization for parameter estimation in hydrology," *Water Resources Research*, vol. 42, no. 7, Article ID W07417, 2006.
- [9] C. Zhang and Y.-Y. Sun, "Automatic calibration of conceptual rainfall-runoff models," *Applied Mechanics and Materials*, vol. 405–408, pp. 2185–2189, 2013.
- [10] J. A. Vrugt, H. V. Gupta, L. A. Bastidas, W. Bouten, and S. Sorooshian, "Effective and efficient algorithm for multiobjective optimization of hydrologic models," *Water Resources Research*, vol. 39, no. 8, pp. SWC51–SWC519, 2003.
- [11] B. A. Tolson and C. A. Shoemaker, "Dynamically dimensioned search algorithm for computationally efficient watershed model calibration," *Water Resources Research*, vol. 43, no. 1, 2007.
- [12] R. J. Zhao, Y. L. Zhuang, L. R. Fang et al., "The Xin'anjiang model in hydrological forecasting," in *Proceeding of the Oxford Symposium*, vol. 129, pp. 351–356, IAHS Publication, April 1980.
- [13] R. J. Zhao, "The Xin'anjiang model applied in China," *Journal of Hydrology*, vol. 135, no. 1–4, pp. 371–381, 1992.
- [14] V. T. Chow, *Handbook of Applied Hydrology*, McGraw-Hill, New York, NY, USA, 1964.
- [15] N. J. Potter and F. H. S. Chiew, "An investigation into changes in climate characteristics causing the recent very low runoff in the southern Murray-Darling Basin using rainfall-runoff models," *Water Resources Research*, vol. 47, no. 10, 2011.
- [16] Q. Y. Duan, V. K. Gupta, and S. Sorooshian, "Shuffled complex evolution approach for effective and efficient global minimization," *Journal of Optimization Theory and Applications*, vol. 76, no. 3, pp. 501–521, 1993.
- [17] Q. Duan, S. Sorooshian, and V. K. Gupta, "Optimal use of the SCE-UA global optimization method for calibrating watershed models," *Journal of Hydrology*, vol. 158, no. 3–4, pp. 265–284, 1994.
- [18] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965.
- [19] W. L. Price, "Global optimization algorithms for a CAD workstation," *Journal of Optimization Theory and Applications*, vol. 55, no. 1, pp. 133–146, 1987.
- [20] J. H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, Mich, USA, 1975.
- [21] J. E. Nash and J. V. Sutcliffe, "River flow forecasting through conceptual models part I—a discussion of principles," *Journal of Hydrology*, vol. 10, no. 3, pp. 282–290, 1970.
- [22] V. P. Singh, *Computer Models of Watershed Hydrology*, Water Resources Publication, Boulder, Colo, USA, 1995.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

