Front. Earth Sci. 2014, 8(2): 264–281
DOI 10.1007/s11707-014-0424-0

RESEARCH ARTICLE

Parameter identification and calibration of the Xin'anjiang model using the surrogate modeling approach

Yan YE[1,5], Xiaomeng SONG (✉)[2,3], Jianyun ZHANG[2,3], Fanzhe KONG[4], Guangwen MA[5]

1 College of Resources and Environment, Southwest University, Chongqing 400716, China
2 State Key Laboratory of Hydrology-Water Resources & Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing 210029, China
3 Research Center for Climate Change, the Ministry of Water Resources, Nanjing 210029, China
4 School of Resource and Earth Science, China University of Mining & Technology, Xuzhou 221008, China
5 College of Water Resource & Hydropower, Sichuan University, Chengdu 610065, China

Abstract Practical experience has demonstrated that single objective functions, no matter how carefully chosen, prove to be inadequate in providing proper measurements for all of the characteristics of the observed data. One strategy to circumvent this problem is to define multiple fitting criteria that measure different aspects of system behavior, and to use multi-criteria optimization to identify non-dominated optimal solutions. Unfortunately, these analyses require running original simulation models thousands of times. As such, they demand prohibitively large computational budgets. As a result, surrogate models have been used in combination with a variety of multi-objective optimization algorithms to approximate the true Pareto-front within limited evaluations for the original model. In this study, multi-objective optimization based on surrogate modeling (multivariate adaptive regression splines, MARS) for a conceptual rainfall-runoff model (Xin'anjiang model, XAJ) was proposed. Taking the Yanduhe basin of Three Gorges in the upper stream of the Yangtze River in China as a case study, three evaluation criteria were selected to quantify the goodness-of-fit of observations against calculated values from the simulation model. The three criteria chosen were the Nash-Sutcliffe efficiency coefficient, the relative error of peak flow, and runoff volume (REPF and RERV). The efficacy of this method is demonstrated on the calibration of the XAJ model. Compared to the single objective optimization results, it was indicated that the multi-objective optimization method can infer the most probable parameter set. The results also demonstrate that the use of surrogate-modeling enables optimization that is much more efficient; and the total computational cost is reduced by about 92.5%, compared to optimization without using surrogate modeling. The results obtained with the proposed method support the feasibility of applying parameter optimization to computationally intensive simulation models, via reducing the number of simulation runs required in the numerical model considerably.

Keywords Xin'anjiang model, parameter calibration, multi-objective optimization, surrogate modeling

1 Introduction

Computer models comprise mathematical relations, data, and a calculation core, in order to simulate the behavior of a real-world system (Song et al., 2012c, 2013). Modeling of the rainfall-runoff process, which is dominated by watershed geomorphic and rainfall characteristics (Lee et al., 2008) owing to the tremendous spatial and temporal variability of watershed characteristics and precipitation patterns, and the number of variables (Jiang et al., 2013), has been the subject of research among hydrologists and engineers for a very long time (Srinivasulu and Jain, 2006; Song et al., 2012b). There is no doubt that a watershed hydrologic model is one of the most important tools in flood simulation, operational forecasting, and the detailed study of watershed response (Song and Kong, 2010; Song et al., 2012d; Yao et al., 2012; Xu et al., 2013). In recent decades, a large number of rainfall-runoff models have been developed, representing the main physical processes involved in the transformation of rainfall into runoff (Jones et al., 2008; Lü et al., 2013). Among these models, conceptual hydrologic models, as the most simple and effective rainfall-runoff modeling tools, have been widely

Received July 17, 2013; accepted October 17, 2013

E-mail: xmsong.cn@gmail.com

applied in the world, such as the Sacramento model, the HBV model (Lindström et al., 1997), and the Xin'anjiang (XAJ) model (Zhao, 1992). Such models are used for various applications, including hydrologic process studies (Bingeman et al., 2006; Li et al. 2009), runoff estimates (Xu, 1999), soil moisture variations (Liu et al., 2007; Yu et al., 2010), and statistical analyses of hydrological processes in ungauged basins (Akhtar et al., 2008).

Generally, conceptual hydrologic models use a number of interconnected mathematical functions to approximate the soil moisture accounting phase of the hydrologic cycle at a basin scale. These models typically have ten or more parameters, some model input, some interim state variables, and some model output (Vrugt et al., 2006). Unfortunately, these parameters cannot be obtained directly, or measured from basin characteristics. A set of model parameters are determined by gradually adjusting the model parameters to match data, which is also known as parameter calibration or optimization (McCabe et al., 2005; Yang et al., 2011a). For the parameter calibration of hydrological modeling, the quality of the final model parameters derived depends primarily on four major factors: 1) the conceptual basis and structure of the model, 2) the power and robustness of the optimization algorithm (Yang et al., 2009), 3) the quality and the amount of information contained in the data available for model calibration and validation, and 4) the evaluation criteria used in the optimization procedure (Gan and Biftu, 1996; Yang et al., 2011b). In addition, the effect of the number or dimension of parameters on calibration, and the problem of equifinality and uncertainty (Song et al., 2011, 2012a) should not be ignored. Therefore, sensitivity and uncertainty analysis, as well as optimization for hydrological model parameters, should be considered simultaneously in the application.

Traditional calibration of hydrologic models is normally done with a single objective function. However, practical experience reveals that single-objective functions are often subjective, and inadequate to properly measure all of the characteristics of a hydrologic system (Vrugt et al., 2003b; Guo et al., 2013a). Therefore, it stands to reason that the calibration of hydrologic models should be made with multiple objective functions, which reflect different aspects of the system's characteristics. In contrast to one optimal solution provided by single-objective optimization, the result of multi-objective optimization will not be a single best solution, but consists of a set of Pareto solutions. The important part for multi-objective optimization is to seek these approximate Pareto solutions. Recently, various multi-objective evolutionary algorithms have been developed in hydrological modeling (e.g., Yapo et al., 1998; Reed et al., 2003; Vrugt et al., 2003b; Khu and Madsen, 2005; de Vos and Rientjes, 2008, Dumedah et al., 2010; Dumedah 2012; Song et al., 2012c; Guo et al., 2013a). Over the years, evolutionary algorithms have been recognized and established as core optimization techniques for handling complex, multi-objective optimization problems, for example, the non-dominated sorting genetic algorithm (NSGA-II), Pareto archived evolutionary strategy (PAES), strength Pareto evolutionary algorithm (SPEA2), multi-objective shuffled complex evolution metropolis algorithm (MOSCEM), multi-objective shuffled complex differential evolution algorithm (MOSCDE), etc.

Nevertheless, calibrating parameters and evaluating their uncertainty in these models requires running the simulation models thousands of times, which also requires prohibitively large computational budgets (Razavi et al., 2012; Song et al., 2012c, 2013; Zhan et al., 2013). For this reason, surrogate modeling, or the meta-model approach, a second level of abstraction of the real-world system, is concerned with developing and utilizing cheaper-to-run "surrogates" of the "original" simulation models. Surrogate modeling has become increasingly popular over the last decade within the water resources community; and this is consistent with the increasing utilization of meta-models in the scientific literature since 1990 (Razavi et al., 2012).

The main motivation of developing surrogate modeling strategies is to make better use of the available, typically limited, computational budget. Many iterative water resources modeling analyses potentially stand to benefit from surrogate modeling (e.g., Khu and Werner, 2003; Broad et al., 2005, 2010; Mugunthan and Shoemaker, 2006; Yan and Minsker, 2006, 2011; Zou et al., 2007; Bliznyuk et al., 2008; Zhang et al., 2009, etc.). An incomplete list of classic or popular iterative modeling analyses in water resources, which are candidates for efficiency enhancement with surrogate modeling, include deterministic model parameter optimization (calibration) studies with evolutionary algorithms (e.g., Wang, 1991; Duan et al., 1992), uncertainty-based or Bayesian model calibration studies (e.g., Beven and Freer, 2001; Kavetski et al., 2006; Mugunthan and Shoemaker, 2006; Bliznyuk et al., 2008), multi-objective optimization algorithms (e.g., Reed et al., 2003; Zou et al., 2007; Razavi et al., 2012; Song et al., 2012c), global sensitivity analysis methods (e.g., Song et al., 2012d, 2013; Zhan et al., 2013), and any traditional Monte Carlo-based reliability or uncertainty analysis (e.g., Skaggs and Barry, 1997). Global optimization algorithms based on surrogate models have been developed to circumvent the computational budget limitations associated with computationally intensive simulation models, such as EGO (Efficient Global Optimization) (Jones et al., 1998), GMSRBF (Global Metric Stochastic Radial Basis Function), MLSLRBF (Multi Level Single Linkage Radial Basis Function) (Regis and Shoemaker, 2007), and Gutmann's method (Gutmann, 2001).

Generally, example surrogate modeling-enabled multi-objective optimization algorithms are formed by 1) fitting response surface surrogate models to only one computa-

tionally expensive objective (or constraint) when other objectives are fast to run (Behzadian et al., 2009), 2) aggregating multiple objectives into one response function to be approximated by a single response surface surrogate (di Pierro et al., 2009; Zhang et al., 2010), and 3) using multiple surrogate models for multiple objectives (Wagner et al., 2010). In our study, the multiple surrogate models for multiple objectives were used to calibrate a conceptual hydrological model. However, the use of multiple surrogate models for multiple objectives has the potential to increase problems with inaccuracy, and definitely increases modeling time (Razavi et al., 2012). It is commonly assumed that for utilizing multiple surrogates, the approximation errors or uncertainties of these multiple surrogates are independent, despite the fact that the objectives are typically conflicting (Wagner et al., 2010). As a result, the uncertainty of the multi-objective optimization results based on surrogate models is also discussed.

The primary objective of our work is: (i) to construct the surrogate models for a conceptually-based distributed XAJ model, (ii) to compare single-objective optimization with multi-objective optimization, and (iii) to efficiently accomplish multi-objective parameter optimization for a complex and time-consuming hydrological model based on surrogate modeling. The remainder of this paper is organized as follows: Section 2 contains a brief description of the study area, the XAJ model, the surrogate modeling approach, and the evaluation criteria. Single objective and multi-objective optimization results are described in Section 3; and subsequently, Section 4 gives some conclusions of the study.

2 Materials and methods

2.1 The study area

The Yanduhe catchment (Fig. 1) of the Three Gorges is an upper tributary of the Yangtze River, with a drainage area of 601 km^2. The area, which is mountainous, with elevations ranging from 138 m above sea level at the watershed outlet to 3,031m in the upstream site, is covered by dense and thick vegetation. The vegetation cover accounts for approximately 80% of this area. There are five rain stations and one hydrological station in the catchment, as shown in Fig. 1. The mean annual temperature of this catchment is 11.5°C, the mean annual precipitation is almost 1,650 mm, and the mean annual runoff is 1,240 mm. Due to the dominance of a monsoon climate, more than 60% of annual rainfall occurs during the flood season (May to September). The DEM data on a grid scale of 30 m×30 m was used to extract the drainage network, and to derive topographic attributes of the watershed. The data set is provided by the International Scientific Data Service Platform, the Computer Network Information Center, and the Chinese Academy of Sciences (http://datamirror.csdb.cn). The catchment can be divided into five sub-catchments, in order to capture the spatial variations of precipitation and the underlying surface, based on the rain stations and the natural channel drainage network. The watershed characteristics for each sub-catchment are shown in Table 1 (Kong and Li, 2006). In this study, fifteen flood events from 1981 to 1987 were selected for analysis. The first 10 flood events were used to calibrate; and the other 5 flood events to validate.

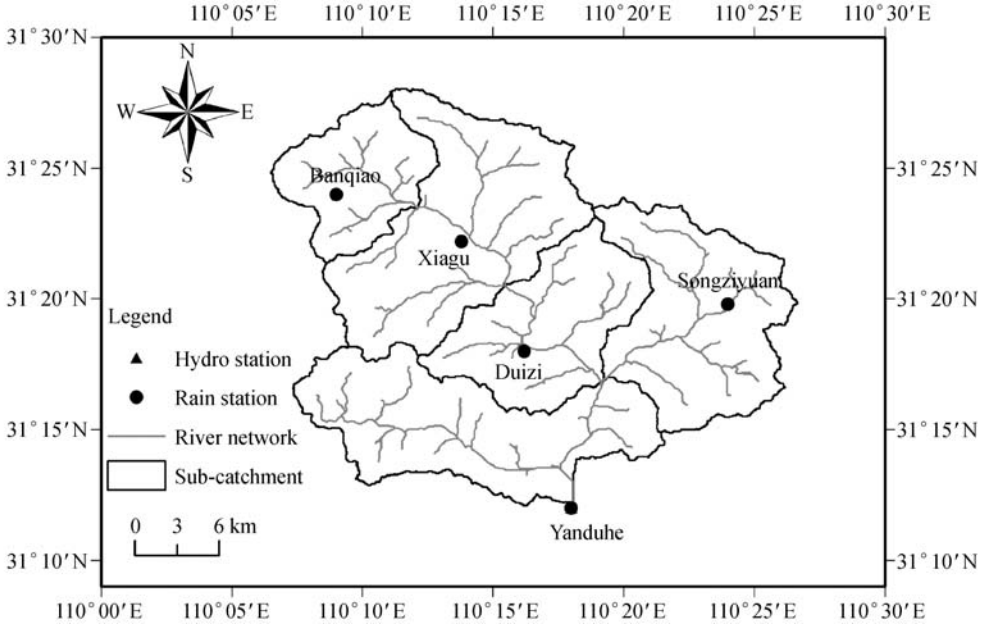


Fig. 1 Distribution of rain station, drainage network, and sub-catchments of Yanduhe catchment.

Table 1 Characteristics of each sub-catchment

Rain station	Latitude and longitude	Area/km^2	Weight of area	Number of reach	Slope/(°)	Distance to watershed outlet/km
Banqiao	110.15°E, 31.40°N	67.15	0.11	4	0.571	30.18
Xiagu	110.23°E, 31.37°N	163.25	0.28	2	0.683	19.84
Duizi	110.27°E, 31.30°N	98.6	0.16	2	0.711	7.88
Songziyuan	110.40°E, 31.33°N	129	0.21	3	0.603	7.88
Yanduhe	110.30°E, 31.20°N	143.24	0.24	0	0.576	0.96

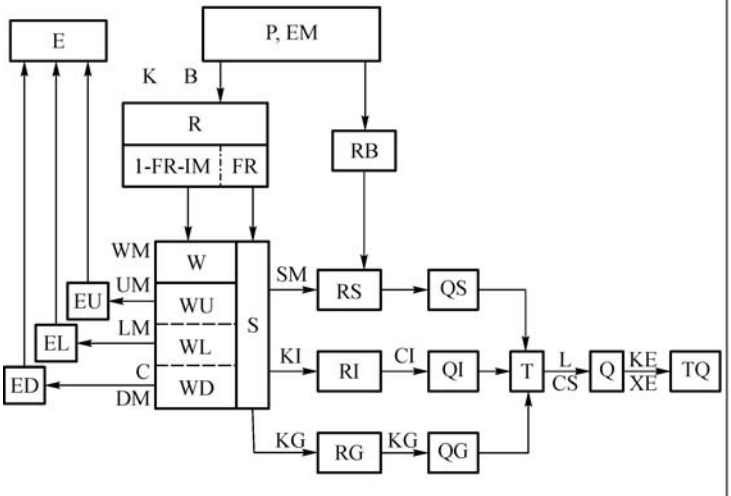
2.2 XAJ model

The XAJ model is a conceptual watershed model developed in the 1980s, which can be used for forecasting flood, runoff, and stream flow over a catchment (Qu et al., 2012). The XAJ model has been widely used in many catchments since it was proposed, and much subsequent experience supports its validity for humid and semi-humid regions. A detailed description of the XAJ model can see in the work of Zhao (1984, 1992) and Zhao and Liu (1995). The model structure, inputs, outputs, and state variables of the XAJ model are shown in Fig. 2. The ranges and physical meaning of the model parameters are summarized in Table 2 (Li et al., 2009; Lü et al., 2013).

2.3 Surrogate model and design of experiments

Surrogate modeling has been widely used for uncertainty analysis and parameter optimization (Song et al., 2012c, 2013). Surrogate modeling is a collection of statistical and mathematical techniques useful for developing, improving, and optimizing processes. In general, a surrogate model utilizes some form of simplified function to approximate the underlying system using points that have already been evaluated, and is considered to be a fast surrogate model compared to the exact evaluation model. The simplified function may be a) linked to the processes to be modeled, b) some simplified solutions to a complex problem, c) a mathematical model of the input and output causal relationship, and d) a surrogate generated due to inherency or proximity. In this work, the third type of simplified function was used to construct a surrogate model. Therefore, a response surface model (RSM) was selected to explore the statistical relationships between original model parameters and output response objectives. The main idea of RSM is to obtain an optimal response between parameters (input) and objective functions (output), via a sequence of experiments designed from the original simulation model.

Generally, the formulation of a surrogate model involves the following three steps: a) experimental design — a means to generate data points for evaluation; b) model choice — the selection of a surrogate function as an approximator; and c) model fitting — tuning of the



INPUT

P	Areal mean rainfall
EM	Pan evaporation

OUTPUT

E	Acctual evapotranspirations from the whole basin
EM	Evapotraspirations from the upper soil layer
EL	Evapotraspirations from the lower soil layer
ED	Evapotraspirations from the deeper soil layer
Q	Discharge from a sub-catchment

STATE VARIABLES

R	Runoff from the previous area
RB	Runoff from the imprevious area
RS	Surface runoff
RI	Interflow runoff
RG	Groundwater runoff
FR	The runoff producting area proportion
W	Areal mean tension water storage
WU	Areal mean tension water storage in the upper layer
WL	Areal mean tension water sorage in the lower layer
WD	Areal mean tension water sorage in the deeper layer
S	Areal mean free water storage
QS	Surface flow
QI	Interflow flow
QG	Groundwater flow
T	The total sub-basins inflow to channel network

Fig. 2 Flowchart for the XAJ model.

Table 2 Ranges and physical meaning of parameters for the XAJ model

Parameter Type	Symbol	Physical meaning	Range and units
Evapotranspiration	*K*	Ratio of potential evapotranspiration to pan evaporation	0.2–1.0
	UM	Averaged soil moisture storage capacity of the upper layer	20–30 (mm)
	LM	Averaged soil moisture storage capacity of the lower layer	60–90 (mm)
	DM	Averaged soil moisture storage capacity of the deeper layer	10–100 (mm)
	C	Coefficient of the deeper layer	0.05–0.2
Runoff generation	*WM*	Areal mean tension water capacity	100–150 (mm)
	B	Exponential of the distribution to tension water capacity	0.1–0.4
	IM	Percentage of impervious and saturated areas in the catchment	0.01–0.02
Runoff separation	*SM*	Areal mean free water capacity of the surface soil layer	5–50 (mm)
	EX	Exponent of the free water capacity curve	1–1.5
	KI	Outflow coefficients of the free water storage to interflow relationships	KI + KG = 0.7
	KG	Outflow coefficients of the free water storage to groundwater relationships	
Runoff routing	*CS*	Recession constant in the lag and route method	0.01–0.9
	CI	Recession constant of the lower interflow storage	0.01–0.95
	CG	Recession constant of the groundwater storage	0.9–0.999
	L	Lag in time empirical value	0–3 (h)
	KE	Muskingum routing method parameter	Δt
	XE	Muskingum routing method parameter	0–0.5

surrogate function to match the data points. First, most surrogate modeling enabled optimizers developed in the literature start with the formal Design of Experiments (DoE), in order to generate an initial set of design sites that is uniformly distributed in the explanatory variable space. A well-distributed initial set helps the surrogate models better represent the underlying (original) function (Razavi, 2013). However, as a priori knowledge about the underlying response surface is usually unavailable, DoE tends to assume uniformity of the distribution in the variable spaces. There are a wide variety of DoE methods available in the literature, such as full factorial design, fractional factorial design, central composite design, Latin hypercube sampling, LP-τ method, and so on. In our study, the LP-τ method is selected based on the PSUADE (Problem Solving environmental for Uncertainty Analysis and Design Exploration) platform to construct the surrogate model (Song et al., 2011), which was developed by the Lawrence Livermore National Laboratory (https://computation.llnl.gov/casc/uncertainty_quantification/).

The number of samples or design sites is another problem for constructing the surrogate model. The optimal size of an initial set of design sites is highly dependent on the shape and complexity of the original response surface, as well as the computational budget available. A small sample may not yield accurate statistical data, while a large sample will consume unnecessary computing resources (Zhan et al., 2013). The term "optimal" here reflects the fact that, for a given original response function, increasing the number of initial design sites would enhance the accuracy of fit. However, after some point, this enhancement would be at the expense of unnecessarily increasing the computational budget having to be initially spent on DoE, while it could have been spent more effectively in the next steps (Razavi, 2013). According to our previous work (Song et al., 2011, 2012a, 2012c, 2013), 4,000 samples based on the LP-τ method are suitable for surrogate modeling. The first 2,000 samples were used to construct the surrogate model; and the other 2,000 samples were selected to validate.

In this work, the multivariate adaptive regression splines (MARS) method (Friedman, 1991) was selected to construct a surrogate model of the XAJ model due to its advantages (Song et al., 2013, Zhan et al., 2013). MARS is a nonparametric regression procedure that makes no assumption about the underlying functional relationship between the dependent and independent variables. Instead, it constructs this relation from a set of coefficients and basis functions that are entirely "driven" from the regression data. In a sense, it is based on the "divide and conquer" strategy, which partitions the input space into regions, each with its own regression equation (Quirós et al., 2009). This makes it particularly suitable for problems with higher input dimensions, where the curse of dimensionality would likely create problems for other techniques[1)]. The resulting regression surface is piecewise linear and continuous. Basis functions are defined in pairs, using a knot or value of a variable that defines an inflection

1) http://documentation.statsoft.com/STATISTICAHelp.aspx?path = Gxx/MARSplines/MARSplinesIntroductoryOverview (2013-08-26)

point along the range of a predictor. The basis functions together with the model parameters are combined to produce the predictions given the inputs. The detailed information for MARS can see the work of Friedman (1991).

2.4 Multi-objective optimization problem

2.4.1 Pareto solutions

In a single-objective model calibration approach, model performance is expressed by a single-objective function that reflects a subjective choice, highlighting a specific aspect of the hydrograph (Vrugt et al., 2003a; de Vos and Rientjes, 2008). This objective function is then optimized to find what is regarded as the optimal model parameters. Multi-objective methods, on the other hand, reveal a set of Pareto solutions that represent the trade-off between the objectives involved. The benefit of this approach is that more information from the data is used in the evaluation of the model. If a model performs well on multiple objectives it implies performance consistency; and thus, the model is likely to be more reliable.

The considerations of multi-objective calibration imply the design of a calibration strategy that has the ability to incorporate several objective functions (Reichert et al., 2011) simultaneously. A strategy that can address this challenge is multi-objective optimization, and can be stated as follows:

$$\min F(\theta) = \min\{F_1(\theta), F_2(\theta), ..., F_m(\theta)\}, \theta \in \Theta, \quad (1)$$

where θ represents the model parameter set, $F_i(\theta)$ is the i^{th} of m objective functions, and Θ is feasible parameter space. The solution to this problem will, in general, no longer be a single "best" parameter set, but will consist of a set of Pareto solutions in the feasible parameter space Θ corresponding to various trade-offs among the objectives (Vrugt et al., 2003b; Tang et al., 2006; Guo et al., 2013a).

In this work, the surrogate modeling based multi-objective optimization method was used to estimate and optimize the parameters of the XAJ model. The flowchart of the integration method involves the following steps as shown in Fig. 3.

2.4.2 Multi-objective evaluation criteria

The selection of the length and information content of the experimental data series used in the model evaluation is a critical step, since it has strong influence on the model's performance. As a result, methods for identifying critical time periods in the observations, which contain most of the information for parameter identification, can provide useful guidance when selecting the evaluation data series (Singh and Bárdossy, 2012). For addressing model performance evaluation, the use of qualitative and quantitative criteria is recommended, because both approaches allow for capturing distinct aspects of model performance (e.g., Clark et al., 2011; Kavetski and Clark, 2011; McMillan et al., 2011; Biondi et al., 2012; Pushpalatha et al., 2012; Ritter and Muñoz-Carpena, 2013). Success in the use of computer models for simulating environmental variables and processes requires objective model calibration and verification procedures. Several methods for quantifying the goodness-of-fit of observations against model-calculated values have been proposed (Song et al., 2012a, b). In this work, a combination of four evaluation tools was used to calibrate the models' parameter values and assess the performance of their simulations.

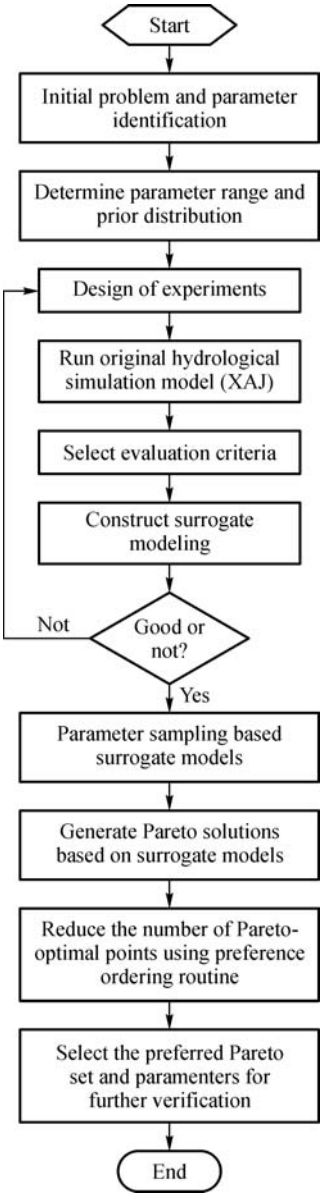


Fig. 3 Flowchart of the meta-model-based optimization for the XAJ model.

First, in order to evaluate model performance, a plot of observations against the calculated values illustrates the degree to which the points match the identity line (denoted as the 1 ∶ 1 line). This scatter plot allows for a visual inspection of model performance, such that the higher the agreement between calculated and observed values, the more the scatters tend to concentrate close to the 1 ∶ 1 line. Another interesting aspect of this graphical method is its ability to show 1) if model performance is homogeneous along the prediction range, or depends on the magnitude of the calculated values; 2) other unexpected relationships between the two data sets: non-linear or a linear correlation different to the identity line; 3) the presence of potential outliers; and 4) appreciable model bias (Ritter and Muñoz-Carpena, 2013).

Secondly, as a dimensionless goodness-of-fit indicator, the Nash-Sutcliffe efficiency (*NSE*) coefficient (Nash and Sutcliffe, 1970) was selected. *NSE* represents the complement to unity of the ratio between the mean square error of observed vs. predicted values and the variance of the observations. This coefficient takes values $-\infty \leqslant NSE \leqslant 1$. $NSE = 1$ indicates a perfect fit, while $NSE \leqslant 0$ suggests that the mean of the observed values is a better predictor than the evaluated model itself (Ritter and Muñoz-Carpena, 2013). This is calculated as follows:

$$NSE = 1 - \frac{\sum_{i=1}^{N} (O_i - S_i)^2}{\sum_{i=1}^{N} (O_i - \overline{O})^2}, \quad (2)$$

where O_i and S_i represent the sample (of size N) containing the observations and the model estimates, respectively; and $\overline{O}$ is the mean of the observed values.

Thirdly, for quantifying prediction error in terms of the units of the variable calculated by the model, relative error of peak flow (*REPF*), and runoff volume (*RERV*) were selected (Madsen, 2000; Song et al., 2012b). These indicators are frequently used and their definitions are given by

$$REPF = \frac{O_p - S_p}{O_p} \times 100\%, \quad (3)$$

$$RERV = \frac{V_{obs} - V_{sim}}{V_{obs}} \times 100\%, \quad (4)$$

where O_p and S_p represent the observed peak flow and the model estimates, while V_{obs} and V_{sim} mean the observed and simulated total runoff volume, respectively. These indicators range from 0 to ∞, where the value 0 indicates a perfect fit.

2.4.3 Performance of the multi-objective calibration problem

Section 2.4.1 reveals that multi-objective optimization problems can be stated as a minimum or maximum optimal problem for many evaluation criteria. However, the above three objective functions (*NSE*, *REPF*, and *RERV*) could not construct a minimum or maximum problem directly since these functions are not all good for smaller and smaller, or vice versa. As a result, the multi-objective optimization problem can be stated as follows:

$$\min F(\theta) = \min\{F_1(\theta), F_2(\theta), F_3(\theta)\}, \theta \in \Theta, \quad (5)$$

where $F_1(\theta) = 1 - NSE, \theta \in \Theta$, $F_2(\theta) = |REPF|, \theta \in \Theta$ and $F_3(\theta) = |RERV|, \theta \in \Theta$.

The solution of Eq. (5) will not, in general, be a single unique set of parameters but will consist of the so-called Pareto set of solutions, according to various trade-offs between the different objectives. A member of the Pareto set will be better than any other member with respect to some of the objectives, but because of the trade-off between the different objectives, it will not be better with respect to other objectives. When solving a multi-objective calibration problem, the problem is usually transformed into a single-objective optimization problem by defining a scalar that aggregates the various objective functions. For example, some previous studies turned the multi-objective optimization problem into a single-objective optimization problem by taking the linear weighted sum of the multiple objectives (Bennett et al., 2004; Huang et al., 2013). In practical applications, the entire Pareto set may be too expensive to calculate, and one is only interested in part of the Pareto optimal solutions. In this case, it is proposed that an aggregated objective function that puts equal weights on the different objectives is used. The aggregated objective function is the Euclidean distance (Madsen, 2000)

$$F_{agg}(\theta) = \left[\left(F_1(\theta) + A_1 \right)^2 + \left(F_2(\theta) + A_2 \right)^2 + \left(F_3(\theta) + A_3 \right)^2 \right]^{1/2}, \quad (6)$$

where A_i are transformation constants assigned to the different objectives. A balance measure can be defined by assigning transformation constants in Eq. (6) such that all $(F_i + A_i)$ have about the same distance to the origin. The minimum values of F_i ($F_{i,\min}$) are estimated from this initial population, and each of the objective functions is transformed to having the same distance to the origin as the objective function with the largest minimum value of F_i, i.e.,

$$A_i = \max\{F_{j,\min}\} - F_{i,\min} \quad i,j = 1,2,3. \quad (7)$$

3 Results and discussion

3.1 Parameter identification and selection

As stated above, the curse of dimensionality is increasingly hard to prevent due to the development and application of

distributed hydrological models. Their model structures are much more complex than that of empirical models, with many sub-processes considered. Wheater (2002) noted that "a simple model structure does not reflect the complexity of the rainfall-runoff response; and a complex model structure is not always supported by the available data. A balance between the complexity of the model and available information is crucial for successful model identification". Model complexity can be reduced by an appropriate degree through identification statistics (Dunn et al., 2008) or sensitivity analysis (Fenicia et al., 2008; van Werkhoven et al., 2009), and by holding insensitive parameters constant or formally re-structuring the model (McIntyre and Al-Qurashi, 2009). Therefore, the first step is to select an appropriate number of parameters for optimization (e.g., sensitivity analysis to reduce the dimensionality of model parameters).

A sensitivity analysis of model parameters was achieved in former work (Song et al., 2013). According to the results of the sensitivity analysis, ten parameters (e.g., *K*, *WM*, *SM*, *KI*, *KG*, *CS*, *CI*, *CG*, *KE*, *XE*) of the XAJ model are important for hour-scale simulation processes (Song et al., 2013). Actually, some sensitive parameters have clear physical meanings, such as Muskingum method parameters *KE* and *XE*, which can be initially determined by hydraulic formulas. *KE* may be approximately the travel time over a sub-reach; and *XE* may be determined by the special length and the length of the sub-reach. However, these parameters should also be estimated through manual or automatic calibration due to lack of available data and information. In addition, *WM* is usually considered as non-sensitive for output response (Li et al., 2009), whereas it is also important for the objective of water balance from our previous work (Song et al., 2013). By the degeneration principles mentioned before, taking a fixed value (e.g., 0.7) for the sum of *KI* and *KG*, optimizes the two parameters by a two-dimensional joint search. That is, one of the two parameters can be determined once the other parameter is estimated. In summary, nine parameters were selected for analysis in this study. The other non-sensitive parameters can be given fixed values, in accordance with prior knowledge and practice, for example, $UM = 20$ mm, $LM = 70$ mm, $B = 0.3$, $IM = 0.01$, $C = 0.15$, $EX = 1.5$, $L = 1$ hr.

3.2 Establishment and validation of the surrogate model

In the MARS method, the number of basis functions is 1,000 and the number of knots is 500. In this study, the two parameters *K* and *SM* were selected as an example to construct the response surface models with different objectives, as shown in Fig. 4. The *x*-axis is parameter *K*, the *y*-axis is the parameter *SM*, and the *z*-axis shows the objective functions. From the response surface relationship between two parameters and the three different objective functions in Fig. 4, we can see that there are different response surfaces and optimal sites of parameters for different objectives. In a sense, it also indicates that the model calibration for single-objective functions is often inadequate to measure all of the characteristics of the model.

Another vital and indispensable step is to validate and test the response surface models. Two approaches were applied to check the responses for goodness-of-fit by examining the interpolation error and prediction error. The first method is cross-validation (*k*-fold cross-validation method) (McLachlan et al., 2004), the other is an experimental test. The statistical histograms of interpolation errors based on the *k*-fold cross-validation method ($k = 500$) for all the three objectives are shown in Fig. 5. The results demonstrate that all of the responses give acceptable interpolation errors, i.e., the sample is adequate to represent the input-output relationship. In addition, Fig. 6 depicts scatter plots between the surrogate-models' simulated objective function values and the corresponding original XAJ model generated values. Note that the objective function values for these 4,000 samples for validation and test correspond to the values from the surrogate modeling. The data points do not deviate greatly from the 1 ∶ 1 line for the four objective functions, which is evident from the high R^2 values. The plots exhibit a very intense scatter, implying that the surrogate models mimic the performance of the original XAJ model fairly well.

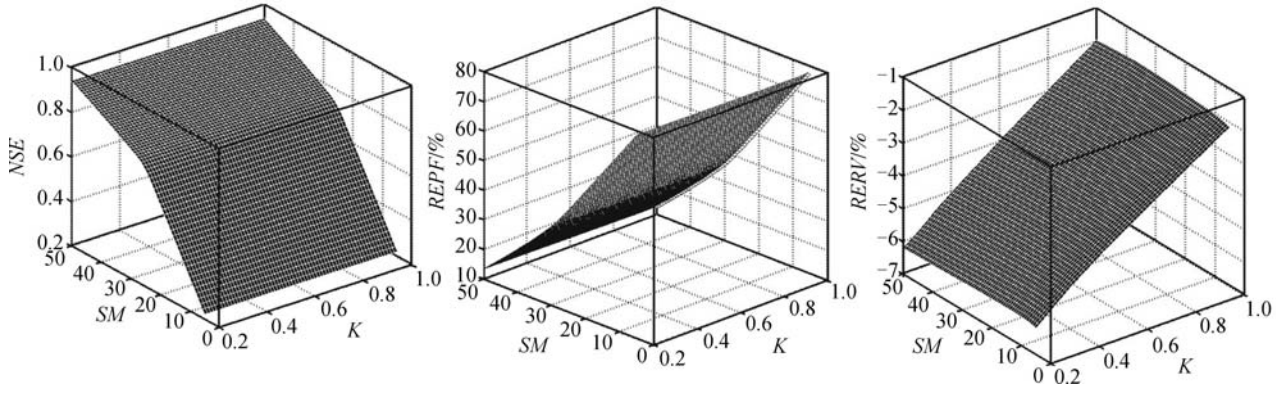


Fig. 4 Response surface plots for the three objective responses with parameters *K* and *SM*.

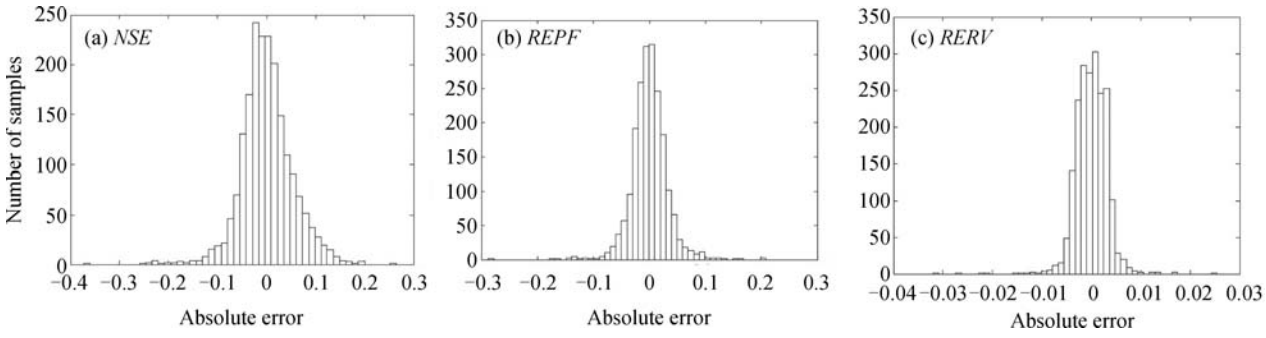


Fig. 5 Cross-validation for the surrogate modeling.

Once the surrogate models are deemed satisfactory, parameter optimization can rely on these response surface models, which are inexpensive to evaluate.

3.3 Single-objective optimization results

This section intends to estimate the efficiency of parameter calibration with a single-objective function. When the response surface models are acceptable and reasonable as a surrogate, the ten parameters of the XAJ model are optimized by searching the optimization values around the response surface models. As mentioned before, the parameters are optimized by the surrogate modeling using *NSE*, *REPF*, or *RERV* as the single-objective function. A population size of 100,000 from the three response objectives is used to estimate the optimal solution sets with respect to the 10 parameters, respectively. Table 3 gives the parameter values based on a single-objective optimal scheme for the three objectives. To illustrate a comparison between the optimal simulated discharge and the observed discharge, Figure 7 provides scatter plots for the optimization results in calibration and validation. For the calibration process, single-objective parameter optimization qualifies relative to the aspect of corresponding objective functions. However, it is not simultaneously perfect for all three of the objective functions in the calibration and validation periods. In other words, the models show large discrepancies in the values of one criterion when the other one is used as the objective function or evaluation criterion. For example, the *NSE* value drops below 0.8 when using *REPF* as an objective, while they are largely acceptable (above 0.9) when using *NSE* as the objective function. Unfortunately, the *REPF* value in the validation based on the *REPF* optimal parameters is –59.5%, which is far below the limit value of acceptability for model prediction (±20%). Furthermore, the simulation in the validation based on the *RERV* optimal parameters is also very poor, with an *REPF* value of –85.65%.

The reasons include the selection of the objective function and the problem of split sample. In this work, the first 10 flood events were taken as a seriate flood event to calibrate model parameters, which may be a key reason for the poor performance of the flood peak simulation. The results, as shown in Fig. 7, also reveal that single-objective functions, no matter how carefully chosen, are often inadequate to properly measure all of the characteristics of the observed data. In the validation period, most of the criteria are degraded; but the calibration using *NSE* as the objective function appears to be more robust, since the degradation on *NSE* from the calibration to validation mode is less pronounced than when using *REPF* and *RERV* as objective functions. To conclude, it is shown that the criteria based on *REPF* or *RERV* cannot be used as a single-objective function, since it may produce large discrepancies in water balance simulations, or erroneous estimation of peak discharges. However, these two criteria are usually considered as hydrological metrics for model calibration and application, because these indicators are suitable for describing rainfall-runoff response at a basin scale. *NSE*, as a statistical metric, focuses on the hydrograph (i.e., errors and trends). Therefore, it is evident that all of these metrics should be considered in multi-objective optimization.

3.4 Multi-objective calibration and uncertainty assessment

The defining limit of acceptability for model predictions depends on the model applications (Beven, 2006). Although subjective, we proposed four model performance classes based on the three objectives for model calibration in this work, which are denoted as Unsatisfactory, Acceptable, Good, and Very good (as shown in Table 4). In accordance with the accuracy standard (Standard for Hydrologic Information and Hydrological Forecasting, GB/T 22482-2008) established by the Ministry of Water Resources of the People's Republic of China, if the absolute value of *RERV* and *REPF* is less than 20%, the simulation result qualifies relative to peak flow and total runoff volume. In addition, the criteria for our goodness-of-fit evaluation was proposed according to various recommendations that can be found in the literature (Moriasi et al., 2007; Biondi et al., 2012; Pushpalatha et al., 2012; Ritter and Muñoz-Carpena, 2013).

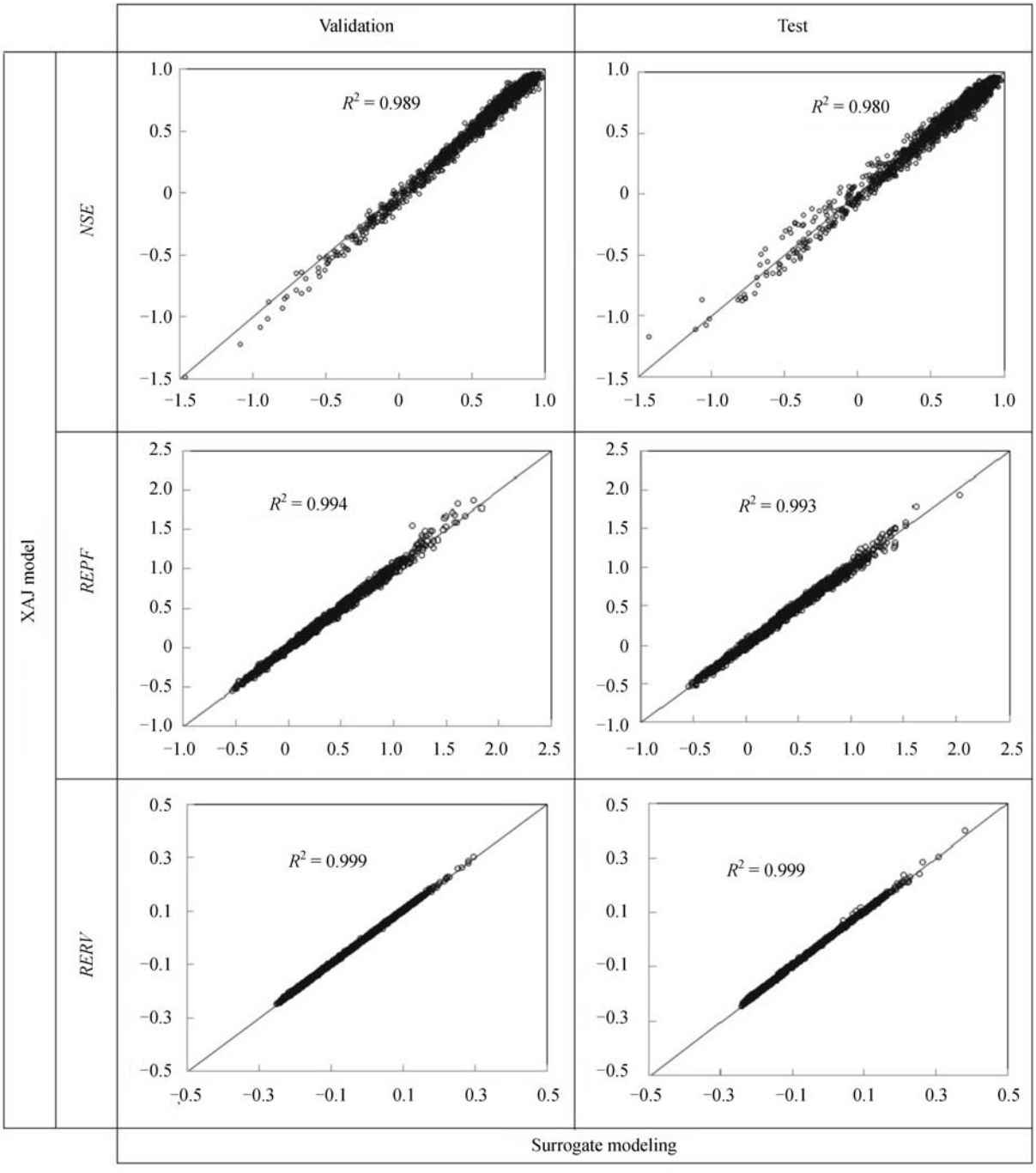


Fig. 6 Validation and test for the surrogate modeling.

Table 3 Optimization results for ten parameters

Scheme	*K*	*WM*	*SM*	*KI*(*KG*)	*CS*	*CI*	*CG*	*KE*	*XE*
NSE (optimum)	0.52	132	45	0.17(0.53)	0.79	0.63	0.97	0.79	0.17
REPF (optimum)	0.37	149	33	0.40(0.30)	0.38	0.79	0.99	1.40	0.18
RERV (optimum)	0.75	131	40	0.52(0.18)	0.24	0.88	0.93	0.65	0.12

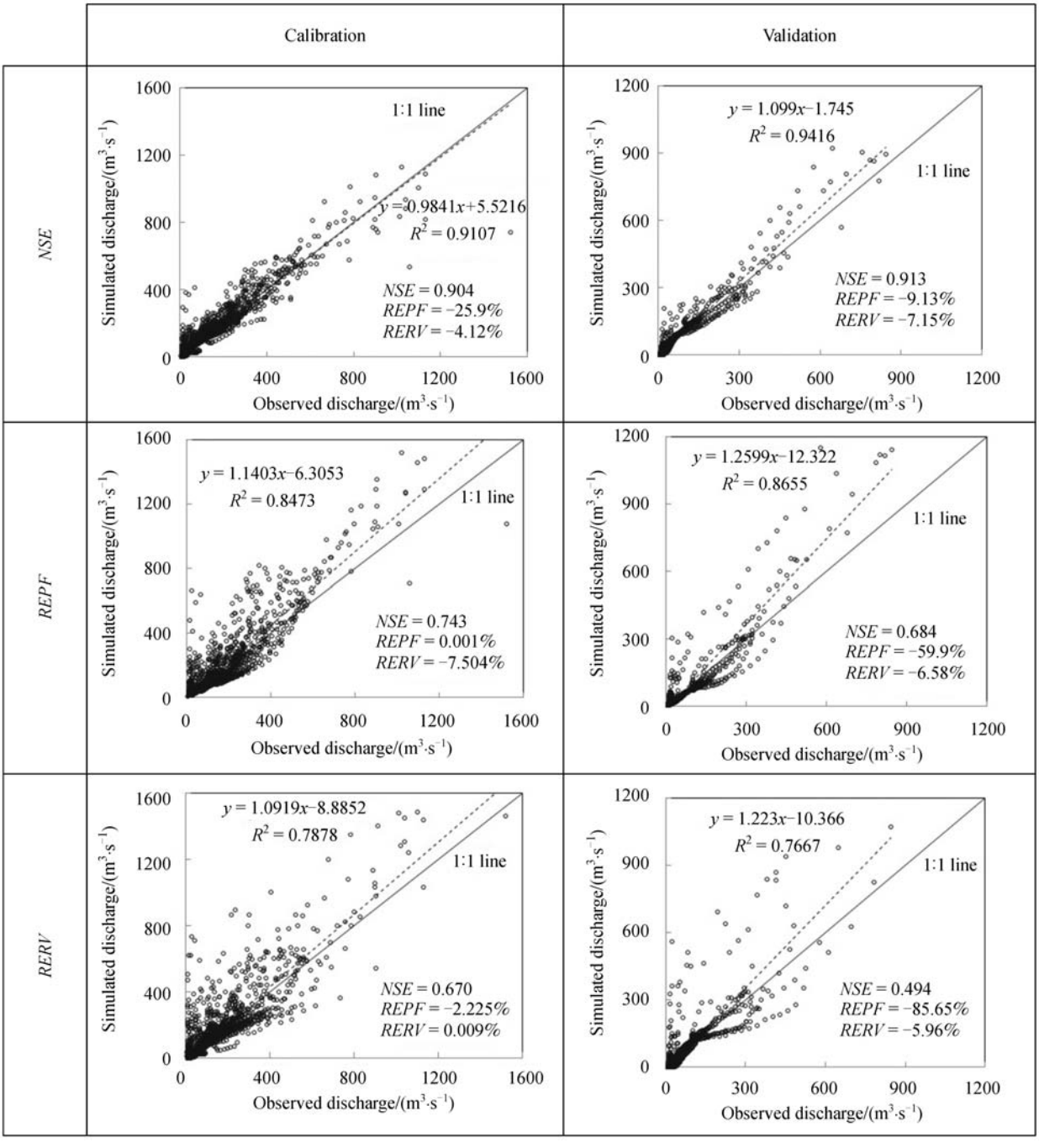


Fig. 7 Comparison of single-objective optimal value using surrogate modeling.

Table 4 Criteria for the goodness-of-fit evaluation

Performance rating	*NSE*	Absolute value of *REPF*	Absolute value of *RERV*
Very good	⩾0.90	⩽5%	⩽5%
Good	0.80–0.90	5%–10%	5%–10%
Acceptable	0.70–0.80	10%–20%	10%–20%
Unsatisfactory	< 0.70	> 20%	> 20%

Because any conceptual watershed model will, in general, be unable to match all of the different aspects of the watershed's behavior observed in the measured hydrograph, all three transition objective functions (F_1, F_2, F_3) are used to calibrate the XAJ model using surrogate modeling. Generally, if the performance rating is "good" or "very good" for a single-objective function, the optimal scheme is considered a perfect performance associated with the corresponding objective function. Besides, when the classes for the three objective functions are good, i.e., they all meet the performance thresholds as shown in Table 4, the optimal scheme could be considered to be a good solution; and it can be defined as an efficient solution for analyzing the Pareto front. A number of samples and tests were carried out in order to estimate the Pareto front and to analyze the trade-off solutions. Because in a multi-objective problem, several objectives are to be considered simultaneously, the concept of inferiority-superiority in multi-objective theory (Yapo et al., 1998) was used to rank the population and to obtain the Pareto optimal solution.

In this work, considering the accuracy standard (GB/T 22482-2008) and the concept of Pareto solutions, the maximum 100,000 samples were based on surrogate modeling and were used to estimate the Pareto sets. The single-objective optimizations provide the tails of the Pareto front, and the optimization based on the balanced aggregated measure approximate the balanced central part of the Pareto front (Madsen, 2000). The estimated Pareto front for the calibration of the three-objective problem is shown in Fig. 8. In this case the compromise solution corresponds to a break point on the Pareto front; that is, moving along the front in either direction implies only a small decrease of one of the objective functions, at the expense of a pronounced increase of the other objective function. In addition, many studies have demonstrated that the Pareto front for a three-objective problem should be a two-dimensional surface (e.g., Goel et al., 2007; Cao et al., 2011; Trautmann et al., 2013). Previous work presented by Guo et al. (2013b) demonstrated that the Pareto front for three-objective functions based on MSLE, M4E, and MSDE for the Xin'anjiang model is approximately a curve, the same as in Fig. 8.

In fact, this Pareto front is not a one-dimensional curve because the front is not consequent. That is, it is a two-dimensional surface. The variation of the optimum model parameter sets along the Pareto front is shown in Fig. 9. It presents normalized parameter plots for each of the sensitive parameters of the XAJ model using the surrogate modeling based on the 100,000 trials. The parameter values are normalized with respect to the upper and lower limits given in Table 2, so that the feasible range of all parameters is between 0 and 1. The ten XAJ model parameters are list along the x axis, while the y axis corresponds to the parameter values scaled according to their prior uncertainty ranges to yield normalized ranges. A remarkably large variability for some sensitive parameters

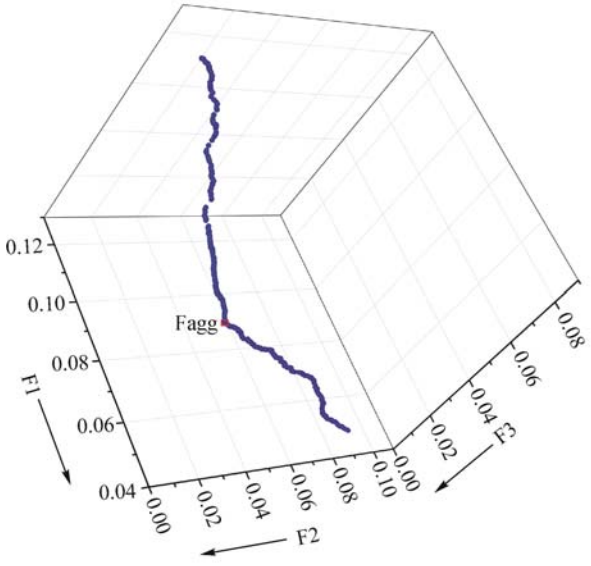


Fig. 8 Illustration of estimated Pareto front for the optimization of a three-objective function attained across surrogate modeling. The black arrows indicate direction of increasing preference. Marked optimal point corresponds to using a balanced aggregated objective function.

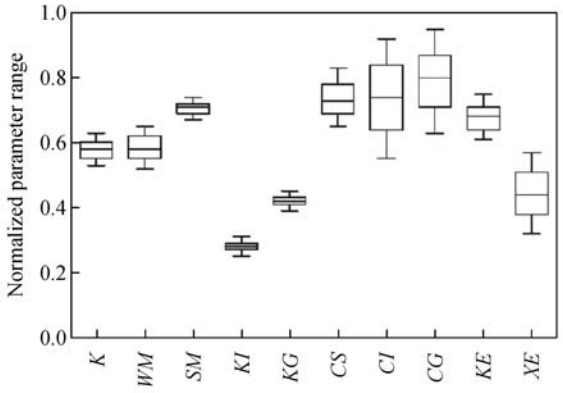


Fig. 9 Normalized range of parameter values along the Pareto front shown in Fig. 8.

(e.g., *CS*, *CI,* and *CG*) is observed in the parameter values when moving along the Pareto front. The variation of parameter combinations along the Pareto front also implies a large variability on the simulated hydrograph, as shown in Fig. 10.

The traditional concept of model calibration based on a single objective is built on the hypothesis that a unique optimum set of parameter values exists. It should be clear from the analysis above that such a unique global solution does not exist. In a multi-objective context, there is a multitude of parameter combinations that are "equally good". Therefore, a calibrated model should produce a range of simulated hydrographs corresponding to the

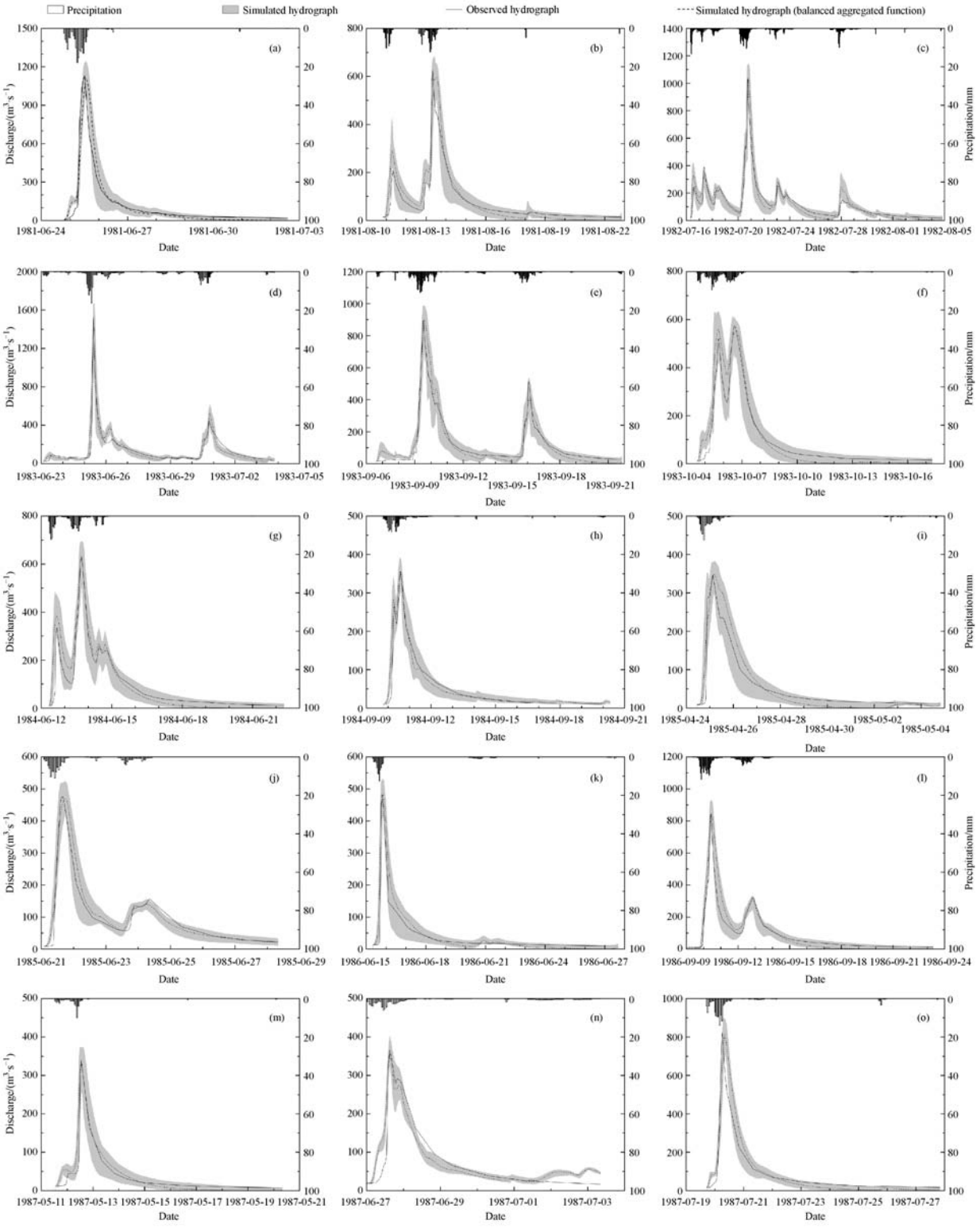


Fig. 10 Hydrograph prediction uncertainty ranges associated with the Pareto solution set estimated from the surrogate modeling: (a)–(j) for the 10 flood events in the calibration, and (k)–(o) for the 5 flood events in the validation. Black solid lines indicate the observed hydrograph for 15 flood events. Short dash dot lines indicate the simulated hydrograph based on the optimization of balanced aggregated function. Black columns indicate the precipitation. Shaded area indicates the simulated hydrograph from Pareto front for three-objective function.

"equally good" parameter sets along the Pareto front. The hydrograph uncertainty ranges associated with these good solutions for ten sensitive parameters of the XAJ model in the calibration and validation periods are displayed in Fig. 10. The observed discharges are indicated with solid areas; while the good solutions for the three objective functions are indicated with shaded areas.

However, another task is much more challenging, that of finding a unique optimal solution from the Pareto front for model application. This is because the computational complexity is increased by generating multiple solutions, and also because of the difficulty associated with maintaining a desirable distribution of the alternative solutions (Huang et al., 2013). To overcome those difficulties, Bennett et al. (2004) and Huang et al. (2013) turned the multi-objective optimization problem into a single-objective optimization problem by taking the linear weighted sum of the multiple objectives. In this work, the balanced aggregated function method (Madsen, 2000) was used to find a "best" unique solution from the Pareto front, as shown in Fig. 10. The results presented in Fig. 10 emphasize that the solution based on the balanced aggregated function is good for the goodness-of-fit evaluation criteria, as shown in Fig. 11.

3.5 Efficiency gains of surrogate-enabled analyses

The most important question in assessing a surrogate-enabled analysis is how efficient or effective it is compared with other efficient alternative tools without surrogate modeling; because the computational efficiency achieved is the main factor motivating the research and application of surrogate modeling. Surrogate-enabled analyses typically sacrifice accuracy for efficiency; as they utilize approximate models to more efficiently achieve the analysis objectives. Broad et al. (2005) pointed out that surrogate models should only be used where time constraints prohibit the possibility of optimizing a problem with a simulation model. To be clear, a benchmark alternative analysis should be available for at least the same number of times of original model simulations as utilized in the surrogate-enabled analysis (Song et al., 2012d; Razavi, 2013).

In an optimization context, the most tangible measure of efficiency gains over a benchmark alternative is computational saving. Generally, the computational saving time can be obtained from the difference value between the computational budget, or time based on an original model, and surrogate modeling with the same quality. Computational budgets may be quantified as the total CPU clock time (e.g., Broad et al., 2005; Behzadian et al., 2009; Kourakos and Mantoglou, 2009), or the number of original function evaluations (Mugunthan and Shoemaker, 2006; Zou et al., 2007, 2009).

In any surrogate-enabled analysis, the available computational budget or time is divided between three main parts: 1) budget or time required to run the original model; 2) budget or time required to develop, run, and update the surrogate model; and 3) budget or time the analyst needs to identify and create an appropriate surrogate-enabled analysis framework. Parts 2 and 3, which are referred to as "surrogate-modeling time" and "analyst time", respectively, should consume a small portion of the available computational budget, leaving the majority for part 1 (Razavi, 2013). However, analyst time is hardly quantified by total CPU clock time. As a result, in the majority of the water resources related surrogate modeling studies, we recognize that the analyst time based on surrogate-model optimizers is equivalent to the analyst time using an optimizer without surrogate models. Therefore, in our work, part 1 and part 2 can be considered to validate the efficiency of the proposed surrogate-enabled analyses.

First, the computational time required to run the original model for constructing the surrogate model should be considered. In this work, it took approximately 28 hours for the 4,000 runs of the original model. That is, a single

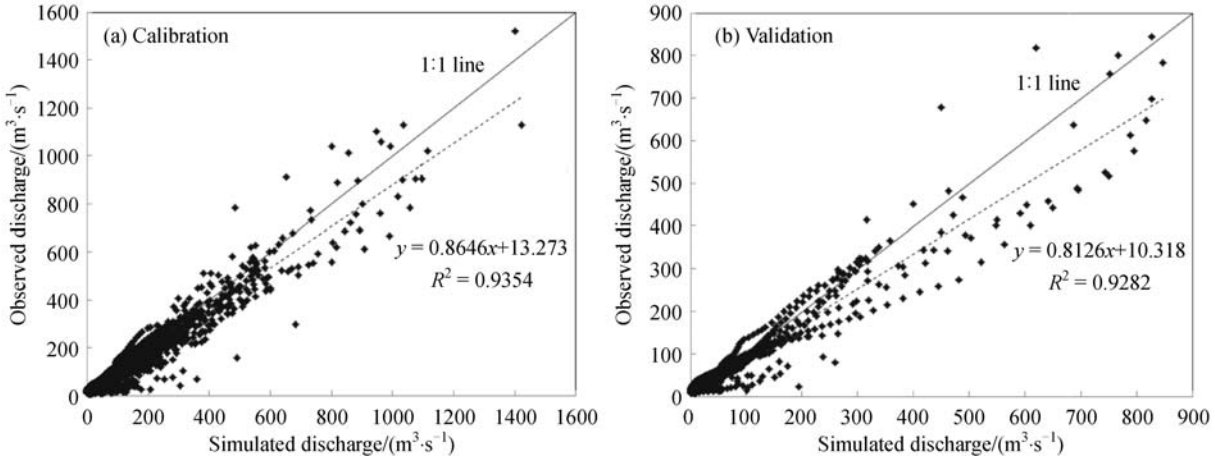


Fig. 11 Scatter plots of observed and simulated discharge with optimization based on balanced aggregated function.

evaluation of the original model takes about 25 seconds on average to execute on a 3.2 GHz Intel Core i5-3470 processor with 4 GB of RAM, and running the Windows 7 operating system. Based on this, if the number of runs is 100,000, it may be about 694 hours (approximately equal to 29 days) on the same computer, which might be practically infeasible. In addition, as mentioned above, constructing and validating the surrogate models also incurs computational cost. In our work, the surrogate modeling time including the development time, validation time, and analysis time, was about 8 hours for one objective response (e.g., 20–30 minutes for constructing the model, 420–450 minutes for cross-validation of the surrogate model, and 10–20 minutes for analyzing time). That is to say, for the three objective responses, it takes approximate 24 hours for model optimization by surrogate modeling. The MARS model can be developed based on PSUADE software with the running environment: Centos 5.5, Inter Pentium IV CPU, 1.80 GHz, 256 MB RAM. In summary, the total time of runs to optimize by surrogate modeling is reduced from the original 694 hours to 52 hours. Furthermore, the results confirmed that surrogate modeling is an effective and efficient approach when the original model is relatively complex and the computational cost is very high (e.g., larger than several hours or days). It can be seen that there is a significant advantage to using surrogate models for hydrological process simulation, design, and calibration.

4 Conclusions

Parameter optimization is a necessary and vital process within the application of conceptual or distributed hydrological models (Song et al., 2012c). However, the practice of model calibration in hydrological models also reveals that single objective functions are often inadequate to properly measure all of the characteristics of the observed values. Consequently, single-objective calibration methods do not usually provide parameter estimates that are considered to be acceptable by practicing hydrologists. As a result, more and more multi-objective optimization methods have been considered to estimate parameters for hydrological models or environmental models (Guo et al., 2013a), such as MOSCEM, NSGA-II, etc. Unfortunately, these methods exhibit some disadvantages, such as being time-consuming, having enormous computation cost, or being relatively difficult to implement in practice.

In this study, the surrogate modeling method based on multiple objective functions was used to calibrate the XAJ model. Surrogate-enabled analyses usually sacrifice accuracy for efficiency, as they utilize approximate models to achieve analysis objectives more efficiently, which is the main factor motivating the research and application of surrogate modeling, especially for complex computer models. Additionally, some criteria cannot be used or recommended as a single-objective function to calibrate model parameters for hydrological models, such as *REPF* and *RERV*. This result is understandable, since criteria based on *REPF* do not meet several key requirements for a discharge simulation, e.g., bias in discharge values (except peak flow).

In summary, the multi-objective calibration of hydrological models is a complex and non-linear optimization problem. The example presented in this paper indicates that the Pareto solutions of multi-objective functions can provide valuable information in model calibration and application, compared with parameter optimization based on single-objective functions. Clearly, it is difficult and precarious to generalize beyond the results presented here; and more research is needed on this aspect, especially for complex hydrological or environmental models with a high-dimension parameter space.

Acknowledgements The work was supported by the National Basic Research Program of China (No. 2010CB951103), the National Natural Science Foundation of China (Grant Nos. 41330854, 41371063 and 51309155) and the National Science & Technology Pillar Program during the 12th Five-year Plan Period (2012BAC21B01 and 2012BAC19B03). We are also thankful to anonymous reviewers and editors for their helpful comments and suggestions.

References

Akhtar M, Ahmad N, Booij M J (2008). The impact of climate change on the water resources of Hindukush-Karakorum-Himalaya region under different glacier coverage scenarios. J Hydrol (Amst), 355(1–4): 148–163

Behzadian K, Kapelan Z, Savic D, Ardeshir A (2009). Stochastic sampling design using a multi-objective genetic algorithm and adaptive neural networks. Environ Model Softw, 24(4): 530–541

Bennett D A, Xiao N C, Armstrong M P (2004). Exploring the geographic consequences of public policies using evolutionary algorithms. Ann Assoc Am Geogr, 94(4): 827–847

Beven K, Freer J (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology. J Hydrol (Amst), 249(1–4): 11–29

Beven K J (2006). A manifesto for the equifinality thesis. J Hydrol, 320: 18–36

Bingeman A, Kouwen N, Soulis E D (2006). Validation of hydrological processes in a hydrological model. J Hydrol Eng, 11(5): 451–463

Biondi D, Freni G, Iacobellis V, Mascaro G, Montanari A (2012). Validation of hydrological models: conceptual basis, methodological approaches and a proposal for a code of practice. Phys Chem Earth A/B/C, 42–44: 70–76

Bliznyuk N, Ruppert D, Shoemaker C, Regis R, Wild S, Mugunthan P (2008). Bayesian calibration and uncertainty analysis for computationally expensive models using optimization and radial basis function approximation. J Comput Graph Statist, 17(2): 270–294

Broad D R, Dandy G C, Maier H R (2005). Water distribution system

optimization using metamodels. J Water Resour Plan Manage, 131 (3): 172–180

Broad D R, Maier H R, Dandy G C (2010). Optimal operation of complex water distribution systems using metamodels. J Water Resour Plan Manage, 136(4): 433–443

Cao K, Batty M, Huang B, Liu Y, Yu L, Chen J F (2011). Spatial multi-objective land use optimization: extensions to the non-dominated sorting genetic algorithm-II. Int J Geogr Inf Sci, 25(12): 1949–1969

Clark M P, McMillan H K, Collins D B G, Kavetski D, Woods R A (2011). Hydrological field data from a modeller's perspective. Part 2: process-based evaluation of model hypotheses. Hydrol Processes, 25 (4): 523–543

de Vos N J, Rientjes T H M (2008). Multiobjective training of artificial neural networks for rainfall-runoff modeling. Water Resour Res, 44 (8): W08434, doi: 10.1029/2007WR006734

di Pierro F, Khu S T, Savic D, Berardi L (2009). Efficient multi-objective optimal design of water distribution networks on a budget of simulations using hybrid algorithms. Environ Model Softw, 24(2): 202–213

Duan Q Y, Sorooshian S, Gupta V (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resour Res, 28(4): 1015–1031

Dumedah G (2012). Formulation of the evolutionary-based data assimilation, and its implementation in hydrological forecasting. Water Resour Manage, 26(13): 3853–3870

Dumedah G, Berg A A, Wineberg M, Collier R (2010). Selecting model parameter sets from a trade-off surface generated from the non-dominated sorting genetic algorithm-II. Water Resour Manage, 24 (15): 4469–4489

Dunn S M, Freer J, Weiler M, Kirkby M J, Seibert J, Quinn P F, Lischeid G, Tetzlaff D, Soulsby C (2008). Conceptualization in catchment modeling: simply learning? Hydrol Processes, 22(13): 2389–2393

Fenicia F, Savenije H H G, Matgen P, Pfister L (2008). Understanding catchment behavior through stepwise model concept improvement. Water Resour Res, 44(1): W01402, doi: 10.1029/2006WR005563

Friedman J H (1991). Multivariate adaptive regression splines. Ann Stat, 19(1): 1–67

Gan T Y, Biftu G F (1996). Automatic calibration of conceptual rainfall-runoff models: optimization algorithm, catchment conditions, and model structure. Water Resour Res, 32(12): 3513–3524

Goel T, Vaidyanathan R, Haftka R T, Shyy W, Queipo N V, Tucker K (2007). Response surface approximation of Pareto optimal front in multi-objective optimization. Comput Methods Appl Mech Eng, 196 (4–6): 879–893

Guo J, Zhou J, Zou Q, Song L, Zhang Y (2013b). Study on multi-objective parameter optimization of Xin'anjiang model. Journal of China Hydrology, 33(1): 1–7 (in Chinese)

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

Gutmann H M (2001). A radial basis function method for global optimization. J Glob Optim, 19(3): 201–227

Huang K, Liu X, Li X, Liang J, He S (2013). An improved artificial immune system for seeking the Pareto front of land-use allocation problem in large areas. Int J Geogr Inf Sci, 27(5): 922–946

Jiang Y, Li X Y, Huang C C (2013). Automatic calibration a hydrological model using a master-slave swarms shuffling evolution algorithm based on self-adaptive particle swarm optimization. Expert Syst Appl, 40(2): 752–757

Jones D R, Schonlau M, Welch W (1998). Efficient global optimization of expensive black-box functions. J Glob Optim, 13(4): 455–492

Jones J P, Sudicky E A, McLaren R G (2008). Application of a fully-integrated surface-subsurface flow model at the watershed-scale: a case study. Water Resour Res, 44(3): W03407, doi: 10.1029/2006WR005603

Kavetski D, Clark M P (2011). Numerical troubles in conceptual hydrology: approximations, absurdities, and impact on hypothesis-testing. Hydrol Processes, 25(4): 661–670

Kavetski D, Kuczera G, Franks S W (2006). Bayesian analysis of input uncertainty in hydrological modeling: 1. theory. Water Resour Res, 42: W03407, doi: 10.1029/2005WR004368

Khu S T, Madsen H (2005). Multiobjective calibration with Pareto preference ordering: an application to rainfall-runoff model calibration. Water Resour Res, 41(3): W03004, doi: 10.1029/2004WR003041

Khu S T, Werner M G F (2003). Reduction of Monte-Carlo simulation runs for uncertainty estimation in hydrological modeling. Hydrol Earth Syst Sci, 7(5): 680–692

Kong F Z, Li L L (2006). Application of digital elevation model in Xinanjiang model. Journal of China University of Mining & Technology, 35(3): 393–396 (in Chinese)

Kourakos G, Mantoglou A (2009). Pumping optimization of coastal aquifers based on evolutionary algorithms and surrogate modular neural network models. Adv Water Resour, 32(4): 507–521

Lee K T, Hung W C, Meng C C (2008). Deterministic insight into ANN model performance for storm runoff simulation. Water Resour Manage, 22(1): 67–82

Li H, Zhang Y, Chiew F H S, Xu S (2009). Predicting runoff in ungauged catchments by using Xinanjiang model with MODIS leaf area index. J Hydrol (Amst), 370(1–4): 155–162

Lindström G, Johannson B, Persson M, Gardelin M, Bergström S (1997). Development and test of the distributed HBV-96 hydrological model. J Hydrol (Amst), 201(1–4): 272–288

Liu D, Yu Z, Hao Z, Yang C, Ju Q (2007). Groundwater simulation in the Yangtze River basin with a coupled climate-hydrologic model. Journal of China University of Geosciences, 18(Special issue): 155–157

Lü H S, Hou T, Horton R, Zhu Y, Chen X, Jia Y, Wang W, Fu X (2013). The streamflow estimation using the Xinanjiang rainfall runoff model and dual state-parameter estimation method. J Hydrol (Amst), 480: 102–114

Madsen H (2000). Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. J Hydrol (Amst), 235(3–4): 276–288

McCabe M F, Franks S W, Kalma J D (2005). Calibration of a land surface model using multiple data sets. J Hydrol (Amst), 302(1–4): 209–222

McIntyre N, Al-Qurashi A (2009). Performance of ten rainfall-runoff models applied to an arid catchment in Oman. Environ Model Softw, 24(6): 726–738

McLachlan G J, Do K-A, Ambroise C (2004). Analyzing Microarray

Gene Expression Data. Hoboken, New Jersey: John Wiley & Sons, Int, 213–216

McMillan H K, Clark M P, Bowden W B, Duncan M, Woods R A (2011). Hydrological field data from a model's perspective. Part 1: diagnostic tests for model structure. Hydrol Processes, 25(4): 511–522

Moriasi D N, Arnold J G, Van Liew M W, Bingner R L, Harmel R D, Veith T L (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans ASABE, 50(3): 885–900

Mugunthan P, Shoemaker C A (2006). Assessing the impacts of parameter uncertainty for computationally expensive groundwater models. Water Resour Res, 42(10): W10428, doi: 10.1029/2005WR004640

Nash J E, Sutcliffe J V (1970). River flow forecasting through conceptual models part I — A discussion of principles. J Hydrol (Amst), 10(3): 282–290

Pushpalatha R, Perrin C, Moine N L, Andréassian V (2012). A review of efficiency criteria suitable for evaluating low-flow simulations. J Hydrol (Amst), 420–421: 171–182

Qu S, Bao W, Shi P, Yu Z, Li P, Zhang B, Jiang P (2012). Evaluation of runoff responses to land use changes and land cover changes in the upper Huaihe river basin, China. J Hydrol Eng, 17(7): 800–806

Quirós E, Felicísimo Á M, Cuartero A (2009). Testing multivariate adaptive regression splines (MARS) as a method of land cover classification of TERRA-ASTER satellite images. Sensors (Basel), 9 (11): 9011–9028

Razavi S (2013). Developing efficient strategies for automatic calibration of computationally intensive environmental models. Thesis of University of Waterloo, Waterloo, Ontario, Canada

Razavi S, Tolson B A, Burn D H (2012). Review of surrogate modeling in water resources. Water Resour Res, 48(7): W07401, doi: 10.1029/2011WR011527

Reed P, Minsker B S, Goldberg D E (2003). Simplifying multi-objective optimization: an automated design methodology for the nondominated sorted genetic algorithm–II. Water Resour Res, 39(7): 1196, doi: 10.1029/2002WR001483

Regis R G, Shoemaker C A (2007). A stochastic radial basis function method for the global optimization of expensive functions. INFORMS J Comput, 19(4): 497–509

Reichert P, White G, Bayarri M J, Pitman E B (2011). Mechanism-based emulation of dynamic simulation models: concept and application in hydrology. Comput Stat Data Anal, 55(4): 1638–1655

Ritter A, Muñoz-Carpena R (2013). Performance evaluation of hydrological models: statistical significance for reducing subjectivity in goodness-of-fit assessments. J Hydrol (Amst), 480: 33–45

Singh S K, Bárdossy A (2012). Calibration of hydrological models on hydrologically unusual events. Adv Water Resour, 38: 81–91

Skaggs T H, Barry D A (1997). The first-order reliability method of predicting cumulative mass flux in heterogeneous porous formations. Water Resour Res, 33(6): 1485–1494

Song X, Kong F (2010). Application of Xin'anjiang model coupling with artificial neural networks. Bulletin of Soil and Water Conservation, 30(6): 135–138 (in Chinese)

Song X, Kong F, Zhan C, Han J (2012a). Sensitivity analysis of hydrological model parameter using a statistical theory approach. Advances in Water Science, 23(5): 642–649 (in Chinese)

Song X, Kong F, Zhan C, Han J (2012b). Hybrid optimization rainfall-runoff simulation based on Xin'anjiang model and artificial neural network. J Hydrol Eng, 17(9): 1033–1041

Song X, Kong F, Zhan C, Han J, Zhang X (2013). Parameter identification and global sensitivity analysis of Xin'anjiang model using meta-modeling approach. Water Sci Eng, 6(1): 1–17

Song X, Zhan C, Kong F, Xia J (2011). Advances in the study of uncertainty quantification of large-scale hydrological modeling system. J Geogr Sci, 21(5): 801–819

Song X, Zhan C, Xia J (2012c). Integration of a statistical emulator approach with the SCE-UA method for parameter optimization of a hydrological model. Chin Sci Bull, 57(26): 3397–3403

Song X, Zhan C, Xia J, Kong F (2012d). An efficient global sensitivity analysis approach for distributed hydrological model. J Geogr Sci, 22 (2): 209–222

Srinivasulu S, Jain A (2006). A comparative analysis of training methods for artificial neural network rainfall-runoff models. Appl Soft Comput, 6(3): 295–306

Tang Y, Reed P, Wagener T (2006). How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration? Hydrol Earth Syst Sci, 10(2): 289–307

Trautmann H, Rudolph G, Dominguez-Medina C, Schutze O (2013). Finding evenly spaced Pareto fronts for three-objective optimization problems. In: Schutze O, Coello C A, Tantar A, Tantar E, Bouvry P, Moral P D, Legrand P eds. EVOLVE–A Bridge between Probability Set Oriented Numeric, And Evolutionary Computation II. Heidelberg: Springer-Verlag, 89–105

van Werkhoven K, Wagener T, Reed P, Tang Y (2009). Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. Adv Water Resour, 32(8): 1154–1169

Vrugt J A, Gupta H V, Bastidas L, Bouten W, Sorooshian S (2003b). Efficient and efficient algorithm for multiobjective optimization of hydrologic model. Water Resour Res, 39(8), doi: 10.1029/2002WR001746

Vrugt J A, Gupta H V, Bouten W, Sorooshian S (2003a). A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrological model parameters. Water Resour Res, 39(8), doi: 10.1029/2002WR001642

Vrugt J A, Gupta H V, Nuallain B, Bouten W (2006). Real-time data assimilation for operational ensemble streamflow forecasting. J Hydrometeorol, 7(3): 548–565

Wagner T, Emmerich M, Deutz A, Ponweiser W (2010). On expected-improvement criteria for model-based multi-objective optimization. In: Schaefer R, Cotta C, Kolodziej J, Rudolph G, ed. Parallel Problems Solving from Nature (PPSN) XI, Part I, LNCS 6238. Berlin: Springer, 718–727

Wang Q J (1991). The genetic algorithm and its application to calibrating conceptual rainfall-runoff models. Water Resour Res, 27(9): 2467–2471

Wheater H S (2002). Progress in and prospects for fluvial flood modeling. Phil Trans R Soc Lond A, 360(1796): 1409–1431

Xu C Y (1999). Estimation of parameters of a conceptual water balance model for ungauged catchments. Water Resour Manage, 13(5): 353–368

Xu D M, Wang W C, Chau K W, Chen C T, Chen S Y (2013). Comparison of three global optimization algorithms for calibration of the Xinanjiang model parameters. J HydroInf, 15(1): 174–193

Yan S, Minsker B (2006). Optimal groundwater remediation design using an adaptive neural network genetic algorithm. Water Resour Res, 42, W05407, doi: 10.1029/2005WR004303

Yan S, Minsker B (2011). Applying dynamic surrogate models in noisy genetic algorithms to optimize groundwater remediation designs. J Water Resour Plan Manage, 137(3): 284–292

Yang X H, Mei Y, She D X, Li J Q (2011b). Chaotic Bayesian optimal prediction method and its application in hydrological time series. Comput Math Appl, 61(8): 1975–1978

Yang X H, She D X, Yang Z F, Tang Q H, Li J Q (2009). Chaotic Bayesian method based on multiple criteria decision making (MCDM) for forecasting nonlinear hydrological time series. International Journal of Nonlinear Sciences and Numerical Simulation, 10 (11–12): 1595–1610

Yang X H, Zhang X J, Hu X X, Yang Z F, Li J Q (2011a). Nonlinear optimization set pair analysis model (NOSPAM) for assessing water resource renewability. Nonlinear Process Geophys, 18(5): 599–607

Yao C, Li Z, Yu Z, Zhang K (2012). A prior parameter estimates for a distributed, grid-based Xinanjiang model using geographically based information. J Hydrol (Amst), 468–469: 47–62

Yapo P O, Gupta H V, Sorooshian S (1998). Multi-objective global optimization for hydrologic models. J Hydrol (Amst), 204(1–4): 83–97

Yu Z, Lü H, Zhu Y, Drake S, Liang C (2010). Long-term effects of vegetation on soil hydrological processes in vegetation-stabilized desert ecosystem. Hydrol Processes, 24(1): 87–95

Zhan C S, Song X M, Xia J, Tong C (2013). An efficient integrated approach for global sensitivity analysis of hydrological model parameters. Environ Model Softw, 41: 39–52

Zhang Q, Liu W, Tsang E, Virginas B (2010). Expensive multiobjective optimization by MOEA/D with Gaussian process model. IEEE Trans Evol Comput, 14(3): 456–474

Zhang X S, Srinivasan R, Van Liew M (2009). Approximating SWAT model using artificial neural network and support vector machine. J Am Water Resour Assoc, 45(2): 460–474

Zhao R J (1984). Watershed Hydrological Modeling. Beijing: Water Conservancy and Electric Power Press, 106–130 (in Chinese)

Zhao R J (1992). The Xin'anjiang model applied in China. J Hydrol (Amst), 135(1–4): 371–381

Zhao R J, Liu X R (1995). The Xin'anjiang model. In: Singh V P ed. Computer Models of Watershed Hydrology. Colorado: Water Resources Publications, 215–232

Zou R, Lung W S, Wu J (2007). An adaptive neural network embedded genetic algorithm approach for inverse water quality modeling. Water Resour Res, 43(8): W08427, doi: 10.1029/2006WR005158

Zou R, Lung W S, Wu J (2009). Multiple-pattern parameter identification and uncertainty analysis approach for water quality modeling. Ecol Modell, 220(5): 621–629