Quantifying dynamic sensitivity of optimization algorithm parameters to improve hydrological model calibration

Wei Qi[a,b], Chi Zhang[a*], Guangtao Fu[b] and Huicheng Zhou[a]

[a] School of Hydraulic Engineering, Dalian University of Technology, Dalian 116024, China

[b] Centre for Water Systems, College of Engineering, Mathematics and Physical Sciences, University of Exeter, North Park Road, Harrison Building, Exeter EX4 4QF, UK

[*] Corresponding author: email: czhang@dlut.edu.cn; Tel: +86 411 84708517; Fax: +86 411 84708517

Abstract It is widely recognized that optimization algorithm parameters have significant impacts on algorithm performance, but quantifying the influence is very complex and difficult due to high computational demands and dynamic nature of search parameters. The overall aim of this paper is to develop a global sensitivity analysis based framework to dynamically quantify the individual and interactive influence of algorithm parameters on algorithm performance. A variance decomposition sensitivity analysis method, Analysis of Variance (ANOVA), is used for sensitivity quantification, because it is capable of handling small samples and more computationally efficient compared with other approaches. The Shuffled Complex Evolution method developed at the University of Arizona algorithm (SCE-UA) is selected as an optimization algorithm for investigation, and two criteria, i.e., convergence speed and success rate, are used to measure the performance of SCE-UA. Results show the proposed framework can effectively reveal the dynamic sensitivity of algorithm parameters in the search processes, including individual influences of parameters and their interactive impacts. Interactions between algorithm parameters have significant impacts on SCE-UA performance, which has not been reported in previous research. The

proposed framework provides a means to understand the dynamics of algorithm parameter influence, and highlights the significance of considering interactive parameter influence to improve algorithm performance in the search processes.

Keywords Algorithm; Optimization; SCE-UA; Sensitivity; TOPMODEL; Variance decomposition

1 Introduction

Many optimization algorithms have been proposed to solve hydrological model optimization problems, such as the Shuffled Complex Evolution algorithm developed at the University of Arizona (SCE-UA) (Duan et al., 1992; Duan et al., 1993; Duan et al., 1994), various Genetic algorithms (Deb et al., 2002; Kollat and Reed, 2006; Tang et al., 2006; Fu et al., 2012), and the dynamically dimensioned search algorithm (Tolson and Shoemaker, 2007; Tolson et al., 2009; Asadzadeh and Tolson, 2013). Many studies have been carried out to investigate the strengths and weaknesses of various algorithms, because algorithm performance is of significant concern to users (Duan et al., 1992; Duan et al., 1993; Sorooshian et al., 1993; Bäck, 1996; Thyer et al., 1999; Kollat and Reed, 2006; Tolson and Shoemaker, 2007; Zhang et al., 2008; van Werkhoven et al., 2009; Wang et al., 2010; Fu et al., 2012; Arsenault et al., 2014; Chao et al., 2015; Qi et al., 2015).

It is widely recognized that algorithm parameters have a significant influence on algorithm performance, but quantifying the influence is very complex and difficult due to high computational demands and dynamic nature of search parameters (Giorgos et al., 2015). Many optimization applications use trial and error to determine parameter values, or simply use default parameter values without investigating their influence on algorithm performance

(Deb et al., 2002; Tolson and Shoemaker, 2007). However, attempts have been made to find optimal parameter combinations. For example, Duan et al. (1994) analyzed the performance of SCE-UA under different parameter combinations for a hydrological model calibration problem, and suggested that many combinations could produce good performance in terms of success rate which was defined as the ratio of success among a number of algorithm runs. However, it has been pointed out that the parameter values suggested by Duan et al. (1994) may be inefficient, when other algorithm performance criteria: for example, convergence speed, are considered (Behrangi et al., 2008a; Tolson and Shoemaker, 2008). More importantly, Duan et al. (1994) did not considered the interactions among parameters, that is, only the individual impacts of algorithm parameters were considered.

Hadka and Reed (2011) proposed a framework to assess the influence of multi-objective algorithm parameters based on Sobol'’s global sensitivity analysis method (Sobol', 2001). However, the proposed framework has a huge computational demand, due to the use of Sobol'’s method. In the study of Hadka and Reed (2011), 280 million algorithm runs were executed on a CyberStar computing cluster which consists of 512 2.7 GHz processors and 1536 2.66 GHz processors. This huge computational burden is not affordable with commonly available computational resources. Further, Hadka and Reed (2011) did not show the dynamic sensitivity of optimization algorithm parameters which is particularly useful to understand the convergence speed in hydrological model calibration.

A variance decomposition-based method - Analysis of Variance (ANOVA) has been used to quantify the influence of uncertain contributors in a process in many studies. It allows for the analysis of individual and interactive impacts of contributors, and therefore allows for the identification of influential contributors and the understanding of parameter interactions. For instance, it has been used to quantify the influence of climate models, statistical downscaling approaches and hydrological models on projected future flows (Bosshard et al., 2013). This method has also been used to investigate the influence of climate change scenarios on water resources, the influence of climate change uncertainties on projected future flows, and the impacts of climate changes on flow frequency (Köplin et al., 2013; Addor et al., 2014; Giuntoli et al., 2015). In these investigations, respective contributions of various uncertainty sources to the overall output variance have been compared, and ANOVA has shown good performance.

Dynamic sensitivity analysis can reveal the changes of the influences of individual parameters and their interactions during a search process. Most recently, it has gained increasing attention in the field of hydrological modeling. For example, the dynamic sensitivity of hydrological model parameters has been studied to understand the variations of modelled hydrological processes, and to verify the modifications of hydrological models (Pfannerstill et al., 2015). In addition, advancements have been made in studying the dynamic effects of hydrological model formulations, dynamic performance of hydrological models and dynamic tuning of algorithm parameters (Rolf, 1982; Sandip et al., 2009; van Werkhoven et al., 2009; Eiben and Smit, 2011; Reusser et al., 2011; Reusser and Zehe, 2011; Garambois et al., 2013; Herman et al., 2013). However, to the best of our knowledge, few studies have been carried out to investigate the dynamic sensitivity of optimization algorithm parameters.

The overall aim of this paper is to provide a global sensitivity analysis-based framework to dynamically quantify individual and interactive impacts of algorithm parameters on optimization performance. ANOVA was employed to quantify the impacts, because it is more computationally efficient compared with Sobol'’s approach. The SCE-UA algorithm was selected as an optimization algorithm to demonstrate the framework. The proposed framework was first tested on five benchmark test functions, with up to 12 dimensions, and then applied to a TOPMODEL hydrological model calibration problem, representing different problems of various levels of difficulty. Two algorithm performance criteria - convergence speed and success rate - were compared in terms of parameter influence. The framework provides an improved understanding of the significant roles of algorithm parameters in the optimization processes, and highlights the importance of considering interactive influence among parameters, which is beyond the information that can be provided by conventional approach. Thus it can assist hydrological model calibration by selecting more appropriate algorithm parameter values to improve calibration efficiency, which is particularly important for a computationally intensive model.

2 Algorithm and materials

2.1 SCE-UA algorithm

SCE-UA algorithm was investigated because the influence of its parameters had been investigated in many studies (Duan et al., 1994; Behrangi et al., 2008a; Tolson and Shoemaker, 2008). The SCE-UA has four main features: (1) combination of deterministic and probabilistic approaches; (2) systematic evolution of complex points; (3) complex shuffling; and (4) competitive evolution. These characteristics stand for a combination of several approaches, including the simplex method (Nelder and Mead, 1965), the control random search (Price, 1987) and evolutionary algorithms (Holland, 1975). The introduction of complex shuffling in SCE-UA is an advanced technique which successfully ensures that the information of all populations is shared by each individual complex. Initially, a set of individuals are randomly sampled from the parameter space, and then selected individuals are divided into several complexes. Each complex evolves using a competitive evolutionary algorithm. All individuals are shuffled and reassigned to new complexes to enable information sharing. As the search progresses, the entire population moves to global optimal solutions. A detailed description of SCE-UA can be found in Duan et al. (1993).

The SCE-UA performance is affected by objective functions, dimensions of decision variables and data used for calibration (Duan et al., 1994; Tolson and Shoemaker, 2007; Behrangi et al., 2008a; Tolson and Shoemaker, 2008). Thus five benchmark test functions, with up to 12 dimensions, and a hydrological model for flood simulations were employed to represent different levels of complexities.

2.2 Benchmark test functions

The five benchmark test functions were Rastrigin, Ackley, Levy and Montalvo 1 (LM1), Levy and Montalvo 2 (LM2) and Levy. These functions are characterized by a large number of local minima and a large search space, and have been chosen by many researchers to evaluate optimization algorithms (Ali et al., 2005; Deep and Thakur, 2007; Tolson and Shoemaker, 2007; Behrangi et al., 2008a; Tolson and Shoemaker, 2008; Chia et al., 2011). The equations of these benchmark test functions were listed in Appendix A.

2.3 Hydrological model calibration problem

The Biliu river basin (2814 km^2), located in a peninsula region between the Bohai Sea and the Huanghai Sea, China, was used for the TOPMODEL calibration. It covers longitudes from 122.29°E to 122.92°E and latitudes from 39.54°N to 40.35°N. This basin is characterized by a monsoon climate, and summer (July to September) is the main rainfall period. The average annual temperature is 10.5 °C, and the lowest and the highest temperature is -4.7 °C in January and 24 °C in August, respectively. The major land cover types are forest and farmland. There are eleven rainfall gauges and one discharge gauge. The basin average rainfall was calculated using the Thiessen method, and six flood data with different flood magnitudes were used in calibration to represent the influence of data on SCE-UA performance.

TOPMODEL is a physically based, variable contributing area model which combines the advantages of a simple lumped parameter model with distributed effects (Beven and Kirkby, 1979). Fundamental of TOPMODEL's parameterization are three assumptions: (1) saturated-zone dynamics can be approximated by successive steady-state representations; (2) hydrological gradients of the saturated zone can be approximated by the local topographic surface slope; and (3) the transmissivity profile whose form exponentially declines along the vertical depth of the water table or storage, is spatially constant. On the basis of above mentioned assumptions, the index of hydrological similarity is represented as the topographic index $\ln(a/\tan\beta)$ where a is the area per unit contour length and β is local slope angle. The greater upslope contributing areas and lower gradient areas are more likely to be saturated. More detailed description of TOPMODEL and its mathematical formulations can be found in Beven and Kirkby (1979). TOPMODEL has been widely used, because of its relatively simple model structure (Blazkova and Beven, 1997; Cameron et al., 1999; Hossain and Anagnostou, 2005; Bastola et al., 2008; Gallart et al., 2008; Bouilloud et al., 2010; Qi et al., 2013). TOPMODEL consists of six parameters, and their ranges and brief descriptions were given in Table 1.

The Nash-Sutcliffe Efficiency (NSE) was selected as a performance metric for TOPMODEL calibration:

$$NSE = 1 - \frac{\sum_{t=1}^{T}\left(Q_{st} - Q_{mt}\right)^2}{\sum_{t=1}^{T}\left(Q_{mt} - Q_m^m\right)^2} \tag{1}$$

where Q_{st} (m^3/s) and Q_{mt} (m^3/s) are the simulated and measured flows at time t; T is the total number of flood data points and Q_m^m (m^3/s) is the average of measured flows. The best theoretical value of NSE is 1.0. As SCE-UA was set up for minimization problems in this study, the following objective function was used in the TOPMODEL calibration

$$f = 1 - NSE \tag{2}$$

The best theoretical value of f is 0.0, while its true minimum value is unknown for real calibration problems since model and data errors exist.

3 Methodology

Fig. 1 shows the flowchart of the proposed framework. The framework includes three main components for an investigated algorithm: (1) selection of concerned parameter values and random combinations (Fig. 1a); (2) selection of performance metrics which should reflect the concerns of algorithm users: for example, convergence speed and success rate, which are illustrated in Fig. 1b; and (3) use of ANOVA to decompose the contributions of parameters and their interactions to reveal the influence of parameters on algorithm performance, as shown in Fig. 1c where the influence on convergence speed and success rate is shown as a three parameter case. It should be noted that the sample number for each parameter can be different, that is, m_1, m_i and m_n are not required to be equal in Fig. 1a.

The remainder of this section will illustrate the framework using SCE-UA algorithm and selected calibration problems.

3.1 SCE-UA parameters and performance metrics

Three parameters of SCE-UA were investigated: (1) complex number (P), (2) reflection parameter (alpha) and (3) contraction parameter (beta), as suggested by several studies (Tolson and Shoemaker, 2007; Behrangi et al., 2008a; Tolson and Shoemaker, 2008). The selected SCE-UA parameters P, alpha and beta are in the ranges of [1, 40], [0.1, 3.0] and [0.05, 1], respectively. It should be noted that P must be an integer. The parameter ranges were defined based on the following studies: Duan et al. (1994), Tolson and Shoemaker (2007) and Tolson and Shoemaker (2008).

In this paper, 11 values for each selected parameter were randomly selected from parameter ranges considering the computational burdens. Fig. 2 depicts the random combinations of algorithm parameters, and every combination was used to optimize objective functions f. In each box of Fig. 2, the number is the selected parameter values, and three values out of the 11 values were shown.

Two algorithm performance criteria, convergence speed and success rate, were studied. These two criteria are of concern for researchers (Duan et al., 1994; Behrangi et al., 2008a; Tolson and Shoemaker, 2008). Convergence speed is assessed by averaging the best objective function value f over several random seed trial runs at every function evaluation (Tolson and Shoemaker, 2007; Tolson and Shoemaker, 2008). In this study, 30 and 10 random seed trial runs were used in benchmark function and TOPMODEL calibration, respectively. Success rate measures the ability to find global optimal solutions (Duan et al., 1994), and was evaluated as

$$Success\ rate = \frac{1}{N}\left\{numnber\ of\ \ f_{end}\ such\ that\ \left| f_{end} - f_{optimal} \right| \le e \right\} \qquad (3)$$

where f_{end} is a best objective function value obtained at the end of optimization; $f_{optimal}$ is a known optimal objective function value which can be a theoretical value or a specified value if theoretical value is unknown; e is an error limit and specified by algorithm users; N is the number of algorithm runs: for example, 30 and 10 runs were used in benchmark function and TOPMODEL calibration problems respectively. The reasons why these numbers of runs were used are explained in Section 4. Each parameter combination in Fig. 2 corresponds to a convergence speed and a success rate, and therefore $11 \times 11 \times 11$ convergence speed data at every function/model evaluation and success rates can be obtained, where number 11 represents the number of selected parameter values. ANOVA was used to decompose the convergence speed and success rate variances resulted from 1331 parameter combinations into contributions of individual SCE-UA parameters and parameter interactions. To relate performance criteria (M) to algorithm parameters, superscripts *j*, *k* and *l* in $M^{j,k,l}$ were used to represent *P*, alpha and beta, respectively, in the equations below.

3.2 Influence quantification

It has been argued that ANOVA approach is based on a biased variance estimator that underestimates the variance when a small sample size is used (Bosshard et al., 2013). To reduce the effects of the biased estimator on contribution quantification, Bosshard et al. (2013) proposed a subsampling method, which was also used in this study. This subsampling approach does not need extra optimization trials; therefore it can reduce the computational burden. In each subsampling iteration *i*, we selected two *P* values out of all *P* values, and the superscript j in calculating $M^{j,k,l}$ was replaced with $g(h,i)$. The total number of 2-combination is 55 in this study, and correspondingly, the superscript *g* is a 2×55 matrix as follows

$$g = \begin{pmatrix} 1 & 1 & \cdots & 1 & 2 & 2 & \cdots & 8 & 8 & 8 & 9 & 9 & 10 \\ 2 & 3 & \cdots & 11 & 3 & 4 & \cdots & 9 & 10 & 11 & 10 & 11 & 11 \end{pmatrix} \quad (4)$$

Based on ANOVA, the total sum of squares (*SST*) can be divided into sums of squares due to the individual and interactive effects:

$$SST = SSA + SSB + SSC + SSI \quad (5)$$

where *SSA* is the contribution of *P*; *SSB* is the contribution of alpha; *SSC* is the contribution of beta; and *SSI* is the contribution of their interactions.

The terms can be estimated using the subsampling procedure as follows (Bosshard et al., 2013):

$$SST_i = \sum_{h=1}^{H} \sum_{k=1}^{K} \sum_{l=1}^{L} \left(M^{g(h,i),k,l} - M^{g(o,i),o,o} \right)^2 \quad (6)$$

$$SSA_i = K \cdot L \cdot \sum_{h=1}^{H} \left(M^{g(h,i),o,o} - M^{g(o,i),o,o} \right)^2 \quad (7)$$

$$SSB_i = H \cdot L \cdot \sum_{k=1}^{K} \left(M^{g(o,i),k,o} - M^{g(o,i),o,o} \right)^2 \quad (8)$$

$$SSC_i = H \cdot K \cdot \sum_{l=1}^{L} \left(M^{g(o,i),o,l} - M^{g(o,i),o,o} \right)^2 \quad (9)$$

$$SSI_i = \sum_{h=1}^{H} \sum_{k=1}^{K} \sum_{l=1}^{L} \left(M^{g(h,i),k,l} - M^{g(h,i),o,o} - M^{g(o,i),k,o} - M^{g(o,i),o,l} + 2 \cdot M^{g(o,i),o,o} \right)^2 \quad (10)$$

where symbol ^{o} indicates the averaging over the particular index. Then the contribution of each influential source η^2 is calculated as follows:

$$\eta_P^2 = \frac{1}{I} \sum_{i=1}^{I} \frac{SSA_i}{SST_i} \quad (11)$$

$$\eta_{alpha}^2 = \frac{1}{I} \sum_{i=1}^{I} \frac{SSB_i}{SST_i} \quad (12)$$

$$\eta_{beta}^2 = \frac{1}{I} \sum_{i=1}^{I} \frac{SSC_i}{SST_i} \quad (13)$$

$$\eta^2_{interaction} = \frac{1}{I}\sum_{i=1}^{I}\frac{SSI_i}{SST_i} \tag{14}$$

η^2 has a value between 0 and 1, which represents the respective contribution to the overall variations of M.

4 Results and discussion

4.1 Benchmark functions

In the simulations, SCE-UA algorithm was stopped when the total number of function evaluations reached a prescribed value. In the flowing subsections, the contributions of individual SCE-UA parameters and parameter interactions to the variance of convergence speed at every function evaluation and success rate are quantified for the selected benchmark functions.

4.1.1 Convergence speed analyses

Fig. 3 shows the contributions of individual SCE-UA algorithm parameters and their interactions in terms of convergence speed in benchmark function calibration, where average best function values over 30 random seed trial runs were used. The 30 random seed trial runs were used considering computational burden, and were the same as many other studies: for example, Deep and Thakur (2007), Tolson and Shoemaker (2007) and Chia et al. (2011). The benchmark functions were optimized under 6, 8, 10 and 12 dimensions. The contributions of individual parameters and their interactions are represented by color strips varying with the function evaluation number shown in the x-axis.

For the 6-dimensinal Rastrigin function, the influence of *P* increases and then decreases, while the impacts of beta and alpha increase with an increase in function evaluation number. The influence of alpha is larger than beta, and the influence of *P* at early stages is larger than alpha and beta. The interactions among *P*, beta and alpha have significant influence, decreasing with an increase in function evaluations. Interactive impacts are larger than those from any individual parameter at initial search stages, and have approximately the same influence as *P* and alpha, but have a slightly larger influence than beta at later optimization stages. For other 6-dimensional functions, similar results are shown; except that, for LM1, LM2 and Levy at later stages, the influence of beta becomes larger than *P*, alpha and interactions, and that the influence of alpha becomes the smallest. The differences result from differences in benchmark functions, which implies that objective functions have influence on algorithm performance and that using several test functions is necessary.

Comparing different dimensions at later stages, with a dimension increase, influence of *P* increases but influence of beta decreases, whilst alpha influence and interactive influence remain approximately the same, which indicates with an increase in dimensions the importance of *P* increases but the importance of beta decreases. This information implies that dimensions have influence on the performance of parameters, and that optimal parameter values derived from low dimensional problems may not have optimal performance for high dimensional problems. All results show that the contributions from various sources become almost constant at the end of the search process, indicating that 1000 function evaluations are sufficient.

4.1.2 Success rate analyses

The contributions to success rate based on the 30 random seed trial runs are shown in Fig. 4 under an error level of 0.001 in terms of benchmark function calibration. The error level represents the absolute differences between an optimal objective function value found at the end of the optimization and a real optimal value, and is subjectively selected: for example, Duan et al. (1994) has used 0.001 as an error limit, and Deep and Thakur (2007) and Chia et al. (2011) have used 0.01 as error limits.

For the 6-dimensional Rastrigin function, *P*, beta and alpha all have significant contributions, and alpha contributes more than *P* and beta, while interactions account for the majority. Comparing different dimensions, with a dimension increase, the contributions of *P*, beta and alpha decrease, but interactive contribution increases. Compared with other functions, similar results can be obtained, except that the contribution of alpha is smaller than beta for 6-dimensional LM1 function. The differences may result from the limited number of random trials. These results indicate that, for the success rate, interactions among parameters are most important, and good combinations of parameters are more important than individual parameters. The results are different from convergence speed analyses. This difference indicates parameters have a different influence when algorithm performance criteria change. It should be noted that the contributions actually includes influence of initial random seeds, but this influence should be very small after many function evaluations (Wang et al., 2010). In addition, the success rate is influenced by the number of function evaluations, but in our study the investigations of convergence speed and success rate used the same number of function evaluations: thus the comparison results are free of influence. Another error limit 0.005 was also analyzed, and similar results are obtained.

4.2 TOPMODEL

Every parameter combination can generate a convergence speed line and a success rate in TOPMDOEL calibration, and therefore 1331 convergence speed lines and success rates were obtained. They are shown in Fig. 5 using flood 1984-06-15 as an examples. Different colors are used to distinguish lines in Fig. 5a. Because the theoretically optimal objective function values were not known, the optimal values obtained from all the 1331×10 optimization runs were used. The variations of histogram heights represent the variance of success rate, as shown in Fig. 5b. Fig. 5a shows there are many vertical lines before 500 function evaluations which are resulted from larger *P* values. This information implies that *P* has larger influence before 500 function evaluations. Significant differences exist in convergence speed and success rate, which can be attributed to the variations of parameter values. Thus it is necessary to analyze the parameter influence. In the flowing subsections, the contributions of individual SCE-UA parameters and parameter interactions are quantified for all six flood events.

4.2.1 Convergence speed analyses

Fig. 6 shows the contributions of SCE-UA algorithm parameters and interactions in terms of convergence speed in TOPMODEL calibration, where average best function values over 10 random seed trial runs were used considering the computational burdens. The number of random seed trial runs are similar to the study by Duan et al. (1994). Each panel represents the results from a flood event. The contributions of individual parameters and their interactions are represented by strips varying with the model evaluations shown in the x-axis.

Fig. 6a shows the influence of *P* increases first and then decreases. However, the influence of alpha grows with an increase in model evaluations, and the contribution of beta slightly increases. Interactions among *P*, alpha and beta have significant contributions, decreasing with an increase in model evaluations. Interactive impacts are larger than beta in all model evaluations, while significantly higher than *P* at initial stages and at later stages. Compared with alpha, interactive influence is larger at initial stages, and is a little smaller at later stages. This information implies that without considering interactions the calibration of parameters may not be effective for improving algorithm performance. For other flood events, similar results can be obtained. These results are consistent with the convergence speed variance in Fig. 5a where *P* has greater influence at early stage. This is because that larger *P* values can slow the information exchange among different complexes. Consequently, larger *P* values have few positive efforts in improving convergence speed at early optimization stage. This information implies that larger influence does not suggest greater convergence speed.

Comparing results in each panel, differences can be attributed to the different roles that parameters play in the SCE-UA calibration processes, while differences among panels result from the influence of data. The complex number *P* controls information exchange among complexes; with an increase in model evaluations, information exchange among complexes doesn’t provide more positive influence in searching for optimal solutions compared with early stages, which implies the complex number has significant influence on the searching speed at early stage. However, for alpha, much more positive influence arises with an increase in model evaluations. Comparing Fig. 3 and Fig. 6, the influence of beta is the smallest in Fig. 6, which is different from the results of the 6-dimensinal functions in Fig. 3. This difference results from objective functions and errors in data used in Fig. 6, which implies that objective functions and data have significant influence besides variable dimensions. All results show patterns are clearly revealed at the end of optimization, and thus 1000 model evaluations are sufficient.

4.2.2 Success rate analyses

The contributions to success rate in the 10 runs under an error level of 0.001 in terms of TOPMODEL calibration are shown in Fig. 7.

Fig. 7a shows the contribution of beta is the smallest, and the contribution of alpha is the largest among individual parameter contributions. However, interactions among parameters contribute the most. Similarly, other five cases show the contributions of alpha are the greatest among individual parameter contributions, or at least not smaller than individual parameter contributions. Comparing the differences among different flood data, Figs. 7d, 7e and 7f show the contribution of beta is larger than P, and Fig. 7b shows contributions of beta and *P* are equal. These differences may result from different flood data and optimal objective function values: for example, the optimal objective function value is 0.0223 for Fig. 7a, and is 0.193 for Fig. 7d. This implies that calibration data have impacts on the parameter influence, and therefore using several flood data sets is necessary. Compared with Fig. 4, similar results can be obtained, which indicates that the results could be applicable to other calibration problems.

In Fig. 5b, the success rate has several peaks, and these peaks are the results of some good parameter combinations that have relatively small *P* values (smaller than 5), which may be because the smaller dimension 6 and limited model evaluations (Duan et al., 1994). When dimension increases, required *P* and model evaluation number should increase to obtain high success rate (Duan et al., 1994). This information implies *P* has large influence on greater success rate, which is different from Fig. 7a where interactions contribute the majority of the variance. This difference is resulted from the differences in definitions of success rate and variance: success rate measures the ability of finding optimal results, but variance measures the changes of this ability along the variations of parameter values. This information implies that larger influence does not guarantee greater success rate. Another error limit 0.005 was also analyzed, and similar results are obtained.

4.3 Discussion

There has been a trend to develop parsimonious algorithms and adaptive parameter control schemes for users' convenience and reduction in algorithm complexity (Gao et al., 2014; Wu et al., 2014; Yu et al., 2014; Goldman and Punch, 2015). However, the proposed framework in this study provides a means to understand the performance of optimization algorithms by revealing the dynamics of parameter sensitivity in the search processes. In addition, the dynamic sensitivity can provide information to set dynamic algorithm parameter values, which could provide a method to improve algorithm efficiency (Eiben and Smit, 2011; Rui et al., 2015). Furthermore, the dynamic sensitivity information could provide evidence for assigning appropriate parameter values in different optimization stages to improve the fitness of optimization algorithms (Giorgos et al., 2015).

In the study by Tolson and Shoemaker (2007), the convergence speed of the SCE-UA was assessed based on adjustments of parameter *P*, and the results were problematic because other parameters: such as, beta and alpha, were not considered, as was pointed out by Behrangi et al. (2008a). Although Behrangi et al. (2008a) realized the influence of other parameters, they did not quantitatively show the influence nor explicitly indicated interactions among parameters. In contrast, the results of this study do quantitatively compare the influence of parameters and explicitly show the dynamic impacts of interactions along the number of function evaluations. This information could guide algorithm development and applications: for example, if an algorithm parameter is not sensitive, it would be helpless to tune this parameter to change algorithm performance; if a parameter has greater sensitivity than the sum of other parameters and interactions, the calibration efficiency may be mainly determined by this parameter and calibrating other parameters may be ineffective to change algorithm performance.

In the study by Duan et al. (1994), the importance of *P* was stressed, and it suggested that *P* should increase with an increase in the difficulty of model calibration problems to obtain a high success rate. However, our study reveals that alpha could have a larger influence than *P* on success rate, and more importantly, the interactions could play an important role in success rate. This information will help optimization algorithm parameter selections in hydrological model calibration, and promote further development in searching for optimal parameters for SCE-UA given consideration of parameter interactions.

It should be noted that the success rate is influenced by the number of function evaluations and error limits. There are several parameter combinations that are failed to success within 1000 function/model evaluations under an error limit 0.001. More function/model evaluations are needed if it is needed to make sure all parameter combinations are successful. In addition, the SCE-UA parameter ranges and the random seed trial runs could also have influence on results. However, the case study of this research shows that *P* is not always the most influential parameter; the developed framework can provide a means to quantify the influences of function evaluation number, error limits, parameter ranges and random seed trial runs on the parameter sensitivity, which can be done by comparing the sensitivity differences of several numbers of these influential variables. It should also be noted that the variance decomposition results revel the variations of convergence speed and success rates along parameter variations, but larger influence does not guarantee faster convergence speed and greater success rate. Larger influence just suggests convergence speed and success rate can be significantly changed when parameter values are altered.

Convergence speed and success rate have to be considered carefully in the model calibration process in practice. Essentially, the selection of algorithm parameter values is based on modellers' preference to convergence speed or success rate, and the computational demand of a hydrological model also plays a key role. Duan et al. (1994) provided guidance for model calibration but it can be applied to success rate only (Behrangi et al., 2008b). However, in this study, we showed how the convergence speed is affected by the parameters and a need to balance convergence and success rate. The value of P should be carefully selected to improve convergence speed at an early stage during optimization; the values of beta and alpha should have more attention in order to improve the convergence speed at a later stage. For success rate, alpha can be more influential than P.

Using the Rastrigin function with up to 12 dimensions as an example, Fig. 8 shows comparison of the convergence speed curves (black bold line) from a set of default parameter values suggested by Duan et al. (1994) and the lower convergence speed boundray curves (red bold line) from the 1331 parameter combinations. Three points (A, B and C) from the lower convergence speed boundray lines are selected and corresponding parameter values are shown as well. Points A, B and C correspond to 100, 400 and 700 function evaluations, respectively.

In the three cases of varying dimensions, the best combination of paramter values is different at different function evaluation numbers, implying that one combination of parameter values can not maintain good performance during the search process. It should be noted that, in the cases of 6- and 8-dimensions, although the *alpha* values are the same at points A, B and C, the *P* and *beta* values are different: thus the paramter value combinations are different at points A, B and C. Because the best parameter values that have the best convergence speed vary at differet function evaluation numbers, it is difficult to provide a set of parameter values that can maintain the best convergence speed during the search process. However, in this study, we provide useful information on the parameter influence on convergence speed in the search process, including interactive influences of parameter values, and therefore we provide an enhanced understanding of SCE-UA algorithm parameter value setting. Future research is encourgaed to develop dynamic parameter values in the search process to improve the convergence speed.

In Fig. 8 it can be seen that there is a gap between the two bold convergence lines, indicating that an improvemrnt can be achieved by changing the default parameter values. In addition, it can be seen that the gaps become wider with an increase in the dimension, and this implies that higher gains in the convergence speed improvements can be obtained for high dimension optimization problems compared with low dimension problems. Thus, quantfiying dynamic sensitivity of parameters reveals useful information for model calibration.

It should be noted that hydrological models such as TOPMODEL have the equafinality problem, which is defined as that many sets of different parameter values are acceptable and result in the same objective function values (Beven and Binley, 1992; Beven and Freer, 2001). However, the equafinality problem does not include the influence on the variations of objective function values, and therefore its influence is negligible in algorithm performance assessment (Tolson and Shoemaker, 2007; Tolson and Shoemaker, 2008; Zhang et al., 2008; Arsenault et al., 2014).

5 Conclusions

The diverse control mechanisms of algorithm parameters in algorithm performance should be investigated, which can provide users with the information on which parameter is most influential and on how influence changes along function evaluation number and algorithm performance criteria. This study developed a new framework to quantify dynamic sensitivity of optimization algorithm parameters and their interactions based on ANONA, and investigated the influence of the parameters of SCE-UA using a suite of benchmark functions and a hydrological model calibration problem. The major findings are as follows.

First, the proposed framework can effectively reveal the dynamic sensitivity of algorithm parameters in the search process, including individual influences of parameters and their interactive impacts on algorithm performance. This provides an effective tool to gain an improved understanding of the significant roles of algorithm parameters.

Second, the value of *P* should be carefully selected to improve convergence speed at early optimization stage; beta and alpha should draw much more attention to improve the convergence speed at later optimization stage. For success rate, alpha can be more influential than *P*.

Third, parameter combinations could have significant influence on algorithm performance, which highlights the importance of considering interactive influence among parameters.

The proposed framework can guide efforts to calibrate algorithm parameters to improve computational efficiency in hydrological model calibration processes. In the future, a sensitivity-based parameter auto-adjusting approach will be studied for SCE-UA.

Acknowledgements:

This study was supported by the National Natural Science Foundation of China (Grant No. 51320105010 and 51279021). The first author gratefully acknowledges the financial support provided by the China Scholarship Council. The data from Biliu basin were obtained from the Biliu reservoir administration.

Appendix A: Benchmark functions

No.	Function	Definition	Parameter Space	Optimal Value
1	Rastrigin (1974)	$f(x)=\sum_{i=1}^{D}\left[x_i^2-\cos(2\pi x_i)\right]$	$[-2,2]^D$	$f^*=-D$
2	Ackley (1987)	$f(x)=-20\exp\sqrt{\frac{1}{D}\sum_{i=1}^{D}x_i^2}-\exp\left[\frac{1}{D}\sum_{i=1}^{D}\left(\cos(2\pi x_i)\right)\right]$	$[-1,3]^D$	$f^*=-20-e$
3	Levy and Montalvo 1 (LM1)	$f(x)=\frac{\pi}{n}\left(10\sin^2(\pi y_1)+\sum_{i=1}^{n-1}(y_i-1)^2\left[1+10\sin^2(\pi y_{i+1})\right]+(y_n-1)^2\right), y_i=1+\frac{1}{4}(x_i+1)$	$[-10,10]^D$	$f^*=0$
4	Levy and Montalvo 2 (LM2)	$f(x)=0.1\left(\sin^2(3\pi x_1)+\sum_{i=1}^{n-1}(x_i-1)^2\left[1+\sin^2(3\pi x_{i+1})\right]+(x_n-1)^2\left[1+\sin^2(2\pi x_n)\right]\right)$	$[-5,5]^D$	$f^*=0$
5	Levy	$f(x)=\left(\sin^2(\pi y_1)+\sum_{i=1}^{n-1}(y_i-1)^2\left[1+10\sin^2(\pi y_{i+1}+1)\right]+(y_n-1)^2\left(1+10\sin^2(2\pi y_n)\right)\right), y_i=1+\frac{1}{4}(x_i+1)$	$[-10,10]^D$	$f^*=0$

References

Addor, N., Rössler, O., Köplin, N., Huss, M., Weingartner, R., Seibert, J., 2014. Robust changes and sources of uncertainty in the projected hydrological regimes of Swiss catchments. Water Resources Research, 50(10): 7541-7562.

Ali, M.M., Khompatraporn, C., Zabinsky, Z.B., 2005. A numerical evaluation of several stochastic algorithms on selected continuous global optimization test problems. Journal of Global Optimization, 31(4): 635-672.

Arsenault, R., Poulin, A., Côté, P., Brissette, F., 2014. Comparison of Stochastic Optimization Algorithms in Hydrological Model Calibration. Journal of Hydrologic Engineering, 19(7): 1374-1384.

Asadzadeh, M., Tolson, B., 2013. Pareto archived dynamically dimensioned search with hypervolume-based selection for multi-objective optimization. Engineering Optimization, 45(12): 1489-1509.

Bäck, T., 1996. Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms. Oxford University Press.

Bastola, S., Ishidaira, H., Takeuchi, K., 2008. Regionalisation of hydrological model parameters under parameter uncertainty: A case study involving TOPMODEL and basins across the globe. Journal of Hydrology, 357(3-4): 188-206.

Behrangi, A., Khakbaz, B., Vrugt, J.A., Duan, Q., Sorooshian, S., 2008a. Comment on "Dynamically dimensioned search algorithm for computationally efficient watershed model calibration" by Bryan A. Tolson and Christine A. Shoemaker. Water Resources Research, 44(12): W12603.

Behrangi, A., Khakbaz, B., Vrugt, J.A., Duan, Q., Sorooshian, S., 2008b. Comment on "Dynamically dimensioned search algorithm for computationally efficient watershed model calibration" by Bryan A. Tolson and Christine A. Shoemaker. Water Resources Research, 44(12).

Beven, K., Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes, 6(3): 279-298.

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

Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrological Sciences Bulletin, 24(1): 43-69.

Blazkova, S., Beven, K., 1997. Flood frequency prediction for data limited catchments in the Czech Republic using a stochastic rainfall model and TOPMODEL. Journal of Hydrology, 195(1-4): 256-278.

Bosshard, T., Carambia, M., Goergen, K., Kotlarski, S., Krahe, P., Zappa, M., Schär, C., 2013. Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections. Water Resources Research, 49(3): 1523-1536.

Bouilloud, L., Chancibault, K., Vincendon, B., Ducrocq, V., Habets, F., Saulnier, G.-M., Anquetin, S., Martin, E., Noilhan, J., 2010. Coupling the ISBA Land Surface Model and the TOPMODEL Hydrological Model for Mediterranean Flash-Flood Forecasting: Description, Calibration, and Validation. Journal of Hydrometeorology, 11(2): 315-333.

Cameron, D.S., Beven, K.J., Tawn, J., Blazkova, S., Naden, P., 1999. Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty). Journal of Hydrology, 219(3-4): 169-187.

Chao, Z., Ru-bin, W., Qing-xiang, M., 2015. Calibration of Conceptual Rainfall-Runoff Models Using Global Optimization. Advances in Meteorology, 2015.

Chia, J.Y., Goh, C.K., Tan, K.C., Shim, V.A., 2011. Memetic informed evolutionary optimization via data mining. Memetic Computing, 3(2): 73-87.

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 6(2).

Deep, K., Thakur, M., 2007. A new crossover operator for real coded genetic algorithms. Applied Mathematics and Computation, 188(1): 895-911.

Duan, Q., Gupta, V.K., Sorooshian, S., 1993. Shuffled Complex Evolution Approach for Effective and Efficient Global Minimization. Journal of optimization theory and Applications, 73(3).

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

Duan, Q., Sorooshian, S., Gupta, V.K., 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. Journal of Hydrology, 158(3-4): 265-284.

Eiben, A.E., Smit, S.K., 2011. Parameter tuning for configuring and analyzing evolutionary algorithms. Swarm and Evolutionary Computation, 1(1): 19-31.

Fu, G., Kapelan, Z., Reed, P., 2012. Reducing the Complexity of Multiobjective Water Distribution System Optimization through Global Sensitivity Analysis. Journal of Water Resources Planning and Management, 138(3): 196-207.

Gallart, F., Latron, J., Llorens, P., Beven, K.J., 2008. Upscaling discrete internal observations for obtaining catchment-averaged TOPMODEL parameters in a small Mediterranean mountain basin. Physics and Chemistry of the Earth, 33(17-18): 1090-1094.

Gao, W., Yen, G.G., Liu, S., 2014. A cluster-based differential evolution with self-adaptive strategy for multimodal optimization. IEEE transactions on cybernetics, 44(8): 1314-1327.

Garambois, P.A., Roux, H., Larnier, K., Castaings, W., Dartus, D., 2013. Characterization of process-oriented hydrologic model behavior with temporal sensitivity analysis for flash floods in Mediterranean catchments. Hydrology and Earth System Sciences, 17: 2305-2322.

Giorgos, K., Mark, H., Eiben, A.E., 2015. Parameter Control in Evolutionary Algorithms: Trends and Challenges. IEEE Transactions on Evolutionary Computation, 19(2): 167-187.

Giuntoli, I., Vidal, J.P., Prudhomme, C., Hannah, D.M., 2015. Future hydrological extremes: the uncertainty from multiple global climate and global hydrological models. Earth System Dynamics Discussions, 6(1): 1-30.

Goldman, B.W., Punch, W.F., 2015. Fast and Efficient Black Box Optimization using the Parameter-less Population Pyramid. Evolutionary Computation.

Hadka, D., Reed, P., 2011. Diagnostic Assessment of Search Controls and Failure Modes in Many-Objective Evolutionary Optimization. Evolutionary Computation, 20(3): 423-452.

Herman, J.D., Reed, P.M., Wagener, T., 2013. Time-varying sensitivity analysis clarifies the effects of watershed model formulation on model behavior. Water Resources Research, 49: 1400-1414.

Holland, J.H., 1975. Adaptation in Natural and Artificial Syatems. University of Michigan Press, Ann Arbor.

Hossain, F., Anagnostou, E.N., 2005. Assessment of a stochastic interpolation based parameter sampling scheme for efficient uncertainty analyses of hydrologic models. Computers & Geosciences, 31(4): 497-512.

Kollat, J.B., Reed, P.M., 2006. Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design. Advances in Water Resources, 29(6): 792-807.

Köplin, N., Schädler, B., Viviroli, D., Weingartner, R., 2013. The importance of glacier and forest change in hydrological climate-impact studies. Hydrology and Earth System Sciences, 17(2): 619-635.

Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. The Computer Journal, 7(4).

Pfannerstill, M., Guse, B., Reusser, D., Fohrer, N., 2015. Temporal parameter sensitivity guided verification of process dynamics. Hydrology and Earth System Sciences Discussions, 12: 1729-1764.

Price, W.L., 1987. Global optimization algorithms for a CAD workstation. Journal of Optimization Theory and Applications, 55(1).

Qi, W., Zhang, C., Chu, J., Zhou, H., 2013. Sobol''s sensitivity analysis for TOPMODEL hydrological model: A case study for the Biliu River Basin, China. Journal of Hydrology and Environment Research, 1: 1-10.

Qi, W., Zhang, C., Fu, G., Zhou, H., 2015. Global Land Data Assimilation System data assessment using a distributed biosphere hydrological model. Journal of Hydrology, 528: 652-667.

Reusser, D.E., Buytaert, W., Zehe, E., 2011. Temporal dynamics of model parameter sensitivity for computationally expensive models with the Fourier amplitude sensitivity test. Water Resources Research, 47: W07551.

Reusser, D.E., Zehe, E., 2011. Inferring model structural deficits by analyzing temporal dynamics of model performance and parameter sensitivity. Water Resources Research.

Rolf, I., 1982. Parameter adaptive control algorithms—A tutorial. Automatica: 513528.

Rui, W., Maszatul, M.M., Robin, C.P., Peter, J.F., 2015. An analysis of parameter sensitivities of preference-inspired co-evolutionary algorithms. International Journal of Systems Science: 1-14.

Sandip, A., Rajeev, K., Chakrabarti, P.P., 2009. Adaptive parameter control of evolutionary algorithms to improve quality-time trade-off. Applied Soft Computing, 9(2): 527-540.

Sobol', I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation, 55(1-3): 271-280.

Sorooshian, S., Duan, Q., Gupta, V.K., 1993. Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model. Water Resources Research, 29(4): 1185-1194.

Tang, Y., Reed, P., Wagener, T., 2006. How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration? Hydrology and Earth System Sciences, 10(2): 289-307.

Thyer, M., Kuczera, G., Bates, B.C., 1999. Probabilistic optimization for conceptual rainfall-runoff models: A comparison of the shuffled complex evolution and simulated annealing algorithms. Water Resources Research, 35(3): 767-773.

Tolson, B.A., Asadzadeh, M., Maier, H.R., Zecchin, A., 2009. Hybrid discrete dynamically dimensioned search (HD-DDS) algorithm for water distribution system design optimization. Water Resources Research, 45(12): W12416.

Tolson, B.A., Shoemaker, C.A., 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resources Research, 43(1): W01413.

Tolson, B.A., Shoemaker, C.A., 2008. Reply to comment on "Dynamically dimensioned search algorithm for computationally efficient watershed model calibration" by Ali Behrangi et al. Water Resources Research, 44(12).

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

Wang, Y.-C., Yu, P.-S., Yang, T.-C., 2010. Comparison of genetic algorithms and shuffled complex evolution approach for calibrating distributed rainfall–runoff model. Hydrological Processes, 24(8): 1015-1026.

Wu, X., Rozycki, P., Wilamowski, B.M., 2014. A Hybrid Constructive Algorithm for Single-Layer Feedforward Networks Learning. IEEE transactions on neural networks and learning systems.

Yu, W.-J.J., Shen, M., Chen, W.-N.N., Zhan, Z.-H.H., Gong, Y.-J.J., Lin, Y., Liu, O., Zhang, J., 2014. Differential evolution with two-level parameter adaptation. IEEE transactions on cybernetics, 44(7): 1080-1099.

Zhang, X., Srinivasan, R., Zhao, K., Liew, M.V., 2008. Evaluation of global optimization algorithms for parameter calibration of a computationally intensive hydrologic model. Hydrological Processes, 23(3): 430-441.

Table 1 TOPMODEL parameters

Name (units)	Description	Lower bound	Upper bound
SZM (m)	parameter of exponential decline in conductivity	0.005	0.04
LNT0 (m^2 h^{-1})	effective lateral saturated transmissivity	-25	10
RV (m^2 h^{-1})	hill slope routing velocity	3500	8000
SR_{max} (m)	maximum root zone storage	0.001	0.01
SR_0 (m)	initial root zone deficit	0	0.01
TD (m h^{-1})	unsaturated zone time delay per unit deficit	0.5	5

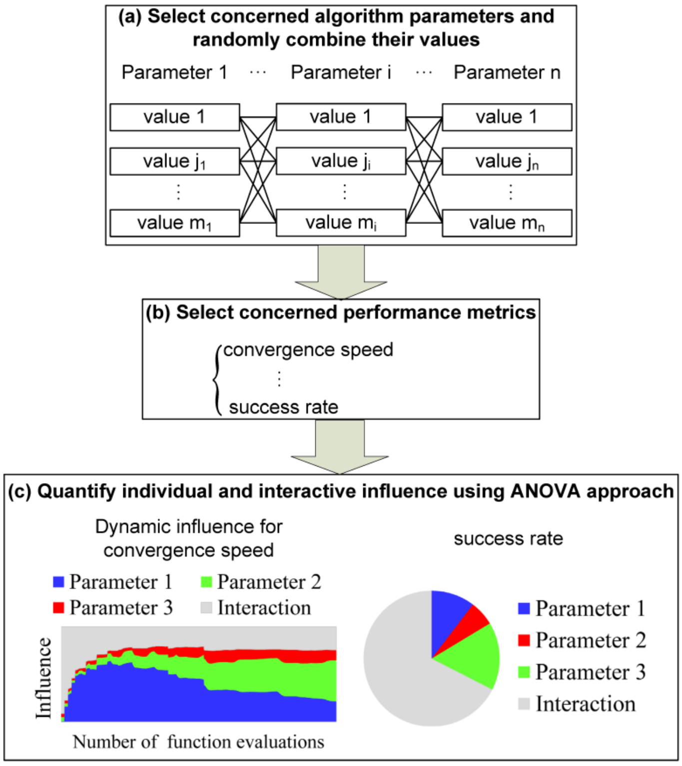


Fig. 1 Flowchart of the proposed framework.

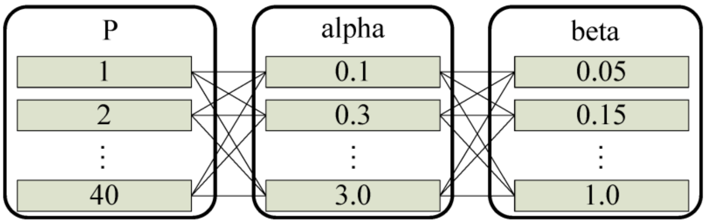


Fig. 2 Combinations of the SCE-UA algorithm parameters: *P*, alpha and beta.

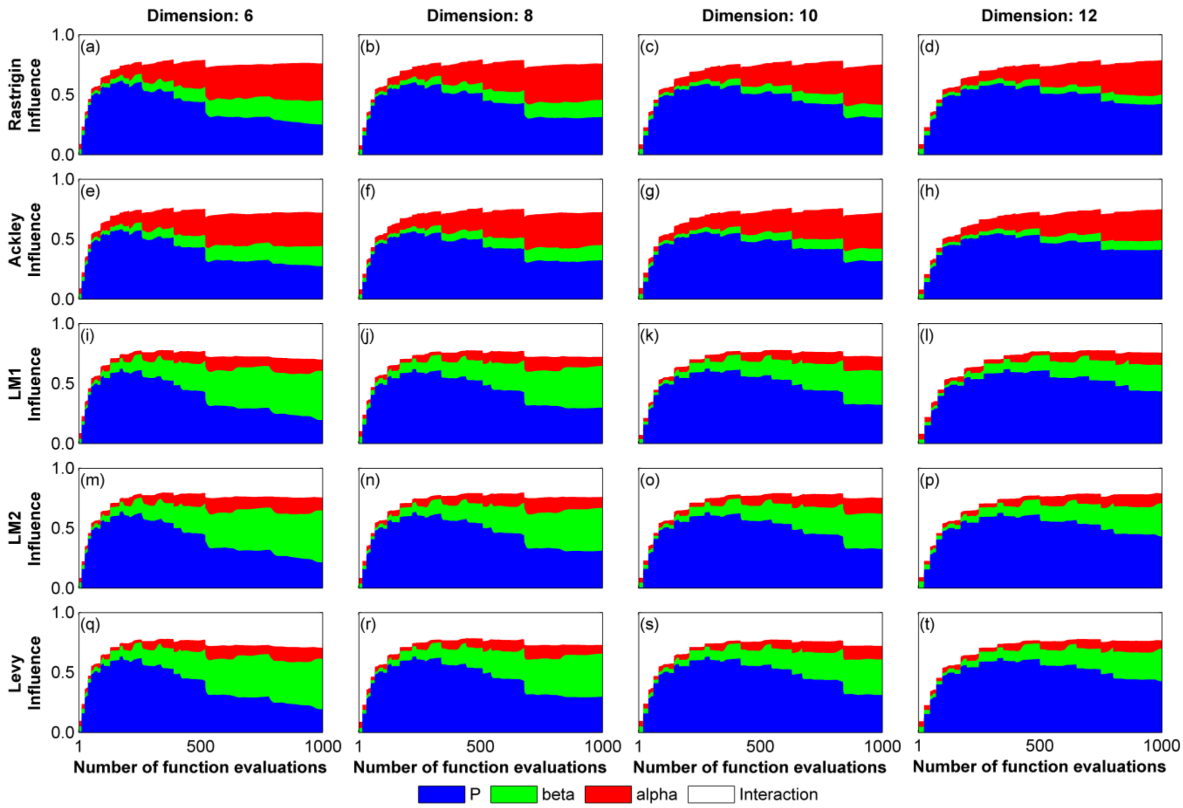


Fig. 3 Contributions of individual SCE-UA parameters and their interactions in terms of convergence speed in benchmark function calibration. Each row represents a benchmark function with 6, 8, 10 and 12 dimensions.

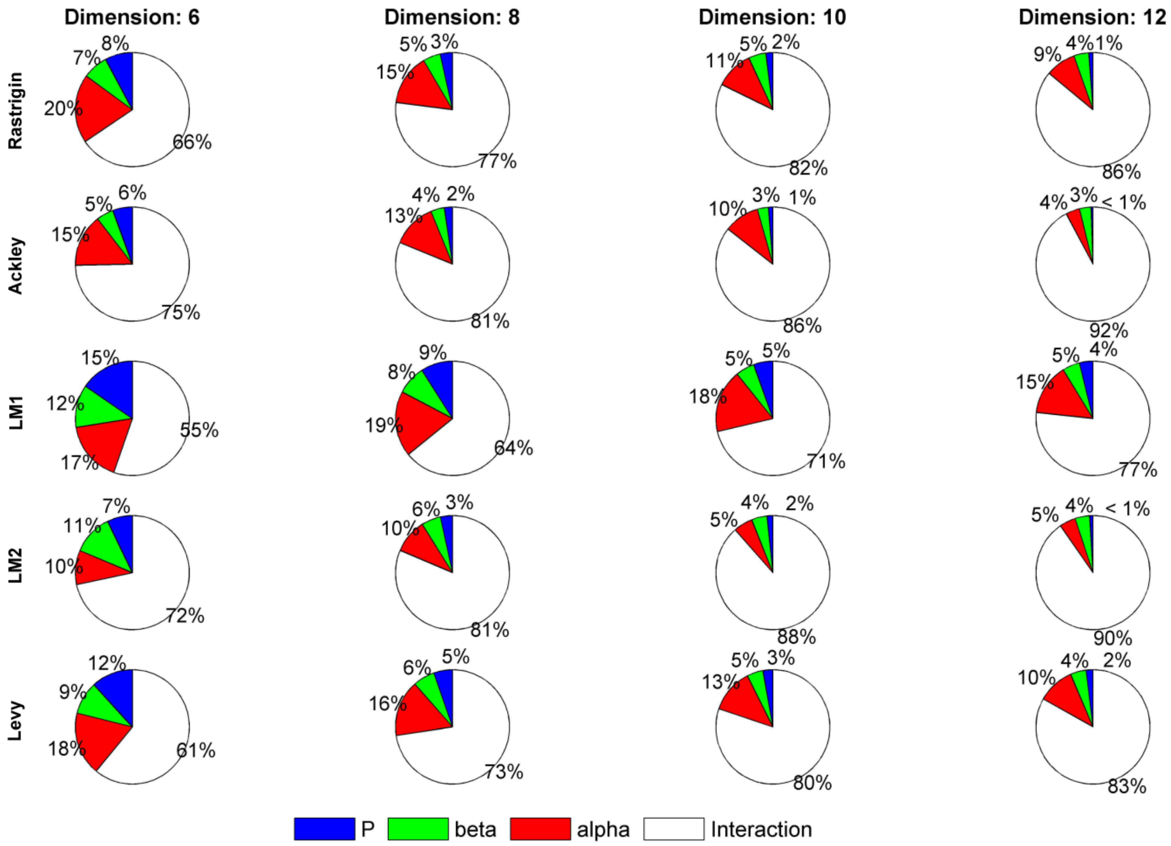


Fig. 4 Contributions of individual SCE-UA parameters and interactions in terms of success rate in benchmark function calibration. Each row represents a benchmark function with 6, 8, 10 and 12 dimensions.

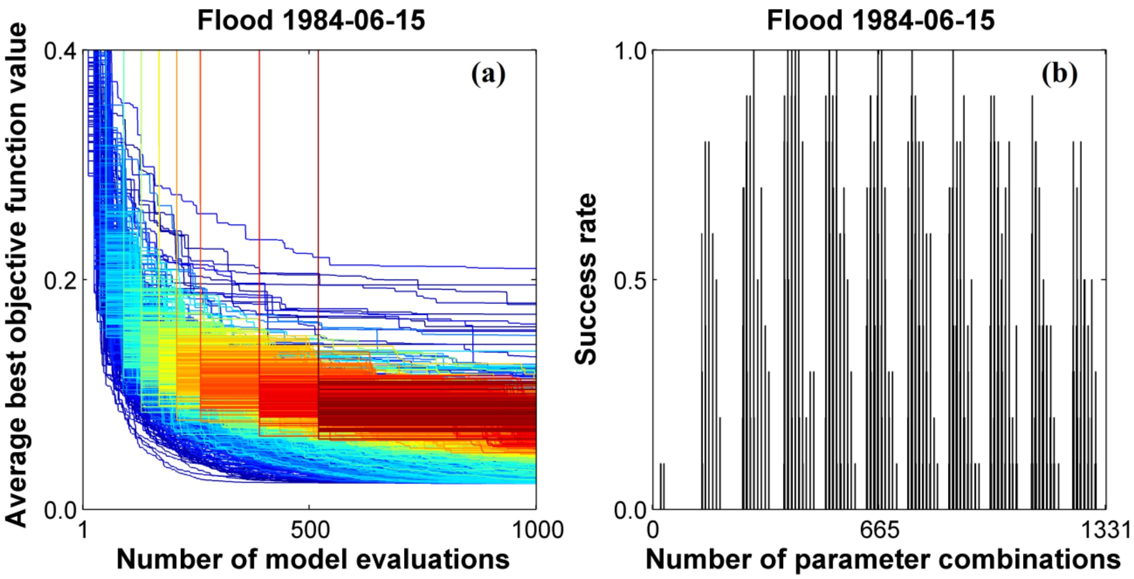


Fig. 5 Convergence speed and success rate variances for flood 1984-06-15, which are generated from 1331 parameter combinations. Different convergence speed lines are represented using different colors in Fig. 5a. The variations of the histogram heights in Fig.5b represent the variance of success rate.

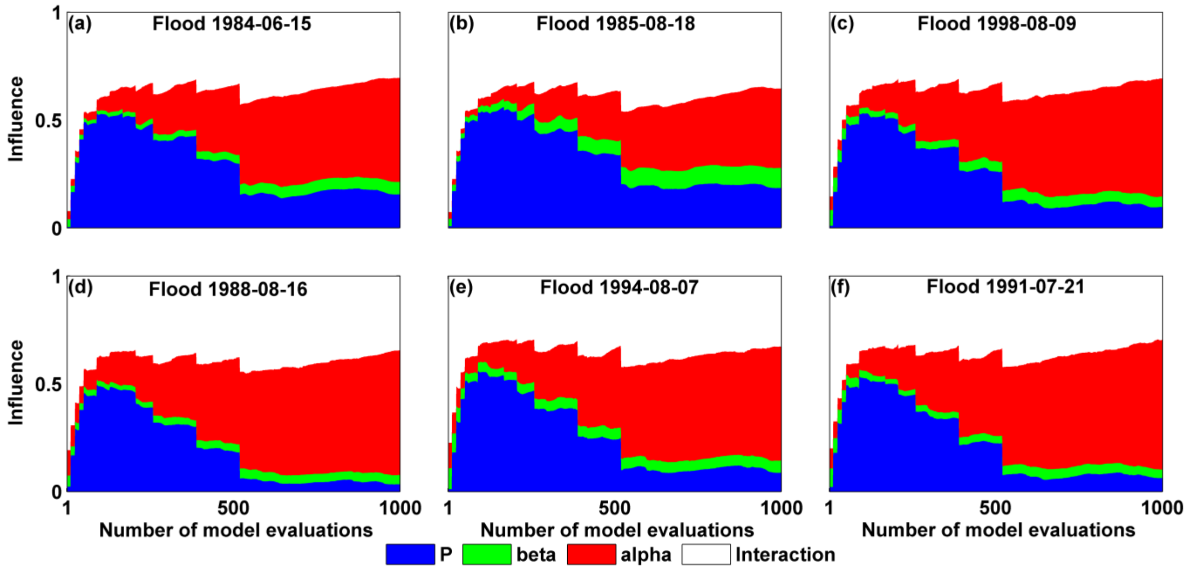


Fig. 6 Contributions of individual SCE-UA parameters and their interactions in terms of convergence speed in TOPMODEL calibration. Each figure represents a flood calibration problem.

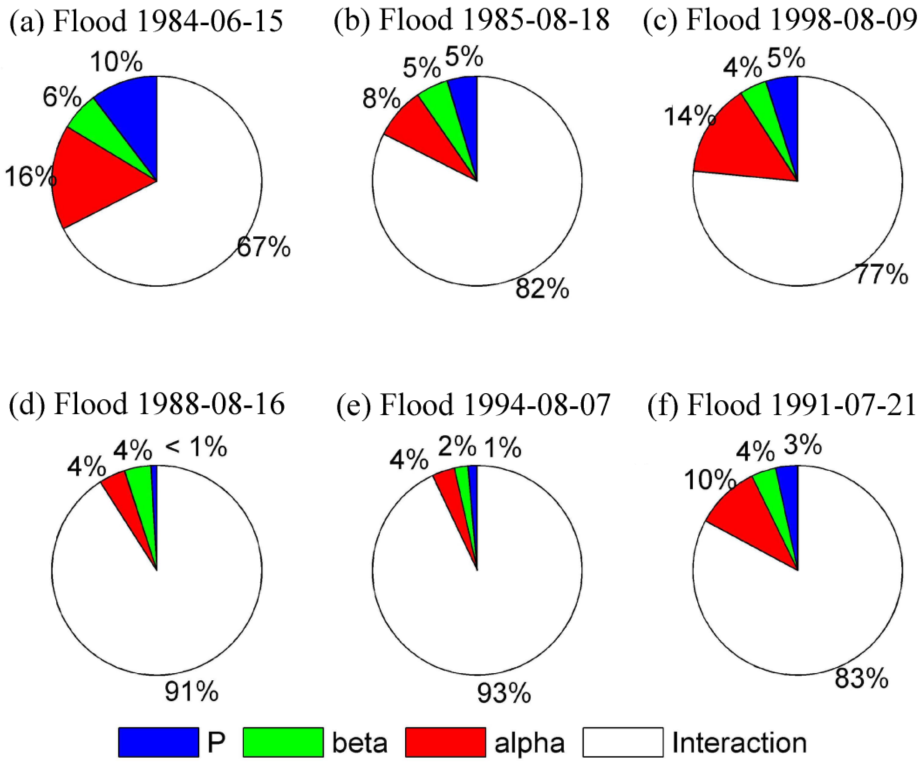
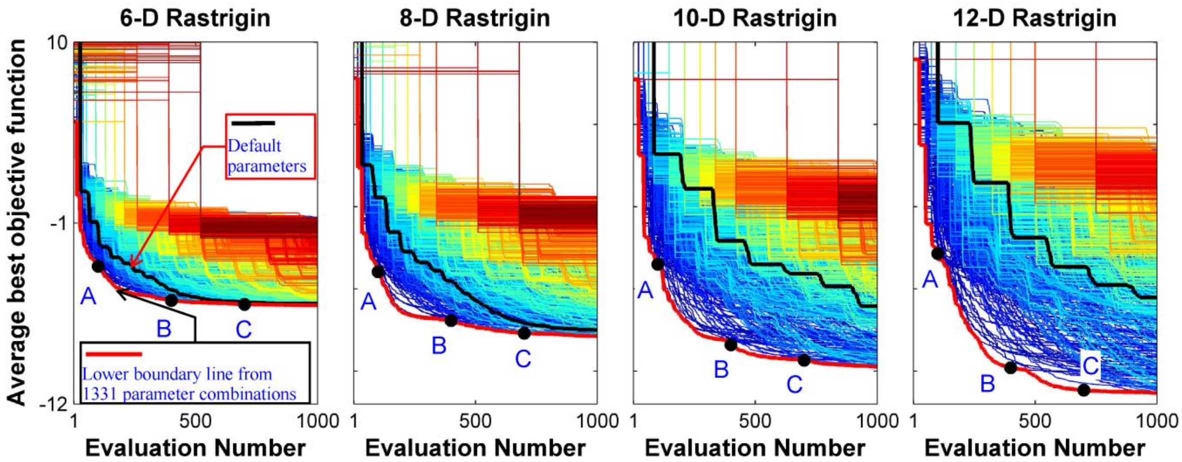


Fig. 7 Contributions of individual SCE-UA parameters and their interactions in terms of success rate in TOPMODEL calibration. Each figure represents a flood calibration problem.



6-D Rastrigin:

	P	beta	alpha
A	1	0.9	0.6
B	2	0.35	0.6
C	4	0.45	0.6

8-D Rastrigin:

	P	beta	alpha
A	1	0.8	0.6
B	2	0.6	0.6
C	3	0.6	0.6

10-D Rastrigin:

	P	beta	alpha
A	1	0.35	0.3
B	2	0.6	0.6
C	3	0.5	0.6

12-D Rastrigin:

	P	beta	alpha
A	1	0.5	0.3
B	1	0.35	0.6
C	2	0.5	0.6

Parameter value combinations of points A, B and C from the lower boundray lines

Fig. 8 Comparison of the convergence speed curves (black bold line) from a set of default parameter values suggested by Duan et al. (1994) and the lower convergence speed boundray curves (red bold line) from the 1331 parameter combinations. Three points (A, B and C) from the lower convergence speed boundray lines and their corresponding parameter values are shown.