

An evaluation of adaptive surrogate modeling based optimization with two benchmark problems



Chen Wang, Qingyun Duan^{*}, Wei Gong, Aizhong Ye, Zhenhua Di, Chiyuan Miao

State Key Laboratory for Earth Surface Processes and Resource Ecology, College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China

ARTICLE INFO

Article history:

Received 3 September 2013

Received in revised form

23 May 2014

Accepted 26 May 2014

Available online 2 July 2014

Keywords:

Adaptive surrogate modeling based optimization

Design of experiment

Adaptive sampling

Global sensitivity analysis

Computationally intensive computer models

ABSTRACT

Surrogate modeling uses cheap “surrogates” to represent the response surface of simulation models. It involves several steps, including initial sampling, regression and adaptive sampling. This study evaluates an adaptive surrogate modeling based optimization (ASMO) method on two benchmark problems: the Hartman function and calibration of the SAC-SMA hydrologic model. Our results show that: 1) Gaussian Processes are the best surrogate model construction method. A minimum Interpolation Surface method is the best adaptive sampling method. Low discrepancy Quasi Monte Carlo methods are the most suitable initial sampling designs. Some 15–20 times the dimension of the problem may be the proper initial sample size; 2) The ASMO method is much more efficient than the widely used Shuffled Complex Evolution global optimization method. However, ASMO can provide only approximate optimal solutions, whose precision is limited by surrogate modeling methods and problem-specific features; and 3) The identifiability of model parameters is correlated with parameter sensitivity.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

1. Introduction

Computer simulation models generally contain model parameters that cannot be measured directly but can only be inferred by calibration, a trial-and-error process that adjusts the parameter values to fit the model simulation outputs to the observations of the real system (Duan et al., 1994). Model calibration problem can be converted to a global optimization problem which aims to minimize the difference between model simulations and corresponding observations. Duan et al. (1992, 1993, 1994) developed a widely used global optimization procedure, entitled the shuffled complex evolution (SCE-UA) algorithm. Many case studies have demonstrated that SCE-UA algorithm is an effective and efficient method for model calibration (Sorooshian et al., 1993; Gan and Biftu, 1996; Hogue et al., 2000; Boyle et al., 2000; Moreno et al., 2012). However, global optimization methods generally require up to tens of thousands of model runs to find the global optimal solution. This may place severe computational constraint on solving such an optimization problem, if the underlying model requires a large amount of CPU time to run. One approach to reduce the

computational burden is to approximate and replace the expensive simulation model with a cheaper-to-run surrogate model. There are two broad families of surrogates: (1) response surface surrogates, which are statistical or empirical data-driven models emulating the responses of a high-fidelity simulation model; and (2) lower-fidelity physical based surrogates, which are simplified models of the original system (Razavi et al., 2012b). In this study, we focus on the former one (referred as surrogate model hereafter). Some fields also refer to the surrogate modeling as function approximation, meta-modeling, response surface method, or model emulation (Blanning, 1975; O'Hagan, 2006).

A surrogate model can be understood as a “model of model”. It is a statistical model of the response surface of a simulation model. A surrogate model describes the relationship between inputs (i.e., model's adjustable parameters) and outputs (i.e., the performance measure of the simulation model). Training an accurate surrogate model needs adequate input–output data, which are obtained by running the simulation model with different sets of parameters selected in the feasible parameter space. Previous studies use the “one-shot” approach (i.e., using a set of samples at once) to obtain input–output data to construct the surrogate model. This method uses data to establish a surrogate model directly. Then it runs global optimization algorithm on the surrogate model. A high number of model runs may be required to ensure that the surrogate model

^{*} Corresponding author. Tel.: +86 (0) 10 5880 4191.

E-mail address: qyduan@bnu.edu.cn (Q. Duan).

represents the response surface of the original simulation model well. One way to economically construct a surrogate model for optimization is to use adaptive sampling. Adaptive sampling means that we sample a certain number of points in the initial stage and then adaptively sample additional points which can most effectively increase the accuracy of the surrogate model. For the purpose of finding an optimum, it is not necessary to map out the whole surface in a surrogate model exploration. An adaptive sampling strategy can quickly move the experiment to a region containing the optimum of the input variables. Only within this region is a thorough exploration of the surrogate model warranted to find the optimum (Wu and Hamada, 2009).

There are various studies of surrogate modeling. Queipo et al. (2005) presented an overview of “one-shot” surrogate based analysis and optimization methods in the aerospace science and engineering field. They covered some of the most popular methods in design of experiments (DoEs), surrogate model construction, model selection and validation, sensitivity analysis, and surrogate modeling based optimization. Forrester and Keane (2009) reviewed progresses made in surrogate modeling research over the past 20 years. Their review includes different adaptive sampling methods and surrogate based multi-objective optimization. Also, some other papers generalized the framework of surrogate modeling and using this for sensitivity analysis and optimization of complex environmental models (Castelletti et al., 2012; Ratto et al., 2012; Carnevale et al., 2012).

In terms of research progress in adaptive surrogate modeling research, we found a number of studies in the field of hydrology and geosciences. For example, Razavi et al. (2012a,b) published a very comprehensive review of adaptive surrogate modeling and its applications to hydrological modeling and water resources management. Also they compared optimization strategies based on surrogate models against other common optimization strategies not involving surrogate models. Mousavi and Shourian (2010) applied adaptive sequentially space-filling (ASSF) meta-modeling method on optimal water quantity allocation problem. Their results show that ASSF method model with fewer actual function evaluations is able to find comparable solutions to other techniques. Castelletti et al. (2010) used a multi-objective adaptive recursive meta-modeling optimization approach to improving water quality planning in lakes and reservoirs. Neelin et al. (2010) made use of a surrogate modeling based parameter optimization and sensitivity analysis method to identify precipitation bias and sensitivity of an atmospheric General Circulation Model. They found that their approach is simple to guide parameter choices and to aid inter-comparison of sensitivity properties among climate models.

The studies cited above have adequately demonstrated that surrogate modeling based optimization methods are promising for solving computationally intensive optimization problems. However, what is the best surrogate model construction method and what is the most efficient adaptive sampling method are not clear. Razavi et al. (2012b) pointed out that the effectiveness and efficiency of surrogate based optimization methods can be impacted by the choice of initial DoEs and initial sample sizes. Even though previous studies have looked at this issue, what initial DoEs and sample sizes should be used for a specific surrogate modeling based optimization problem is not clearly demonstrated (Jones et al., 1998; Sóbester et al., 2005; Loepky et al., 2009). In this study, we aim to contribute toward answering those questions.

This paper is organized as follows. Section 2 presents the surrogate modeling based optimization framework and a brief introduction of methods used in every step. Section 3 presents the procedures and results of two case studies. Section 4 provides the conclusions.

2. The adaptive surrogate modeling based optimization methods

2.1. The procedure in adaptive surrogate modeling based optimization

The adaptive surrogate modeling based optimization (referred as ASMO hereafter) method involves several steps. First, the parameters to be optimized are chosen. Sensitivity analysis is often used to screen out the insensitive parameters and identify the parameters that exert the most influence on model performance for further optimization (Shi et al., 2012). Then a surrogate model based on adaptive sampling is constructed to represent the input–output response surface of the simulation model. Finally an optimization search is conducted on the surrogate model. Assuming that the most sensitive parameters have already been screened out, a specific implementation of how to construct an adaptive surrogate model for optimization is illustrated in Fig. 1 and is described as follows:

Step 1: An initial set of sample points (i.e., parameter sets) is generated using a specific DoE method. These points are then used as inputs to run the simulation model. The objective function values are then computed at all sample points. Generally, at this stage the location of the points is only required to satisfy parameters' probability distribution and some space-filling criterion. However, if more detailed prior knowledge is available on the parameters, it may be incorporated in the sampling design. This study assumes that no other prior knowledge is available except the parameter lower and upper bounds.

Step 2: A surrogate model type is selected for representing the response surface of the simulation model. Several possible surrogate model types can be considered: Gaussian process (GP), polynomial regression, multi-variate adaptive regression splines (MARS), support vector machines (SVMs), among others. Given the surrogate model type, the surrogate model is constructed by fitting a statistical model to the performance measures of the simulation model at the sample points.

Step 3: New sample points are generated using an adaptive sampling strategy. The adaptive strategy makes full use of the

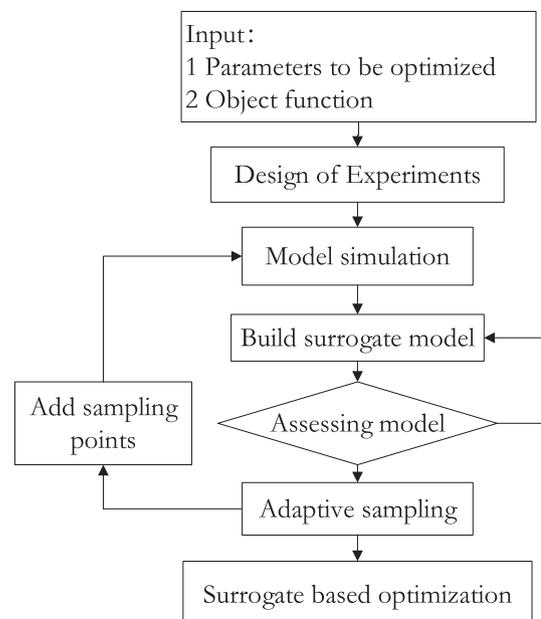


Fig. 1. A schematic description of the ASMO scheme.

information gained from existing sample points to favor new sample points in the promising regions. These new sample points and their simulation results from the simulation model are then used to update the existing response surface.

Step 4: A convergence check is performed on the updated response surface. If one or more pre-specified convergence criteria (see Section 3.1 for definition) are met, then an optimization search is performed on the final response surface. The simulation model is run with this optimal parameter set and the corresponding objective function value is recorded and checked against the best objective function value from the prior search. If it is the new objective function value is better than the previous one, then we accept this parameter set from the surrogate model as the optimal parameter set. Otherwise, new sets of adaptive samples are generated until the resulted response surface is acceptable.

2.2. Design of experiment (DoE)

DoE is a body of techniques that enable an investigator to conduct better experiments, analyze data efficiently and make the connections between the conclusions from the analysis and the original objectives of the investigation (Wu and Hamada, 2009). The choice of the DoE method is an important step in the ASMO method. DoE serves to put an initial set of sample points in the feasible parameter space. There are a wide variety of DoE methods available. Among them, Full Factorial, Fractional Factorial (Box et al., 2005), Central Composite (Myers, 1971), Box–Behnken (Box and Behnken, 1960), Monte Carlo, Latin Hypercube (McKay et al., 1979) and Quasi-Monte Carlo (QMC) (Sobol', 1967; Halton, 1964) designs are the most widely used. However, not all methods are suitable for surrogate based optimization. Some methods need a large number of sampling points and become impractical for high-dimensional problems. For instance, the number of sampling points required for a Full Factorial design is exponentially proportional to the dimensionality of the problem (e.g., a 6-dimensional space with only 10 levels needs 10^6 points). Other methods have low space filling capability, e.g., Box–Behnken and crude Monte Carlo methods. Methods such as Latin Hypercube and QMC are commonly used in ASMO because they allow flexible sample sizes, have a good space filling capability and are capable of representing the whole space with relatively few sample points. Therefore, we evaluate several space-filling DoEs in this study, including Latin hypercube and two QMC designs. A brief description of these designs is given in Appendix I.

2.3. Surrogate model construction methods

Generally, surrogate model construction methods are statistical regression methods that estimate response surface of a simulation model. A variety of approximation techniques have been developed and applied as the surrogates of an original simulation model: polynomial regression (Fen et al., 2009), regression tree method (Tree) (Breiman et al., 1984), Random Forest (RF) method (Breiman, 2001), MARS (Friedman, 1991), Support Vector Machines (SVMs) (Zhang et al., 2009), Artificial Neural Networks (ANNs) (Behzadian et al., 2009) and GP (Rasmussen and Williams, 2006). At the highest level, response surfaces can be differentiated based on whether they are non-interpolating (i.e., it minimizes the sum of squared errors from some predetermined functional form) or interpolating (i.e., it passes through all points). It has been suggested that non-interpolating surfaces, such as fitted quadratic surfaces, are unreliable for surrogate based optimization because the surface may not sufficiently capture the shape of the function (Jones, 2001). On the other hand, interpolating methods can get more and more accurate as new points are added, eventually converging to the true

function. In this study, we test seven different surrogate model construction methods on the Hartman function and then choose the best one to be used for evaluating the DoEs and initial sample sizes for ASMO. A brief description of the seven methods is given in Appendix II.

2.4. Adaptive sampling methods

Adaptive sampling methods (also called sequential design methods) are iterative algorithms that use data acquired from previous iterations to guide future sample selections. Adaptive sampling methods allow significant reduction in the number of simulations of the original simulation model because they only search the area that may contain the optimum of the input variables.

There are many adaptive sampling approaches, including “Minimizing an Interpolating Surface (MIS)” (Jones, 2001), “Minimizing a Statistical Lower Bound (MSL)” (Cox and John, 1997), “Maximizing the Probability of Improvement (MPI)” (Stuckman, 1988) and “Maximizing Expected Improvement (MEI)” (Schonlau et al., 1997; Picheny et al., 2013). All these methods have advantages and disadvantages. Some choose the points where the probability of improving the response surface is maximized, e.g., MSL, MPI and MEI methods. These three methods rely on the standard error of the GP regression method to force the algorithm to go back and explore regions where the sampled points are sparse. However, for surrogate based optimization problem, it is not necessary to describe the whole response surface. Instead, it is enough just to delineate a region containing the optimum. MIS is a method not focused on building the whole response surface but only the region containing the optimum. Because studies have shown that the MEI method is superior to MSL and MPI methods (Jones, 2001), we compare only the MEI method against the MIS method on Hartman function. See Appendix III for a brief description of MIS and MEI.

3. Evaluation of the ASMO method with two case studies

We evaluate the efficiency and effectiveness of the ASMO method presented in Section 2 with two case studies. These two case studies are benchmark problems that are used to in evaluating optimization methods. The first case is a simple mathematical function: the Six-Dimensional Hartman function. This example is used to show the feasibility of the ASMO method and different adaptive strategies. We test different surrogate model construction methods and two adaptive sampling methods in this case. Further, we discuss how different initial DoE methods and initial sample sizes affect the optimization results. The second case study is closer to a real-world hydrological modeling problem, i.e., calibration of the parameters of the Sacramento Soil Moisture Accounting (SAC-SMA) hydrologic model using the ASMO method. The second case study makes use of the findings in the first case study.

3.1. Case study I: the six-dimensional Hartman function

The six-dimensional Hartman function is a commonly used mathematical function used to test optimization algorithms. Its expression is shown as follows:

$$f(x) = - \sum_{i=0}^3 c \cdot \exp \left(- \sum_{j=0}^5 A_{i,j} (x_j - p_{i,j})^2 \right) \quad (1)$$

where

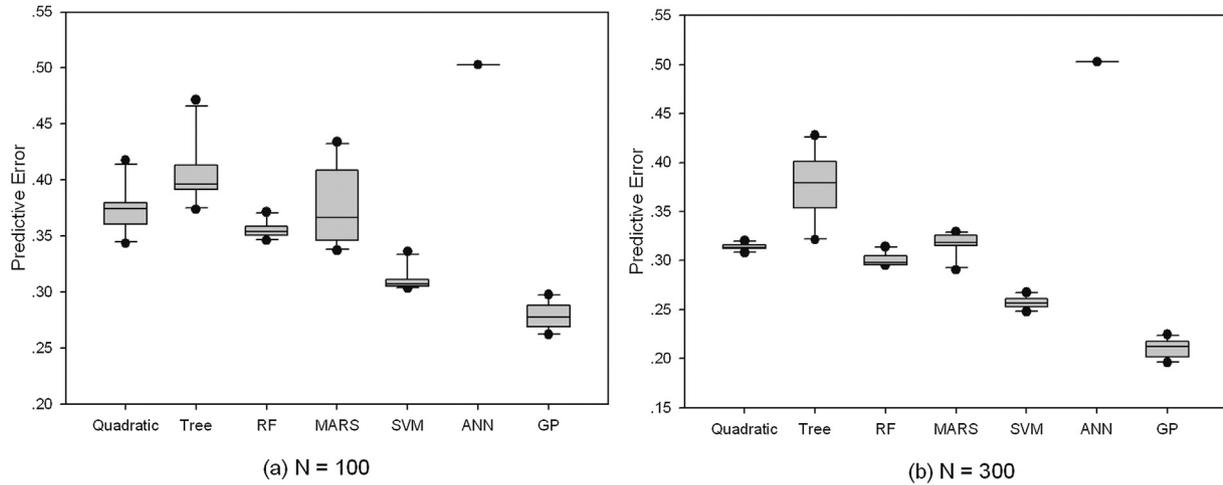


Fig. 2. The predictive errors of different surrogate model methods: (a) $N = 100$ and $K = 10$; and (b) $N = 300$ and $K = 10$. The center lines are the median error; the top and bottom lines of the gray boxes are the average \pm one standard deviation; the top and bottom lines are the maximum and minimum error of the K cross validation tests. The horizontal labels denote quadratic regression, regression tree, random forest, MARS, support vector machine, artificial neural network and Gaussian process regression, respectively.

$$A = \begin{pmatrix} 10 & 3 & 17 & 3.5 & 1.7 & 8 \\ 0.05 & 10 & 17 & 0.1 & 8 & 14 \\ 3 & 3.5 & 1.7 & 10 & 17 & 8 \\ 17 & 8 & 0.05 & 10 & 0.1 & 14 \end{pmatrix}, c = \begin{pmatrix} 1 & 1.2 & 3 & 3.2 \end{pmatrix}^T, p = \begin{pmatrix} 0.1312 & 0.1696 & 0.5569 & 0.0124 & 0.8283 & 0.5886 \\ 0.2329 & 0.4135 & 0.8307 & 0.3736 & 0.1004 & 0.9991 \\ 0.2348 & 0.1451 & 0.3522 & 0.2883 & 0.3047 & 0.6650 \\ 0.4047 & 0.8828 & 0.8732 & 0.5743 & 0.1091 & 0.0381 \end{pmatrix}$$

the six-dimensional Hartman function has six local minima. The ranges of the independent variables are $0 \leq x_j \leq 1$ where $j \in \{0, 1, \dots, 5\}$. The function value at the global minimum is $f(x^*) = -3.32237$, where

$$x^* = (0.20169, 0.150011, 0.476874, 0.275332, 0.311652, 0.6573)$$

is the global minimum point. In this study, we will try to find the optimum by optimizing the surrogate model for the Hartman function.

The specific objectives of this case study are: (a) to choose the best surrogate model construction method and adaptive sampling method for the ASMO method; (b) to determine the most appropriate initial DoE methods and initial sample sizes for the ASMO method; (c) to compare the optimization results of the ASMO method and the direct global optimization algorithm, the SCE-UA method.

We first evaluate which of the following surrogate model construction methods are the best for the ASMO method: (a) Quadratic (polynomial), (b) Tree, (c) RF, (d) MARS, (e) SVM, (f) ANN and (g) GP. The QMC Sobol' method (which is proven the best DoE method for ASMO later in this section) is used to sample N points in the feasible space of the Hartman function. The objective function values at those sample points are then computed. To check the performance of the surrogate model construction methods, we used the K -fold cross validation method, which works as follows. The entire sample set is divided into K subsets. Given a specific surrogate model construction method, a response surface is created by using only $K - 1$ subsets of sample points (i.e., by eliminating one subset). The difference (i.e., the predictive error) between the true response

surface and the response surface built on the $K - 1$ subsets of samples is computed. This procedure iterates K times to obtain the average predictive error. Fig. 2 shows the predictive errors of different surrogate model construction methods for $K = 10$ and $N = 100$ and $N = 300$, respectively. We can see that GP has the smallest predictive error and ANN has the highest error. Some methods show a significant variation among the K trials (e.g., regression tree and MARS), indicating more sample points are needed for a more accurate representation of the response surface. ANN shows no variation among the K trials for both $N = 100$ and $N = 300$, suggesting that it is not sensitive to the sample size. Based on this result, we choose the GP method as the best surrogate model method for further study.

Given GP as the best surrogate model method, we now evaluate whether the MIS or MEI method are the best adaptive sampling methods for the ASMO method. Again we use the QMC Sobol' design to generate $N = 100$ and $N = 300$ random samples, respectively, and use them as the initial sample sizes for the ASMO method, respectively. Fig. 3 shows the optimization convergence process of the two adaptive sampling methods. From Fig. 3, we note clearly that MIS performs better than MEI. For the MIS method, it converges to the true optimum (within <0.001 of the true optimum value of -3.32237) at 127 total sample points when $N = 100$ and at 328 when $N = 300$, suggesting that fewer than 40 additional sample points are needed in the adaptive sampling stage. For the MEI method, the optimization search converges at about 155 for $N = 100$ and about 428 for $N = 300$, suggesting 55–130 additional sample points in the adaptive sampling stage. All two experiments for MEI fail to find the optimum at convergence. Some previous findings have suggested that the MEI method is capable to finding the global minimum asymptotically (Locatelli, 1997). However, it may take a long time to do so (Jones, 2001). The MEI method underperforms the MIS method here probably due to the following reasons: 1) per MEI's algorithm, the adaptive sampling strategy tends to focus more on regions where there is a large predictive uncertainty instead of focusing on the more promising regions. When the initial sample size is small, the predictive uncertainty estimates of the GP response surfaces are large everywhere. This means that MEI needs to do a fairly exhaustive search in the entire feasible parameter space, demanding high computational time (Jones, 2001); 2) The predictive uncertainty of the surrogate model near the boundary of the feasible space tends to be larger than

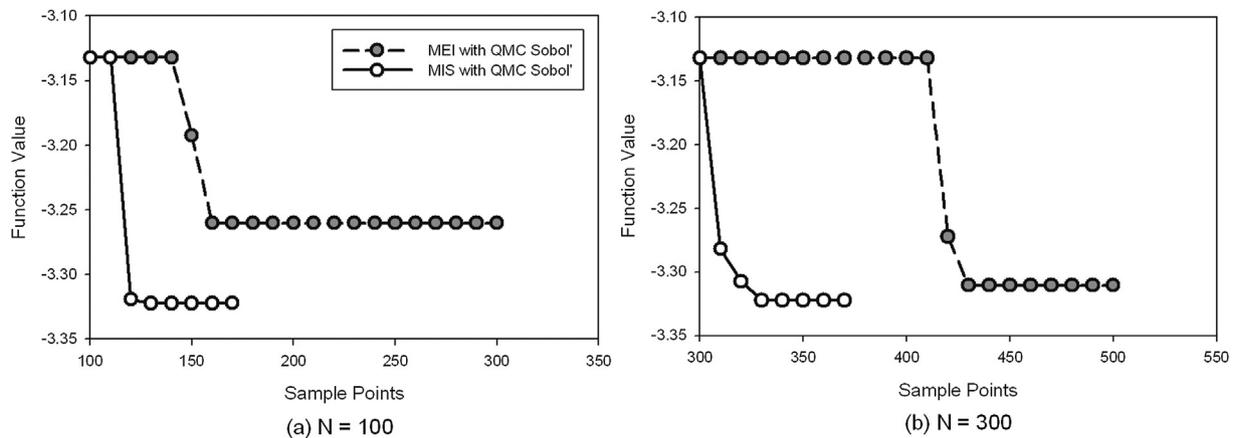


Fig. 3. Comparison of the convergence process by MEI and MIS adaptive sampling methods: (a) $N = 100$ and (b) $N = 300$.

other regions because that is where extrapolation is needed to fit the GP model. For this reason, MEI tends to put more points near the boundary of the feasible space in the initial search stage. A careful examination of the MEI search process confirms that it takes many iterations near the boundary before the search moves toward the optimum (see Fig. 3). Based on above results, we choose MIS adaptive sampling method over MEI for further evaluation of the ASMO method.

After we have chosen GP as the best surrogate model construction method and the MIS method as the best adaptive sampling method for ASMO, we now examine how different DoE methods and initial sample sizes influence the optimization results. Here two space filling DoE methods: LH and QMC are evaluated. Since LH is a random sampling approach, we repeated LH random sampling using different random seeds ten times (labeled with “LH_1”–“LH_10”). QMC method is a pseudo random approach, i.e., the sequences of samples look random, but they are actually deterministic, depending on the sequence generating method. In this study two QMC sequence generating methods are evaluated: QMC-Sobol’ and QMC-Halton approaches (see Appendix III for a brief description). Three initial sample sizes, N_{init} , are tested, i.e., $N_{init} = 50, 100$ and 150 . After the initial sample sets are generated, additional adaptive sampling is then conducted until one or more convergence criteria is reached. One convergence criterion is that the best objective function value is sufficiently close to the true optimal value (in this case, within 0.001 of true optimal value of -3.32237 for the Hartman function). Another convergence criterion is that the search reaches the maximum number of samples, N_{max} , allowed (in this case $N_{max} = 300$). Fig. 4(a)–(c) shows the optimization results for three DoEs and three initial sample sizes, respectively. With any of the three initial sample sizes, the optimization search using some of LH designs and two QMC designs can all converge to the optimum at different convergence rates. When $N_{init} = 50$, it takes QMC-Sobol’ more than 230 and QMC-Halton more than 180 total samples to find the optimum. When $N_{init} = 150$, it takes both QMC-Sobol’ and QMC-Halton about 165 total samples to converge to the optimum. On the other hand, it takes about 120 and 152 total samples for QMC-Sobol’ and QMC-Halton, respectively, to converge to the optimum when $N_{init} = 100$. The best results from the 10 LH designs are similar to or even surpass those of QMC designs. However, the search using LH designs fails to find the true optimum in numerous cases, suggesting that LH is not a robust DoE method for ASMO.

Through this study, we learn that the proper initial sample size is important. The convergence speed might be slow if the initial samples are too few and it would be too inefficient if the initial

sample size is too large. Based on this case study, an initial sample size of $N_{init} = 100$ seems to be the appropriate choice. In terms of choosing the best DoEs for the ASMO method, the two QMC designs are more reliable than the LH design.

Since the ASMO method has been demonstrated to be capable of finding the optimum of the Hartman function, we now compare the ASMO method with the SCE-UA optimization method. The objective is to demonstrate the advantage of the ASMO method over a traditional global optimization approach. The best performing setup for the ASMO method is used in this comparison, i.e., the GP method chosen as the surrogate model method, MIS as the adaptive sampling method, $N_{init} = 100$, and QMC Sobol’ as the DoE for initial sampling. The optimization results are shown in Fig. 5, which shows the convergence process and Fig. 6, which shows both the ASMO and the SCE-UA methods can find the exact true optimum in every parameter dimension. For the ASMO method, the optimization process converges to the optimal function value of <-3.322 after a total sample of 127 (with 100 initial samples and 27 additional adaptive samples). On the other hand, the SCE-UA method requires up to 880 sample points to reach the function value of <-3.322 , which is more than seven times of the ASMO method. This result indicates that ASMO has a clear advantage over SCE-UA in terms of optimization search efficiency at least for the Hartman function case. The probable reason is that SCE-UA is a direct optimization method that explores the entire feasible space. By using the information obtained from the sample points and mathematical approximations of the response surface of the original simulation model, the ASMO method does not try to emulate the entire response surface of the original simulation model. Rather it focuses on searching the region which is more likely to contain the optimum.

3.2. Case study II: calibration of the SAC-SMA model

3.2.1. Description of model, study area, datasets and objective function

In this case study, we evaluate the efficiency and effectiveness of the ASMO method using a hydrologic model used in real-world applications, the SAC-SMA model. The SAC-SMA model is a conceptual rainfall–runoff model that represents the soil column with upper and lower zones of multiple storages (Burnash, 1995). It has been used extensively in both research and operational applications for river forecasting by the National Weather Service River Forecast System (NWSRFS) Centers across the United States. It is also a benchmark problem for evaluation of many parameter estimation methods (Duan et al., 1992, 1994; Yapo et al., 1998; Ajami

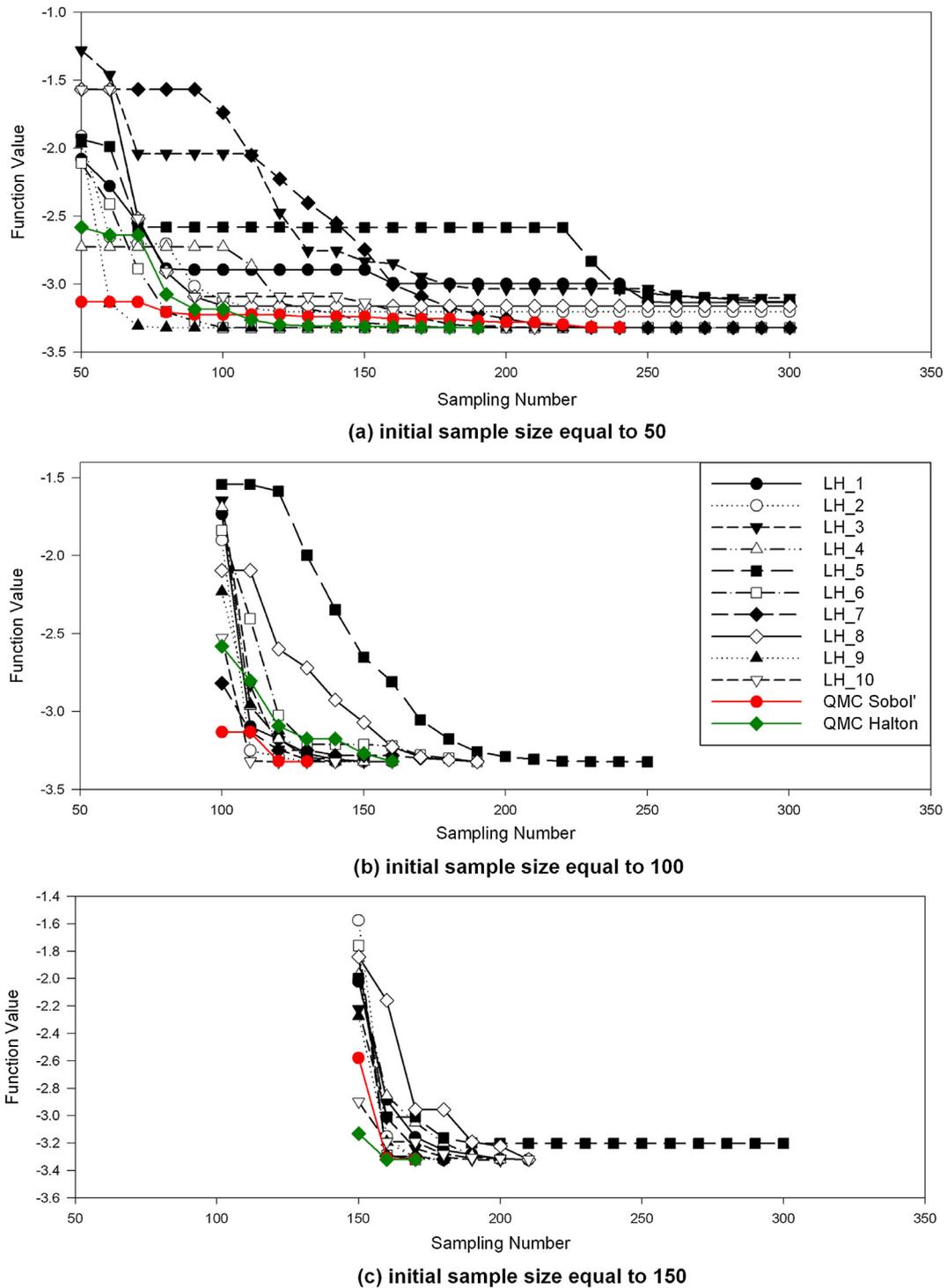


Fig. 4. Optimization results of the Six-Dimensional Hartman function by the AMSO method with different DoEs and initial sample sizes.

et al., 2007). Fig. 7 shows the structure of the SAC-SMA model. There are sixteen parameters in the SAC-SMA model. We consider only thirteen of them as adjustable parameters, whose feasible ranges and descriptions are listed in Table 1. Three parameters RSERV, RIVA, and SIDE are fixed at pre-specified values according to Brazil (1988).

The study area is the South Branch Potomac River basin near Springfield, West Virginia in the U.S. It is one of the 12 experimental watersheds of the Model Parameter Estimation Experiment

(MOPEX) (Duan et al., 2006). The total drainage area of the basin (U.S. Geological Survey Station No. 01608500) is about 3810 km². Historical precipitation, potential evapotranspiration and streamflow observations from January 1st, 1960 to December 31st, 1979 were obtained from the MOPEX database for this study (Duan et al., 2006). The annual average precipitation over this period is 1021 mm, annual average potential evapotranspiration is 762 mm, and annual average streamflow discharge is 39.5 m³/s yr. The hydrological simulations are run with a 6-h time step for each set of

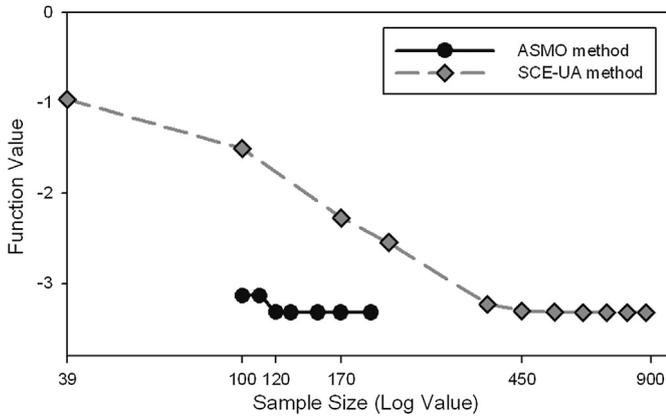


Fig. 5. Comparing the optimization results of the ASMO and SCE-UA methods on the Six-Dimensional Hartman function.

model parameters. Additional physical characteristics of the study area were presented by Duan et al. (2006).

The purpose of a model calibration (i.e., parameter optimization) is to find the optimal parameter set for the SAC-SMA model such that the simulated streamflow would have the best overall match with the observed streamflow through minimizing the objective function value. The observed streamflow data in this study was recreated synthetically according to the following procedure: (1) an artificial set of parameters was chosen as the “true” optimal parameters; (2) the SAC-SMA model was run using this set of “true” parameters, observed precipitation and potential evapotranspiration to obtain the streamflow simulation for the historical period 1960–1979. The simulated streamflow obtained using above procedure is treated as the “observed streamflow”. The advantage of using synthetic data is that we know the “true” parameter values and the “true” streamflow.

$$x^* = \{242.868, 49.5779, 0.4373, 0.011, 0.063, 97.7848, 1.8564, 325.192, 353.817, 61.679, 0.1092, 0.0131, 0.262\}$$

The artificially assigned “true” parameter set is In this study, we used the Root Mean Square Error (RMSE) as the objective function, which is defined as

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Q_{s,t} - Q_{o,t})^2} \quad (2)$$

where n is the number of time steps, $Q_{s,t}$ is the simulated flow for time step t , and $Q_{o,t}$ is the observed flow for time step t . In the model simulation process, we used a warm-up period of 3 months to remove any impact of uncertain initial conditions, i.e., in computing RMSE, the streamflow values for the first three months were removed from consideration. In hydrological modeling, the choice of objective function (i.e., performance metrics) can have a great impact on the identifiability of model parameters (Bennett et al., 2013). Since this is a synthetic case study with a known optimum, the use of RMSE as an objective function is sufficient.

3.2.2. The optimization results and analysis

The ASMO method described in Section 2 is used to optimize the parameters of the 13-dimensional SAC-SMA model. Based on the

findings from Section 3.1, we examined only the two QMC methods (i.e., the Sobol' and Halton sequences) as the possible choice for the initial DoE method. Different initial sample sizes at 100, 250 and 400 are experimented to find the best initial sample size. Fig. 8 displays the results. We note that initial sample size at 250 seems to produce relatively better objective function values than other sample sizes. However the advantage is not significant. For an initial sample size of 100, the ASMO search makes quick improvement initially, but settles at a local optimum which is only slightly worse than the optimum found with an initial sample size of 250. When the initial sample size is 400, the ASMO search also moves rapidly toward the optimum in the adaptive stage. However, the total number of samples is too high (>510) and is therefore regarded too wasteful. The optimization results seem to suggest that the initial sample does play a role in the effectiveness and efficiency of the ASMO search process. But what exact initial sample size should the ASMO method use seems to be not clear-cut. One obvious observation is that for a 13-dimensional optimization problem, no more than 20 times the dimension is adequate. The initial sample size for ASMO seems to be somewhat different from those of other researchers. For example, Söbester et al. (2005) suggested that about 35% of the samples should be spent in the initial stage, while Jones et al. (1998) recommended that the initial sample size should be about ten times the dimensionality of the problem.

Based on above results, we chose QMC Sobol' sequence as the initial DoE method and 250 as the initial sample, and compared the ASMO method with the SCE-UA method. The results in Fig. 9 reveal how the objective function values trend down with increasing sample points for both the ASMO and SCE-UA methods. This figure shows that ASMO has an obvious advantage in convergence speed over SCE-UA, with the former needing about 400 total sample points and the latter needing close to 1600 sample points to reach similar objective function value. The SCE-UA method possesses an edge over the surrogate based optimization in converging to the

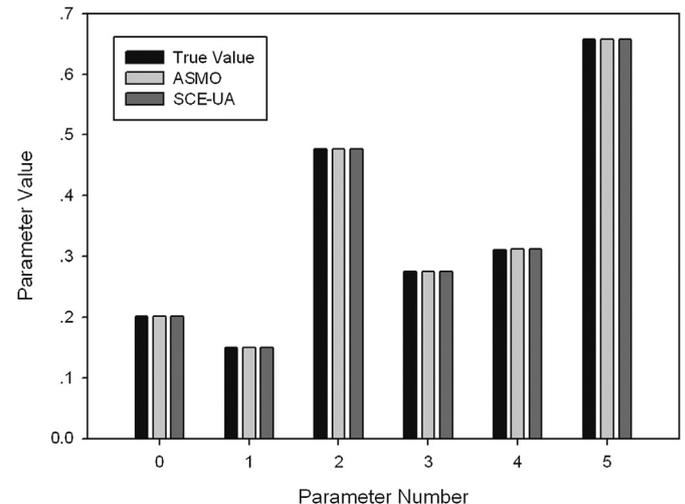


Fig. 6. Comparing the converged parameters of the six-dimensional Hartman function by the ASMO and SCE-UA methods.

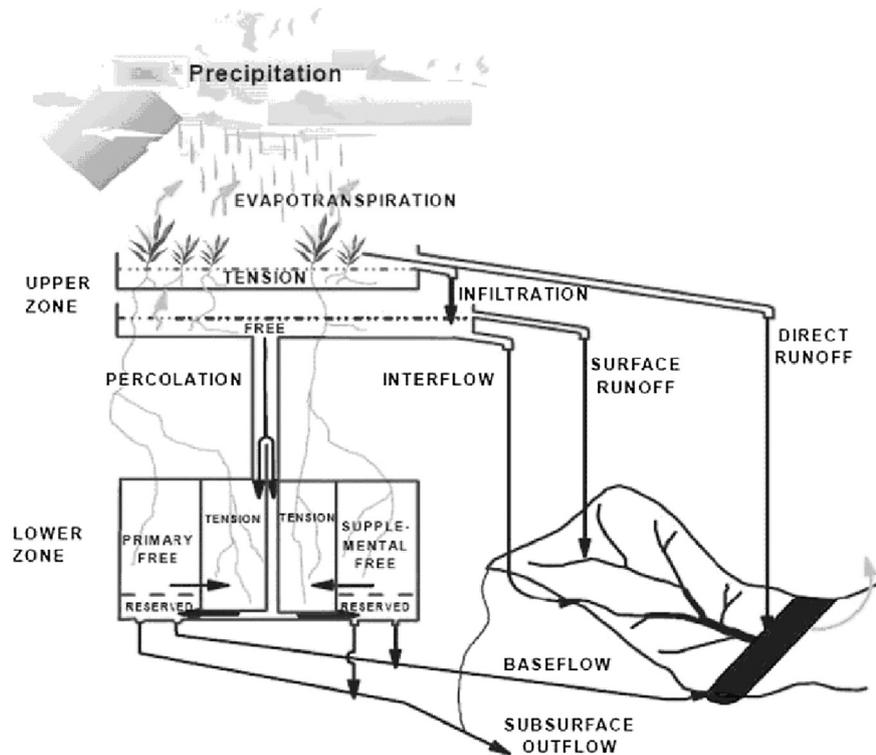


Fig. 7. A schematic of the SAC-SMA model (Ajami et al., 2007).

Table 1

The thirteen parameters of the SAC-SMA model and their feasible ranges (Brazil, 1988).

No	Parameter	Description	Lower bound	Upper bound
1	UZTWM	Maximum capacity of the upper zone tension water storage (mm)	10.00	300.00
2	UZFWM	Maximum capacity of the upper zone free water storage (mm)	5.00	150.00
3	UZK	Lateral drainage rate of upper zone free water (day^{-1})	0.10	0.75
4	PCTIM	Impervious fraction of the watershed area (decimal fraction)	0.00	0.10
5	ADIMP	Additional impervious area (decimal fraction)	0.00	0.20
6	ZPERC	Proportional increase in percolation from saturated to dry condition (dimensionless)	5.00	350.00
7	REXP	Exponent of the percolation equation (dimensionless)	1.00	5.00
8	LZTWM	Maximum capacity of lower zone tension water storage (mm)	10.00	500.00
9	LZFSM	Maximum capacity of lower zone supplemental free water storage (mm)	5.00	400.00
10	LZFPM	Maximum capacity of lower zone primary free water storage (mm)	10.00	1000.0
11	LZSK	Lateral drainage rate of lower zone supplemental free water storage (day^{-1})	0.01	0.35
12	LZPK	Lateral drainage rate of lower zone primary free water storage (day^{-1})	0.001	0.05
13	PFREE	Fraction of the percolated water which is transmitted directly to the lower zone free water aquifers during the period when the lower zone tension water has a soil moisture deficiency (decimal fraction)	0.00	0.80

The three fixed parameter value is: RSERV = 0.3; RIVA = 0.0; SIDE = 0.0.

“true” optimal parameter set if there is no limit on the number of sample points. In other words, the SCE-UA method is capable of finding the exact “true” parameter values, while the ASMO method can only provide approximate optimization results. Additional samples points cannot make ASMO to find the exact “true” optimal values, as illustrated in Fig. 9, which shows that the optimization results would not improve after certain iterations.

3.2.3. The influence of parameter sensitivity on optimization results

From Section 3.2.2, we found that the ASMO method is able of identifying an approximate optimal parameter set, but not the “true” optimum. In theory, when synthetic data were used, we should be able to locate the “true” parameter set if an optimization method works effectively. In this section, we investigate how close the approximate optimal parameters are to the “true” parameters and why they deviate from their “true” parameters.

Fig. 10 shows the optimal parameters obtained by the ASMO and SCE-UA methods. Also shown is the “true” parameter set. This figure shows that some parameters were identified precisely, while others were not. To find out the reasons, we conducted a global sensitivity analysis of the 13 parameters to see if there is any relationship between parameter identifiability and parameter sensitivity. We used the Sobol’ variance decomposition method to evaluate the sensitivity of the parameters (Saltelli et al., 2000). The Sobol’ method (Sobol’, 1993, 2001) provides a quantitative evaluation of the contribution of each parameter to the total variance of the model outputs, including main effect (i.e., the effect of varying one parameter alone), two-factor interaction effect (i.e., the effect of varying two parameters jointly), and total effect (including all the possible synergetic terms between that parameter and all the others) (Saltelli and Annoni, 2010).

Fig. 11 shows the main effects and the total effects of different parameters. The main effects of parameter 1(UZTWM), 4(PCTIM), 5(ADIMP), 8(LZTWM) and 12(LZPK) are relatively high and are thus

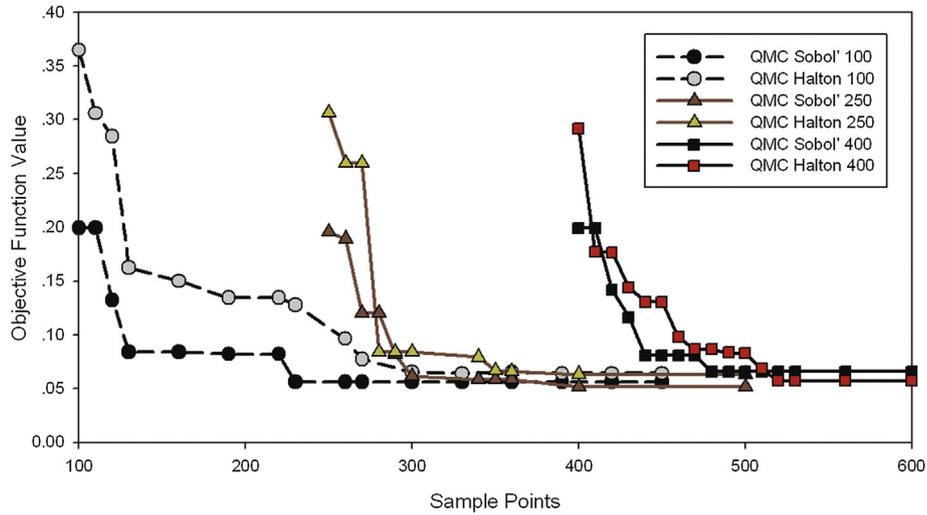


Fig. 8. Optimization results of the SAC-SMA model by the ASMO method using different initial DoE methods and different initial sample sizes.

regarded as highly sensitive. Parameters 2(UZFWM), 3(UZK), 6(ZPERC), 7(REXP), 9(LZFSM) are relatively insensitive parameters based on their main effects. The optimized parameter values for parameters 1, 4, 5, 8 and 13 are close to the “true” values, while the optimized parameter values for parameters 2, 3, 6, 7, 9 are far away from the “true” values. Interestingly, the closeness of the optimized parameters to the “true” values seems to be proportional to the main effects of the parameters. This finding is similar to the one by van Wekhoven et al. (2009). However, van Wekhoven used total effects to assess parameter sensitivity. This observation is particularly obvious for the results from the SCE-UA method, in which all highly sensitive parameters are identified precisely, while the insensitive parameters can be far away from the “true” values (i.e., parameters 2, 3, 6, 7 and 9). The optimized parameter values from the ASMO method deviate from the “true” parameter values more than those from the SCE-UA method at about 1000 samples.

There are a number of reasons for the deviations of the optimized parameters from the “true” values. The main reason is that, for insensitive parameters (parameters 2, 3, 6 and 7), their values are not identifiable because they have little influence on the model

output. There is another group of parameters, which have small main effects, but high total effects (i.e., parameters 3, 9, 10 and 11), consistent with the findings by Wu and Hamada (2009). That means they have strong interaction effect that can be estimated by the difference between the total and the main effects. We suspect that the strong interaction among these parameters resulted in optimization convergence to the wrong parameter values. This is consistent with the findings by other researchers. According to the finding by Pokhrel et al. (2008), some SAC-SMA parameters have strong inter-parameter correlations: e.g., LZSK (parameter 11) co-varies with UZK (parameter 3), LZFP (parameter 10) co-varies with PFREE (parameter 13) and REXP (parameter 7), LZPK (parameter 12) co-varies with LZTWM (parameter 8). The correlations among these parameters make the identification of these parameters very difficult.

4. Conclusions

To provide the practitioners with a clear perspective on how to use the surrogate modeling based optimization properly, we

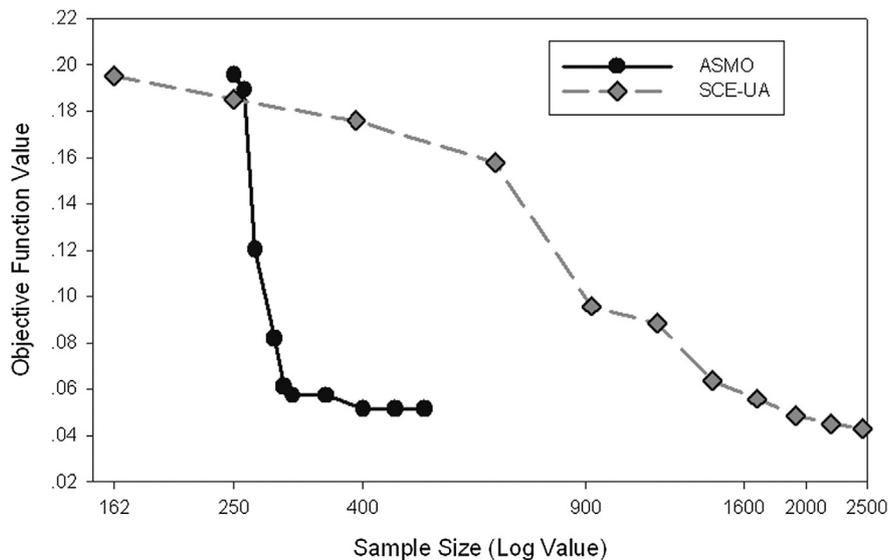


Fig. 9. Optimized objective function values and sampling sizes of the SAC-SMA model by the ASMO and SCE-UA methods.

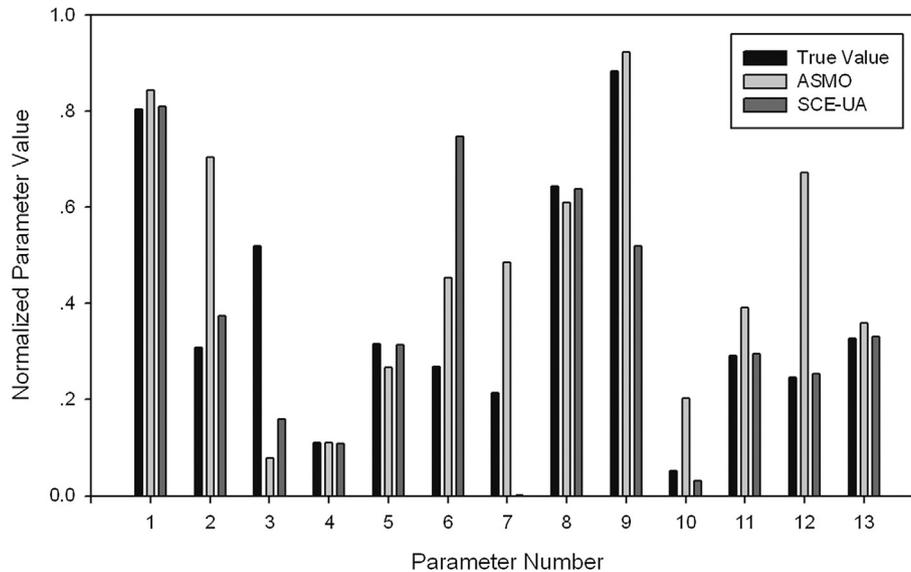


Fig. 10. Comparison of optimal parameter sets of the SAC-SMA model obtained by the ASMO and SCE-UA methods with the “true” parameter set.

conducted a detailed analysis of an ASMO method with respect to the selection of the most appropriate surrogate model construction method, the adaptive sampling method, the initial DoE methods and initial sample sizes. To sum up, we compared seven different surrogate model construction methods and two adaptive sampling methods. Then we explored how different initial DoE methods and different initial sample sizes affect optimization search results in two case studies. The main findings are:

- 1) Generally speaking, the ASMO method is efficient and effective in finding the approximate optimal parameters of the simulation model. It can acquire good optimization results with fewer sample points than the SCE-UA method;
- 2) Surrogate model built using the GP method produces consistently better optimization results than other methods and the MIS method is clearly more efficient than the MEI method as the adaptive sampling strategy. Initial sampling methods can affect the optimization results of surrogate

modeling based optimization. The low discrepancy DoE methods such as QMC-Sobol' and QMC-Halton are better than the LH method. Proper sampling size can make the method more efficient and effective. In our study, we found that the proper initial sample size should not exceed 20 times the dimension of the problem;

- 3) The ASMO method is designed to use a cheaper surrogate to provide approximate solutions to a simulation model. It cannot be expected to provide a precise solution as the direct optimization method such as the SCE-UA method. The precision of the surrogates is limited by many factors, including the initial DoE, initial sample size and the complexity of the original problem;
- 4) The precision of the optimized parameters has a close relationship with the sensitivity indices of the parameters. Highly sensitive parameters have a higher possibility of being identified correctly. Parameter interaction, however, may complicate the identifiability of the parameters, resulting in incorrect parameter estimates.

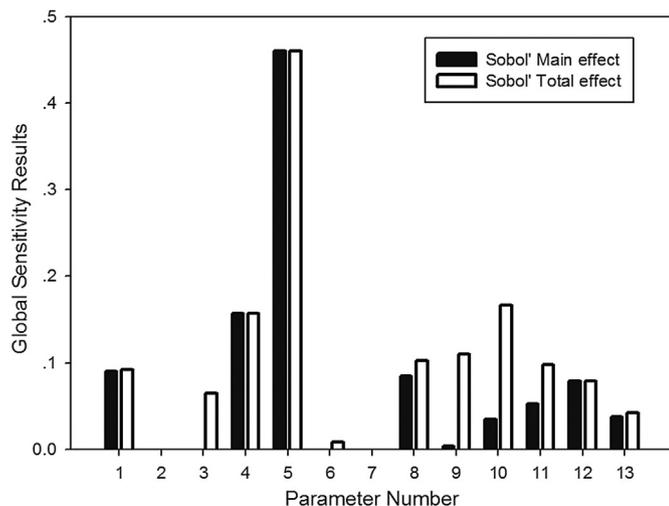


Fig. 11. Sobol' global sensitivity analysis results showing the main effect (Black) and the total effect (White) of different parameters in the SAC-SMA model.

The results from this study are based on two typical case studies, in which we know the exact optimal solutions. In real world applications, the model structure may be in error and the data used in estimating the model parameters may be noisy. Therefore there is no such thing as a “true” solution. In this case, we may have to find ways to account for these uncertainties by developing better models and formulating better objective functions that represent the model errors better. Since this study focuses on whether we can find the approximate solution with limited computer runs, we believe that most of the findings on the proper choice of DoE methods and initial sample sizes should help the research community in dealing with their specific problems. Further research, however, is necessary to validate if the findings from this study can be generalized to other cases.

Acknowledgment

This research is supported by the Ministry of Science and Technology of the People's Republic of China National Science and Technology Support Program (Grant No. 2013BAB05B04) and the Natural Science Foundation of China (No. 41375139).

Appendix I. A brief description of Latin hypercube, quasi Monte Carlo Sobol' and quasi Monte Carlo Halton designs

Latin hypercube design

Latin hypercube (LH) sampling is a statistical method for generating a distribution of plausible collections of parameter values from a multi-dimensional distribution. This method was first described by McKay et al. (1979) and further elaborated by Iman et al. (1981) and Hernandez et al. (2012). In the context of statistical sampling, a square grid containing sample points is named a Latin Square if there is only one sample in each row and each column. A Latin hypercube is a sampling design suited for an arbitrary number of dimensions, where each sample is the only one in each axis-aligned hyper-plane containing it. When sampling M points from a function of N variables, the range of each variable is divided into M equally probable intervals; within each interval, random values are selected according to the probability density function (PDF) of each parameter.

QMC Sobol' design and Halton design

A QMC method can be viewed as a deterministic version of Monte Carlo method (Niederreiter, 1992). It also called a low-discrepancy procedure. A low-discrepancy sequence is a set of s -dimensional points, filling the sample area "efficiently" and has a low discrepancy than straight pseudo-random number set (Krykova, 2003). The exact location of the sample points is determined by the base number, the construction method and the sample size. One popular QMC sequence is called Sobol' sequence, which was first introduced by Russian mathematician Sobol' (1967). For a s -dimensional problem, Sobol' sequence uses different permutations of base-two van der Corput sequences (van der Corput, 1935) for all dimensions to form successively finer uniform partitions of the unit interval, and then reorder the coordinates in each dimension. Halton sequence (Halton, 1964) is another QMC sequence in multiple dimensions. The Halton sequence is a general s -dimensional sequence in the unit hypercube $[0,1]^s$. It is generated using a different prime number as the base for van der Corput sequence in each of the s dimension (Krykova, 2003).

Appendix II. A brief description of seven surrogate model construction methods

Polynomial regression

Polynomial regression models are parametric models that represent input–output relationships through polynomials. Training data will be fitted using a least-square approach. Since polynomial regression models are global methods, a perturbation in any sample data will change all parts of the model. Thus, the presence of outliers can have a disastrous effect on the model itself. Additionally, when the simulation model is sufficient complex, a simple polynomial regression model may not be adequate to approximation the response surface of the simulation model. Usually we use linear, quadratic and cubic polynomial models.

Regression tree

Tree method is a non-parametric supervised learning method used to predict a response from several inputs. It can conveniently be used for Bayesian supervised learning, such as regression and classification. Whereas the outputs for classification are discrete class labels, regression is concerned with the prediction of

continuous quantities (Rasmussen and Williams, 2006). Regression tree method is done by growing a binary tree. At each internal node in the tree, a test is applied to one of the inputs. Depending on the outcome of the test, one can go to either the left or the right sub-branch of the tree. Eventually we come to a leaf node where a prediction is made. However, when the data has lots of features which interact in complicated, nonlinear ways, assembling a single global model can be very difficult. Regression tree method chooses the sub-divide approach or partition to solve this problem. It splits the space into smaller regions, where the interactions are more manageable. It then partitions the sub-divisions again, this is called recursive partition.

Multi-variate adaptive regression splines

Multi-variate Adaptive Regression Splines (MARS) is a procedure for fitting adaptive non-linear regression that uses piecewise basis functions to define relationships between a response variable and some set of predictors (Friedman, 1991). MARS is a non-parametric regression technique and can be seen as an extension of decision tree method. It uses a class of pairwise spline basis function for each input. The general form of the MARS model can be represented by the following expression:

$$\hat{y} = \beta_0 + \sum_{j=1}^M \beta_j B_j(\vec{x}) \quad (A1)$$

where $\vec{x} = (x_1, x_2, \dots, x_p)$ is the vector of inputs, B_j is the j th basis function, which can be a single spline function or a product of two or more basis function, and the coefficients β_j s are estimated by minimizing the sum of squared residuals (Shahsavani et al., 2010). MARS uses a forward-backward procedure to construct the final model. The forward step is very similar to stepwise regression but uses the basis functions (Shahsavani et al., 2010). Generalized Cross Validation (GCV) (Golub et al., 1979) evaluates the performance of the model. MARS chooses the best model with less GCV.

Random forests

Random forests (Breiman, 2001) are one of ensemble methods. The goal of ensemble methods is to combine the predictions of several models built with a given learning algorithm in order to improve generalizability and robustness over a single model. Random forests are learning ensemble consisting of a bagging of un-pruned decision tree learners with a randomized selection of features at each split. The random forests algorithm is as follows: 1) provide n_{tree} bootstrap samples from the original data; 2) for each of the bootstrap samples, grow a regression tree with the following modification: at each node, rather than choosing the best split among all predictors, randomly sample m_{try} of the predictors and choose the best split from among those variables; 3) predict new data by aggregating the predictions of the n_{tree} trees (Liau and Wiener, 2002).

Support vector machines

Support Vector Machines (SVMs) are learning machines implementing the structural risk minimization inductive principle to obtain good generalization on a limited number of learning patterns (Basak et al., 2007). SVM has been first introduced by Vapnik (1995). There are two main categories for support vector machines: support vector classification (SVC) and support vector regression (SVR). We use SVR for constructing surrogate models. The idea of SVR is based on the computation of a linear regression

function in a high dimensional feature space where the input data are mapped via a nonlinear function. Due to the possible high dimensionality of the input variable, SVR usually solves the dual problem instead of the original problem. After the dual problem is solved, using the primal–dual relationship to acquire the final expression.

Artificial neural network

The basic premise of ANN is that biological systems have extraordinary learning ability when exposed to external signals. With proper training of an ANN, it has been used for regression analysis. The complexity of real neurons is highly abstracted when modeling artificial neurons. These basically consist of inputs, which are multiplied by weights, and then computed by a mathematical function which determines the activation of the neuron. Another function computes the output of the artificial neuron. ANNs combine artificial neurons in order to process information.

Gaussian process regression

GP is an interpolating regression method that uses basis functions with tuned parameters to represent the original model $f(x)$ (Jones, 2001). GP regression model is a collection of random variables, any finite number of which has a joint Gaussian distribution (Rasmussen and Williams, 2006). A GP is completely specified by its mean function $m(\vec{x})$ and covariance function $k(\vec{x}, \vec{x}')$.

$$\begin{aligned} m(\vec{x}) &= E[f(\vec{x})], \\ k(\vec{x}, \vec{x}') &= E\left[(f(\vec{x}) - m(\vec{x}))\left(f(\vec{x}') - m(\vec{x}')\right)\right] \end{aligned} \quad (A2)$$

and a GP can be written as

$$f(\vec{x}) \sim \text{GP}\left(m(\vec{x}), k(\vec{x}, \vec{x}')\right) \quad (A3)$$

A GP is shown to be the predictor that minimizes the expected squared prediction error subject to: (1) being unbiased and; (2) being a linear function combination of observations. GP also has a statistical interpretation, which not only allows us to compute a predictor, but also allows us to compute a measure of the possible error in the predictor.

Appendix III. A brief description of two adaptive sampling methods

Minimum interpolating surface method

In Minimum Interpolating Surface (MIS) method, a surrogate modeling method is first fitted to initial sample points. Then the minimum of the response surface is found and is treated as a new sample point to be added to the existing sample set. The simulation model is run at this point to obtain the model performance measure. If the new sample set meets the convergence criterion, the search stops; otherwise, we begin another loop of adaptive sampling.

Maximum expected improvement method

The Maximum Expected Improvement (MEI) method needs to compute how much improvement we expect to achieve if we sample at a given point. Let $Y(\vec{x})$ be a random variable describing the uncertainty about the function's value at a point \vec{x} . If the

current best function value is f_{\min} , then we will achieve an improvement of I if $Y(\vec{x}) = f_{\min} - I$. The likelihood of achieving this improvement is given by the normal density function:

$$\frac{1}{\sqrt{2\pi}s(\vec{x})} \exp\left[-\frac{(f_{\min} - I - \hat{y}(\vec{x}))^2}{2s^2(\vec{x})}\right] \quad (A4)$$

where $y(x)$ is the mean and $s^2(x)$ is the variance given by the kriging predictor. The expected improvement can be integrated over this density function:

$$E(I) = \int_{I=0}^{I=\infty} I \left\{ \frac{1}{\sqrt{2\pi}s(\vec{x})} \exp\left[-\frac{(f_{\min} - I - \hat{y}(\vec{x}))^2}{2s^2(\vec{x})}\right] \right\} dI \quad (A5)$$

Using integration by parts, one can show that

$$E(I) = s(\vec{x})[\Phi(u) + \phi(u)] \quad (A6)$$

where

$$u = \frac{f_{\min} - \hat{y}(\vec{x})}{s(\vec{x})}$$

Φ and ϕ are the normal cumulative distribution function and density function, respectively (Jones, 2001).

References

- Ajami, N.K., Duan, Q., Sorooshian, S., 2007. An integrated hydrologic Bayesian multi-model combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resour. Res.* 43, W01403. <http://dx.doi.org/10.1029/2005WR004745>.
- Basak, D., Pal, S., Patranabis, D., 2007. Support vector regression. *Neural Inform. Process. Lett. Rev.* 11 (10).
- Behzadian, K., Kapelan, Z., Savic, D., Ardeshtir, A., 2009. Stochastic sampling design using a multi-objective genetic algorithm and adaptive neural networks. *Environ. Model. Softw.* 24 (4), 530–541.
- Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. *Environ. Model. Softw.* 40, 1–20.
- Blanning, R.W., 1975. Construction and implementation of metamodells. *Simulation* 24 (6), 177–184.
- Box, G., Behnken, D., 1960. Some new three level designs for the study of quantitative variables. *Technometrics* 2, 455–475.
- Box, G.E., Hunter, J.S., Hunter, W.G., 2005. *Statistics for Experimenters: Design, Innovation, and Discovery*, second ed. Wiley, ISBN 0-471-71813-0.
- Boyle, D.P., Gupta, H.V., Sorooshian, S., 2000. Toward improved calibration of hydrologic models: combining the strengths of manual and automatic methods. *Water Resour. Res.* 36 (12), 3663–3674.
- Brazil, L.E., 1988. *Multilevel Calibration Strategy for Complex Hydrologic Simulation Models*. Colorado State University, Fort Collins, CO.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. *Classification and Regression Trees*. Wadsworth, Belmont, Calif.
- Breiman, L., 2001. Random forests. *Machine Learn.* 45 (1), 5–32. <http://dx.doi.org/10.1023/A:1010933404324>.
- Burnash, R.J.C., 1995. The NWS river forecast system: catchment modeling. In: Singh, V.P. (Ed.), *Computer Models of Watershed Hydrology*. Water Resources Publications, Highlands Ranch, Colorado, pp. 311–366.
- Carnevale, C., Finzi, G., Guariso, G., Pisoni, E., Volta, M., 2012. Surrogate models to compute optimal air quality planning policies at a regional scale. *Environ. Model. Softw.* 34, 44–50.
- Castelletti, A., Pianosi, F., Soncini-Sessa, R., Antenucci, J.P., 2010. A multiobjective response surface approach for improved water quality planning in lakes and reservoirs. *Water Resour. Res.* 46.
- Castelletti, A., Galelli, S., Ratto, M., Soncini-Sessa, R., Young, P.C., 2012. A general framework for dynamic emulation modelling in environmental problems. *Environ. Model. Softw.* 34, 5–18.

- Cox, D.D., John, S., 1997. SDO: a statistical method for global optimization. In: Alexandrov, N., Hussaini, M.Y. (Eds.), *Multidisciplinary Design Optimization: State of the Art*. SIAM, Philadelphia, pp. 315–329.
- Duan, Q., Sorooshian, S., Gupta, H.V., 1992. Effective and efficient global optimization for conceptual rainfall–runoff models. *Water Resour. Res.*
- Duan, Q., Gupta, V.K., Sorooshian, S., 1993. Shuffled complex evolution approach for effective and efficient global minimization. *J. Optim. Theory Appl.* 76 (3), 501–521. <http://dx.doi.org/10.1007/BF00939380>.
- Duan, Q., Sorooshian, S., Gupta, V.K., 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. *J. Hydrol.*
- Duan, Q., Schaake, J., Andreassian, V., Franks, S., Goteti, G., Gupta, H.V., Gusev, Y.M., Habets, F., Hall, A., Hay, L., Hogue, T., Huang, M., Leavesley, G., Liang, X., Nasonova, O.N., Noilhan, J., Oudin, L., Sorooshian, S., Wagener, T., Wood, E.F., 2006. Model parameter estimation experiment (MOPEX): an overview of science strategy and major results from the second and third workshops. *J. Hydrol.* 320 (1–2), 3–17.
- Fen, C.S., Chan, C.C., Cheng, H.C., 2009. Assessing a response surface-based optimization approach for soil vapor extraction system design. *J. Water Resour. Plann. Manag.* 135 (3), 198–207.
- Forrester, A.I.J., Keane, A.J., 2009. Recent advances in surrogate-based optimization. *Progr. Aerosp. Sci.* 45 (1–3), 50–79.
- Friedman, J.H., 1991. Multivariate adaptive regression splines. *Ann. Statist.* 19 (1), 1–67.
- Gan, T.Y., Biftu, G.F., 1996. Automatic calibration of conceptual rainfall–runoff models: optimization algorithms, catchment conditions, and model structure. *Water Resour. Res.* 32 (12), 3513–3524.
- Golub, G.H., Heath, Michael, Wahba, Grace, 1979. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*. <http://dx.doi.org/10.1080/00401706.1979.10489751>.
- Halton, J., 1964. Algorithm 247: radical-inverse quasi-random point sequence. *ACM*, 701. <http://dx.doi.org/10.1145/355588.365104>.
- Hernandez, A.S., Lucas, T.W., Sanchez, P.J., 2012. Selecting random Latin hypercube dimensions and designs through estimation of maximum absolute pairwise correlation. In: *Proceedings of the 2012 Winter Simulation Conference*. <http://informatics-sim.org/wsc12papers/includes/files/con441.pdf>.
- Hogue, T.S., Sorooshian, S., Gupta, H., Holz, A., Braatz, D., 2000. A multistep automatic calibration scheme for river forecasting models. *J. Hydrometeorol.* 1 (6), 524–542.
- Iman, R.L., Helton, J.C., Campbell, J.E., 1981. An approach to sensitivity analysis of computer models, part I. Introduction, input variable selection and preliminary variable assessment, 13 (3), 174–183.
- Jones, D., 2001. A taxonomy of global optimization methods based on response surfaces. *J. Global Optim.* 21, 345–383.
- Jones, D., Schonlau, M., Welch, W., 1998. Efficient global optimization of expensive black-box functions. *J. Global Optim.* 13 (4), 455–492.
- Krykova, I., 2003. Evaluating of path-dependent securities with low discrepancy methods. In: *Financial Mathematics*. Worcester Polytechnic Institute.
- Liaw, A., Wiener, M., December 2002. Classification and regression by random forest. *R News* 2/3, 18–22. URL: http://cran.r-project.org/doc/Rnews/Rnews_2002-3.pdf.
- Locatelli, M., 1997. Bayesian algorithms for one-dimensional global optimization. *J. Global Optim.* 10, 57–76.
- Loeppky, J.L., Sacks, J., Welch, W., 2009. Choosing the sample size of a computer experiment: a practical guide. *Technometrics* 51, 366–376 which assumes Latin Hypercube Sampling (LHS) as the Design Of Experiments (DOE) method, and recommends a sample size of ten times the number of dimensions.
- McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*.
- Moreno, H.A., Vivoni, E.R., Gochis, D.J., 2012. Utility of quantitative precipitation estimates for high resolution hydrologic forecasts in mountain watersheds of the Colorado front range. *J. Hydrol.* 438–439, 66–83. <http://dx.doi.org/10.1016/j.jhydrol.2012.03.019>.
- Mousavi, S.J., Shourian, M., 2010. Adaptive sequentially space-filling meta-modeling applied in optimal water quantity allocation at basin scale. *Water Resour. Res.* 46 (3).
- Myers, R.H., 1971. *Response Surface Methodology*. Allyn and Bacon, Inc., Boston.
- Neelin, J.D., Bracco, A., Luo, H., McWilliams, J.C., Meyerson, J.E., 2010. Considerations for parameter optimization and sensitivity in climate models. *Proc. Natl. Acad. Sci. U. S. A.* 107 (50), 21349–21354.
- Niederreiter, H., 1992. Random number generation and quasi-Monte Carlo methods. In: *CBMS-NSF, vol. 63*. SIAM, Philadelphia.
- O'Hagan, A., 2006. Bayesian analysis of computer code outputs: a tutorial. *Reliab. Eng. Syst. Saf.* 91 (10–11), 1290–1300.
- Picheny, V., Ginsbourger, D., Richet, Y., Caplin, G., 2013. Quantile-based optimization of noisy computer experiments with tunable precision, including comments and rejoinder. *Technometrics* 55 (1), 1–36.
- Pokhrel, P., Gupta, H.V., Wagener, T., 2008. A spatial regularization approach to parameter estimation for a distributed watershed model. *Water Resour. Res.* 44, W12419. <http://dx.doi.org/10.1029/2007WR006615>.
- Queipo, N.V., Haftka, R.T., Shyy, W., Goel, T., Vaidyanathan, R., Kevin Tucker, P., 2005. Surrogate-based analysis and optimization. *Progr. Aerosp. Sci.* 41 (1), 1–28.
- Rasmussen, C.E., Williams, C.K.I., 2006. *Gaussian Processes for Machine Learning*. MIT Press, ISBN 026218253X.
- Ratto, M., Castelletti, A., Pagano, A., 2012. Emulation techniques for the reduction and sensitivity analysis of complex environmental models. *Environ. Model. Softw.* 34, 1–4.
- Razavi, S., Tolson, B.A., Burn, D.H., 2012a. Numerical assessment of metamodelling strategies in computationally intensive optimization. *Environ. Model. Softw.* 34, 67–86.
- Razavi, S., Tolson, B.A., Burn, D.H., 2012b. Review of surrogate modeling in water resources. *Water Resour. Res.* 48 (7).
- Saltelli, A., Chan, K., Scott, E.M., 2000. *Sensitivity Analysis*. Wiley, New York.
- Saltelli, A., Annoni, P., 2010. How to avoid a perfunctory sensitivity analysis. *Environ. Model. Softw.* 25 (12), 1508–1517. <http://dx.doi.org/10.1016/j.envsoft.2010.04.012>.
- Schonlau, M., Welch, W.J., Jones, D.R., December 1997. Global versus local search in constrained optimization of computer models. In: *Flournoy, N., Rosenberger, W.F., Wong, W.K. (Eds.), New Developments and Applications in Experimental Design*. Institute of Mathematical Statistics. Also available as Technical Report RR-97-11, Institute for Improvement in Quality and Productivity, University of Waterloo, Waterloo, Ontario, Canada.
- Shahsavani, D., Tarantola, S., Ratto, M., 2010. Evaluation of MARS modeling technique for sensitivity analysis of model output. *Proc. Soc. Behavior. Sci.* 2 (6), 7737–7738.
- Shi, W., Kleijnen, J.P.C., Liu, Z., 2012. Factor Screening for Simulation with Multiple Responses: Sequential Bifurcation (Center Discussion Paper, 2012-032). <http://arno.uvt.nl/show.cgi?fid=122462>.
- Söbester, A., Leary, S., Keane, A., 2005. On the design of optimization strategies based on global response surface approximation models. *J. Global Optim.* 33 (1), 31–59.
- Sobol', I.M., 1967. On the distribution of points in a cube and the approximate evaluation of integrals. *Zh. Vychisl. Mat. Mat. Fiz.* 7 (4), 784–802.
- Sobol', I.M., 1993. Sensitivity analysis for nonlinear mathematical models. *Math. Model. Comput. Exp.* 1, 407–414.
- Sobol', I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* 55 (1–3), 271–280. [http://dx.doi.org/10.1016/S0378-4754\(00\)00270-6](http://dx.doi.org/10.1016/S0378-4754(00)00270-6).
- Sorooshian, S., Duan, Q.Y., Gupta, V.K., 1993. Calibration of rainfall–runoff models – application of global optimization to the Sacramento soil moisture accounting model. *Water Resour. Res.* 29 (4), 1185–1194.
- Stuckman, B.E., 1988. A global search method for optimizing nonlinear systems. *IEEE Trans. Syst. Man Cybernet.* 18, 965–977.
- van der Corput, J.C., 1935. Verteilungsfunktionen I–VIII. *Proc. Ned. Akad. Wet.* 38, 813–821, 1058–1066; 39 (1936), 10–19, 19–26, 149–153, 339–344, 489–494, 579–590.
- 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. Res.* 32 (6), 1154–1169. <http://dx.doi.org/10.1016/j.advwatres.2009.03.002>.
- Vapnik, V., 1995. *The Nature of Statistical Learning Theory*. Springer.
- Wu, C.F.J., Hamada, M., 2009. *Experiments: Planning, Analysis, and Optimization*, second ed. Wiley, New York.
- Yapo, P.O., Gupta, H.V., Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. *J. Hydrol.* 204, 83–97.
- 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.