Journal of Environmental Informatics XX(X) XX-XX (XXXX)

Journal of Environmental Informatics

www.iseis.org/jei

Assessment of Parametric Sensitivity Analysis Methods Based on A Quasi Two-Dimensional Groundwater Model

Z. H. Di[1]*, A. Z. Ye[1], Q. Y. Duan[1], and X. X. Wang[1]

[1]State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

Received 20 November 2017; revised 05 November 2018; accepted 21 January 2019; published online 21 June 2019

ABSTRACT. Parametric sensitivity analysis (SA) aims to select the sensitive parameters that most significantly affect the model output variables, which helps to improve model optimization efficiency by adjusting a small number of sensitive parameters instead of all adjustable parameters. The qualitative and quantitative SA methods have been commonly used to quantify the sensitive parameters of the models. However, the response surface model based quantitative SA method was rarely used. Taking the simulation of a quasi two-dimensional (quasi-2D) groundwater model as an example, this study systematically assess eight SA methods divided into three categories (qualitative SA, quantitative SA, and the response surface model-based quantitative SA). The study validates the effectiveness of these methods by comparing the parameter sensitivity results, and also demonstrates the efficiency of these methods by determining the minimum sample size required. Using the minimum samples means the least number of model runs. The results show that P1 and P2 are the most sensitive parameters of the quasi-2D model for simulating groundwater table elevation. Except for local method, four global qualitative SA methods obtain reasonable parameter sensitivity rankings using 200 samples, but the parameter sensitivity scores fail. For obtaining accurate sensitivity scores, at least 2000 samples are required by the quantitative SA methods. However, for the response surface model-based quantitative SA method, 60 samples are sufficient to obtain accurate sensitivity scores, demonstrating that the method is an effective and highly efficient, and should be recommended as the primary parametric SA method, especially for the complex models with large computational demand.

Keywords: parametric sensitivity analysis, quasi-2D groundwater model, response surface model, Tarim River

1. Introduction

Improving the simulation ability of groundwater models for water table elevation and soil water content is often achieved by tuning model parameter values. Common parameter estimation methods include the a priori inference method and the parameter calibration method (Duan et al., 2003). The a priori inference method estimates model parameter values using a look-up table that reflects the relationship between parameter values and geographical features (e.g., topography, soil texture, and vegetation cover). The parameter calibration method, also called the inverse problem-solving method, searches for the optimal parameter values that make the simulated results as close as possible to the corresponding observed values by repeatedly tuning the parameters within their variation ranges. Obviously, the latter method can obtain better parameter values to improve model simulation results.

However, parameter calibration methods usually require tens of thousands of model runs to determine the optimal values of the adjustable parameters, where the number of model runs is usually exponentially related to the number of parameters to be optimized. A large number of model runs makes parameter optimization difficult to complete, especially for the models requiring an extremely large computational amount. These are usually classified as (1) models incorporating many physical processes (e.g., run off, evaporation, and groundwater movement etc.). Examples include hydrological models such as the Soil and Water Assessment Tool model (Arnold et al., 1993) and the Variable Infiltration Capacity model (Liang et al., 1994); and (2) models incorporating complex resolution algorithms that require high run-times. The Richards' equation-based soil water model (Richards, 1931) is one such example. The Richards' equation is a nonlinear partial differential equation, and therefore its best solution is obtained using a numerical algorithm; however, numerical algorithms are computationally expensive because they usually require the simulation domain to be split into many interconnected grid cells. The differential equation is then approximated as an equilibrium-type algebraic equation for each grid cell. Because the algebraic equations for the adjacent grids are interactive, all equations (equal to the number of grids) must be solved simultaneously, requiring a high computer memory capacity. This results in high computational cost, especially for domain with more split grids.

* Corresponding author. Tel.: +86 010-58800217; fax: +86 010-58800217.
E-mail address: zhdi@bnu.edu.cn (Z. H. Di).

ISSN: 1726-2135 print/1684-8799 online
 doi:10.3808/jei.201900413

For the above two types of high-cost model, optimization efficiency is very low if all parameters are to be calibrated simultaneously. Many insensitive parameters are optimized, but do not obviously change the simulated results however they are tuned. To enhance optimization efficiency, a small number of the most sensitive parameters should be optimized, rather than all the parameters. Variation in sensitive parameters exerts a significant effect on the simulated results, whereas variation in insensitive parameters does not. Therefore, how to distinguish the sensitive parameters in the set of all adjustable parameters is a critical problem for parametric sensitivity analysis (SA).

Parametric SA mainly investigates how variations in model output can be attributed to variations in the input parameters. According to their different scopes of action, SA can be divided into local SA and global SA. Local SA quantifies only the effect of a single parameter perturbation on model output variation with other parameters unchanged (Griewank and Walther, 2000). However, global SA perturbs all parameters simultaneously (Homma and Saltelli, 1996).

Depending on the computation requirements for model runs, global SA methods are divided into qualitative and quantitative SA methods. Among the qualitative methods, there are three major categories. The first category includes qualitative methods based on the gradient principle, such as the Newton iteration method (Ortega and Rheinboldt, 1970) and the Morris one-at-a-time method (Morris, 1991). The second category includes qualitative methods based on the nearest-neighbor principle, such as the delta test (DT) method (Eirola et al., 2008). The third category comprises qualitative methods based on a regression response surface, such as the multivariate adaptive regression spline (MARS) method (Steinberg et al., 1999), the sum-of-trees (SOT) method (Breiman et al., 1984), and the Gaussian process (GP) method (Rasmussen and Nickisch, 2010). Quantitative SA methods measure the contribution of parameter perturbations to variation in model output based on variance decomposition, which is evidently different from qualitative sensitivity methods. Therefore, quantitative SA methods can assess not only the effect of each parameter on variation in model output, but also the effect of interactions between parameters on model output variation. The McKay method (McKay et al., 1998) and the Sobol' method (Sobol', 2001) are representative of quantitative methods.

The response surface model-based quantitative SA method is another type of SA method. It differs from the common quantitative SA method in that it applies the quantitative SA method on a statistical response surface model rather than a complex physical model. The response surface model is evaluated more rapidly than the original physical model. Common regression methods used to build a response surface model include multiple linear regression (Navarra and Simoncini, 2010), support vector machines (Cortes and Vapnik, 1995), and neural network methods (Bhadeshia, 1999). Note also that the reasonableness of the response surface model should be verified before implementing a quantitative SA on it.

Many SA methods have been used in various fields. For food safety models, Patil and Frey (2004) applied 10 SA methods to a draft Vibrio parahaemolyticus food risk assessment model for obtaining sensitive inputs and finally found that the mutual information index, scatter plot, and analysis of variance methods were more robust than the others. Confalonieri et al. (2010) performed a number of SA methods on a crop model of rice growth to obtain sensitivity rankings of crop parameters. For water quality models, Neumann (2012) combined multiple SA methods and various objective functions for a micropollutant degradation model to obtain more robust parameter sensitivity results. Sun et al. (2012) demonstrated that the regional SA method could provide parameter interactions and hence is more appropriate for complex water quality models. For hydrological and land-surface models, Collins and Avissar (1994) screened out the important sensible heat and latent heat parameters for a land-atmosphere interaction model using the Fourier amplitude sensitivity test SA method. Hou et al. (2012) evaluated the importance of the parameters in the Community Land Model to the hydrologic output variables using a generalized linear model. These studies have demonstrated that these SA methods are appropriate for performing parametric SA experiments. However, the efficiency of the SA methods was not assessed.

Although some studies (e.g., Li et al., 2013, Gan et al., 2014 and Di et al., 2017 in our previous work) referred to the efficiency of SA methods, they were incomplete and unsystematic. The main deficiencies include the following:

(1) Gan et al. (2014) assessed only qualitative and quantitative SA methods, but the response surface model-based quantitative SA method was not reviewed. Additionally, low-uniformity sampling methods were not applied to the SA experiments, reducing the efficiency of SA methods.

(2) Li et al. (2013) and Di et al. (2017) assessed only the effectiveness and efficiency of qualitative SA methods using the results of the response surface model-based quantitative SA method with enough sample points. The efficiency of the quantitative SA method and the response surface model-based quantitative SA method were not assessed. Moreover, the representativeness of the response surface model with respect to the original physical model was not evaluated before conducting the quantitative SA method based on a response model.

(3) The previous work referred only analyzed the characteristics of qualitative and quantitative SA methods, and did not select a highly efficient SA method for the parameter SA of the complex models with expensive computation cost, due to lack of comprehensive comparison on efficiency of the three types of SA methods, especially the efficiency assessment of the response surface model based quantitative SA method.

In response to these deficiencies, this study plans to assess systematically the three categories of SA methods from effecttiveness to efficiency using more uniform samples. Certainly, how to build the best response surface model should be also discussed. In the present study, a quasi-2D groundwater model is used to evaluate the effectiveness and efficiency of parametric SA methods. The reasons for this choice are the following:

(1) The quasi-2D groundwater model fully considers the movement variations of soil water and groundwater, and especially their interaction. The model is a relatively perfect groundwater model, and its practicability has been proved in arid area (Di et al., 2011).

(2) Each run of the model consists of two nonlinear partial differential equations and a relation equation that entails high computation cost due to the numerical algorithm required. The algorithm gives an accurate solution to the nonlinear partial differential equations, but it involves the simultaneous solution of a large number of equations (equal to the number of domain grid cells). This requires large amounts of computer memory, thus increasing computation cost for domains containing more grids. Because the model is nonlinear, the iterative algorithm is also needed in addition to the numerical algorithm, slowing the speed of solution even further. For parameter optimization in models with high computation cost, it is very necessary to conduct the SA on the model to filter in advance the small number of sensitive parameters to be optimized.

(3) The results of sensitive parameters and their optimal values for the model change for the different climate regions being simulated. Therefore, when the model is applied to a large region (e.g., the whole of China), the experiments on the parameter SA and optimization of sensitive model parameters should be separately conducted for the different climatic regions. In such a case, using a highly efficient parametric SA method for the model greatly reduces the number of model runs required to identify the sensitive parameters, and further improves the efficiency of model parameter optimization. This approach allows the implementation of the model over a large area.

Taking as an example a simulation of groundwater table elevation in the Yingsu section of the lower reaches of the Tarim River in China under stream water transfer conditions, the present study assesses the effectiveness of the three types of SA methods, including the eight SA methods by comparing the parameter sensitivity results, and evaluates the efficiency of these SA methods by determining the minimum sample size required. The aim is to find an effective and highly efficient SA method for parameter SA experiments on high-cost computation models. This paper is organized as follows. Section 2 introduces the methodology, including two sampling methods, eight SA methods, and SA experiment design. Section 3 presents the assessment results of three types of SA methods, from effectiveness to efficiency. Conclusions and discussions are provided in the last section.

2. Materials and Methods

An integral parametric SA procedure consists of parameter sampling, model evaluation, and SA. Parameter sampling produces randomly perturb parameter values in an adjustable parameter space.

The perturbed parameter values are then put into the physical model by replacing the corresponding default values, and the models are then run with perturbed parameter values to obtain the corresponding model outputs or simulation errors.

Finally, the sensitive parameters are selected by the SA method based on the sample points consisting of the input perturbed parameter values and corresponding model outputs.

2.1. Sampling Methods

Parameter sampling, also called parameter perturbation, produces several random parameter values (i.e., parameter samples) over the parameter range. A good sampling method should produce evenly distributed samples that effectively fill the parameter space using the fewest possible samples. According to previous studies on comparison of sampling methods (Wang et al., 2014; Gong et al., 2016), the quasi Monte Carlo (QMC) sampling method (Caflisch, 1998) is considered as a uniform sampling method that could provide better space-filling capability than the Monte Carlo (MC) or Latin hypercube sampling methods (McKay et al., 2000). Therefore, the QMC uniform sampling method is used as a primary sampling method in this study. Note also that Sobol' SA method only prescribes the corresponding Sobol' sampling method rather than other sampling methods, which means that Sobol' SA does not produce parametric sensitivity results from the QMC samples. Therefore, both sampling methods are used in this study.

2.1.1. QMC Sampling Method

The QMC sampling method is a deterministic version of the MC method. It produces a low-discrepancy sequence that approximates the integral of a function with minimum error. The sample locations from the sequence are usually determined by base number, construction method, and specified sample size (Caflisch, 1998). There are many ways to build the low discrepancy sequence of QMC samples, such as the Halton, Faure, and Niederreiter sequences (Halton, 1960; Faure, 1982; Nieder-reiter, 1988). Here, we select the Halton sequence widely acknowledged as the result of QMC sampling method. The Halton sequence for the n-dimensional parameters is constructed as follows: (1) The first n primes are firstly chosen as the n bases; (2) An integer m randomly chosen is then represented as a numerical digit with a prime basis in each dimensional parameter space, and the different dimensionalities use the different prime bases; (3) For each dimensional parameter, the number of numerical digit is arranged in reverse order, and the decimal value is then obtained by adding a decimal point in front of the numerical digit; (4) Let m equal to $m + 1$, repeating the previous steps until the whole sequence is obtained. The fixed QMC algorithm produces the fixed sampling sequence, which is obviously different from the regular MC method that produces a pseudorandom sequence. For approximating a function integral, the pseudorandom samples from the MC sampling method produce larger errors than the low-discrepancy samples from the QMC, demonstrating the uniformity of the QMC samples.

2.1.2. Sobol' Sampling Method

The Sobol' sampling method is designed to meet the perturbation analysis requirements of the Sobol' SA method. It starts with two random $r \times k$ sample matrices M_0 and M_{k+1}, where r is the number of repeats and k is the dimensionality of the input parameters. For each of the r parameter arrays from the M_0 matrix, the i^{th} ($i \in [1, k]$) point is generated from both matrices based on the rule that the i^{th} column is replaced by the same column of M_{k+1}, but the other columns are the same as M_0. The i^{th} column of M_{k+1} is also replaced in the same way. Finally, by adding the two original matrices M_0 and M_{k+1}, the total number of sample points is equal to $2 \times (k + 1) \times r$.

2.2. SA Methods

Eight representative SA methods are chosen to conduct parameter SA for the quasi-2D groundwater model in this study. These methods represent the three types of SA methods: qualitative and quantitative SA methods, and the response surface model-based quantitative SA method. For the qualitative SA methods, one common gradient method is defined as the local qualitative SA method, and four global qualitative SA methods including DT, SOT, MARS, and GP methods are chosen. As representatives of quantitative SA methods, the McKay method and the Sobol' method are selected. In addition, the MARS response surface model-based Sobol' method (RS_ Sobol') is used to evaluate the effect of the response surface model-based quantitative SA method on the parameter SA results. These methods are briefly described in the Appendices.

2.3. SA Experimental Design

2.3.1. Model Description

A sketch map of lateral groundwater flow in the river bank section is shown in Figure 1. Assuming that the river bank is perpendicular to the impermeable layer, a rectangular coordinate system is constructed with the river bank as the z-axis, the impermeable layer as the x-axis, and their crossing point as the coordinate origin. The positive directions of the x- and z-axes are designated as rightward and upward, respectively. The simulated domain of the right-hand section is $H \times L$, and the groundwater table (Γ_1) divides the domain into two subdomains of unsaturated soil water and saturated groundwater. In general, only the right-hand domain is considered when constructing the quasi-2D groundwater model because of the symmetry of the river banks.

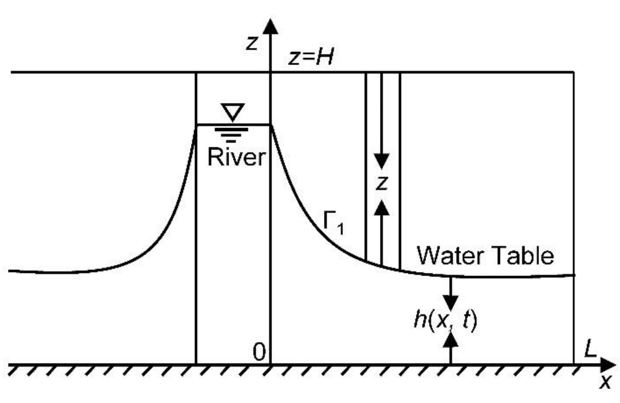

Figure 1. Schematic representation of groundwater lateral flow.

For the unsaturated soil water subdomain, conduction of soil water flow occurs mainly in the vertical direction due to the effect of gravity. For saturated groundwater, the flow is slow, and the water potentials in the vertical direction are almost equal; therefore, vertical flow is ignored, and horizontal flow is assumed. Finally, the simulated domain in the right-hand section is divided into n vertical soil columns defined as $d_1, d_2, \ldots, d_n$. For each column $x \in d_i$, the vertical unsaturated soil water flow is described as:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[D(\theta)\frac{\partial \theta}{\partial Z}\right] + \frac{\partial K(\theta)}{\partial z}, \quad h(x,t) < z < H \tag{1}$$

where θ is the soil water content; h (x, t) is the groundwater table elevation for soil column d_i at time t; H is the ground surface elevation; $K(\theta)$ is the hydraulic conductivity of unsaturated soil; and $D(\theta)$ is the hydraulic diffusivity of unsaturated soil. According to the Clapp and Hornberger formulation (Clapp and Hornberger, 1978), $K(\theta)$ and $D(\theta)$ are respectively parameterized as:

$$K(\theta) = K_{sm}\left(\frac{\theta}{\theta_s}\right)^{2b+3} \text{ and } D(\theta) = \frac{-bK_{sm}\psi_s}{\theta_s}\left(\frac{\theta}{\theta_s}\right)^{b+2} \tag{2}$$

where K_{sm} is the hydraulic conductivity of the saturated soil column in the vertical direction; θ_s is the saturated soil water content; ψ_s is saturated matric potential; and b is the slope of the retention curve. Because Equation (1) is a nonlinear partial differential equation, boundary conditions must be designated to solve it. Zero flux and saturated soil water content are designated as the upper and lower boundary conditions, respectively.

The continuity equation for saturated groundwater flow in the whole simulated domain in the right-hand section $H \times L$ can be written as:

$$n_e \frac{\partial h}{\partial t} = \frac{\partial}{\partial x}\left(K_s h \frac{\partial h}{\partial x}\right) - q_{z=h(x,t)}(x,t), \; 0 < x < L \tag{3}$$

where h (x, t) is the groundwater table elevation; K_s is the horizontal groundwater hydraulic conductivity; n_e is the specific yield of the soil; and $q_{z=h(x,t)}(x, t)$ is the exchange flux between the unsaturated and saturated zones at the groundwater table. The left-hand boundary condition of Equation (3) is the time-varying river stage, and the right-hand boundary condition is zero groundwater conveyance flux.

Combined with Darcy's law, the two sides of Equation (1) are integrated from the groundwater table elevation h (x, t) to the ground surface height H along the vertical z-direction. The exchange flux $q_{z=h(x,t)}(x, t)$ are obtained:

$$q_{z=h(x,t)}(x,t) = \int_{h(x,t)}^{H} \frac{\partial \theta}{\partial t} dz + q_{z=H}(x,t)$$
(4)

where $q_{z=H}(x, t)$ is the infiltration (evaporation) flux on the ground surface. Finally, Equations (1), (3), (4), and the corresponding initial and boundary conditions are made up the quasi -2D groundwater model. The concept and numerical solution of the quasi-2D groundwater model are proposed by Di et al. (2011), who used the model to simulate variations in the groundwater table elevations in the Yingsu section for stream water conveyance in the lower reaches of the Tarim River. The simulated results have demonstrated the reasonableness of the developed model through comparisons between simulation results and observed data. In this study, the parameter sensitivities of the quasi-2D groundwater model are analyzed and various SA methods are assessed using the same experiments as Di et al. (2011).

2.3.2. Study Area

The Yingsu section is located between the Daxihaizi Reservoir and Taitema Lake in the lower reaches of the Tarim River (see Figure 2), where the annual precipitation is 17.1 ~ 42.0 mm (Chen et al., 2010). Due to low precipitation and unreasonable usage of water resources in the upper and middle reaches, the lower reaches of the Tarim River have dried up, the ground-water table is continuously decreasing, and most of the natural vegetation along the river bank has died out (Feng et al., 2001). To solve this problem, a water conveyance project in the lower reaches of the Tarim River has been implemented by the Chinese government since May 2000. Seven water releases took place from Daxihaizi Reservoir to Taitema Lake from 2000 to 2005. For each water release, data on river discharges, river elevations, and the groundwater table for nine sections along the river bank are recorded.

The Yingsu section is the third of nine sections and is 60 km away from the Daxihaizi Reservoir. In the Yingsu section, there are seven groundwater-monitoring wells recording detailed groundwater table elevations, of which four wells (C3 ~ C6) are selected for comparison with the simulated groundwater table. The distances between the river bank and the four wells are 150, 300, 500, and 750 m respectively. Data for river discharges and river water elevations is recorded once every day, but data for groundwater table elevations from the four wells is recorded once every five days or every month. The duration of the simulation experiment (i.e., the second water release) in this study is 81 days (i.e., from 16 November, 2000 to 4 February, 2001). The objective function used to evaluate simulated groundwater table elevation is the mean absolute error

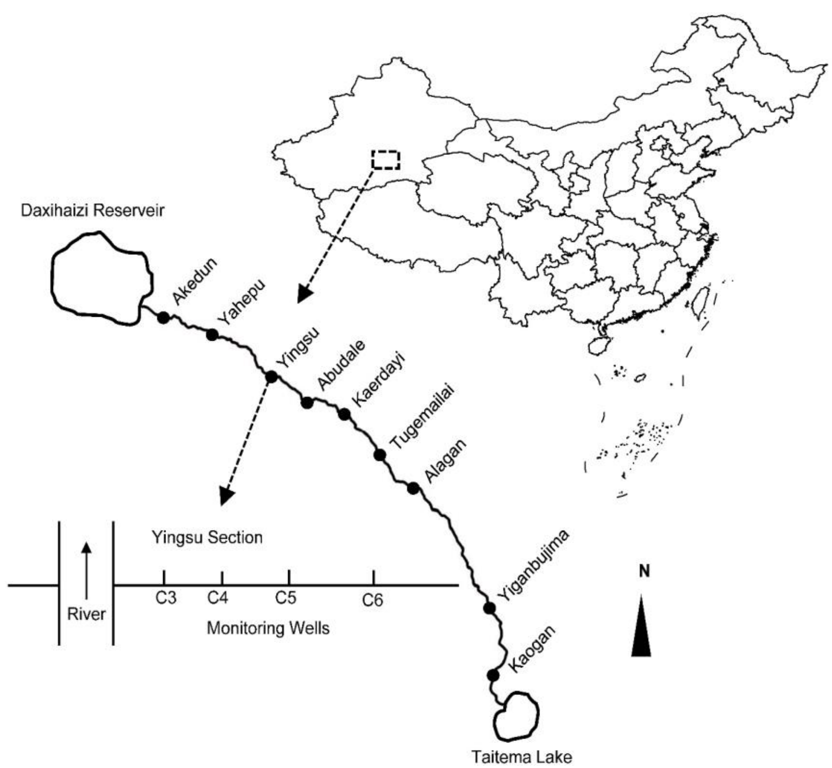

Figure 2. Location of nine sections between the Daxihaizi Reservoir and Taitema Lake in the lower reaches of the Tarim River and the monitoring wells (C3–C6) in the Yingsu section.

(MAE) between the simulated results and the observed data for the four wells. The MAE formula is expressed as follows:

$$\mathrm{MAE} = \frac{1}{MT}\sum_{t=1}^{T}\sum_{i=1}^{M}\left|sim_i^t - obs_i^t\right| \tag{5}$$

where sim_i^t and obs_i^t are simulated and observed groundwater table elevations at the i^{th} observation well and time t; M is the number of observation wells (i.e., equal to 4); and T is the total number of days with observation records. In this study, there are 12 days observation records from 16 November, 2000 to 4 February, 2001, and the observation frequency is usually once every five days or every month.

2.3.3. Adjustable Parameters

The parameter n_e (specific yield of soil) is defined as the change of the volume of water held in the moisture profile above the groundwater table for a groundwater table rise or fall of unit depth. The value of n_e depends on the distribution of pores, their shape, and grain size. The method used to estimate it has been documented by Taheri Tizro et al. (2012) and Durand et al. (2017). The hydraulic conductivity is a measure of the capacity of a soil to transmit water. It is divided into saturated and unsaturated hydraulic conductivities. The value of the unsaturated hydraulic conductivity varies with variation of soil water content, and thus it is not classified as a parameter; however, the saturated hydraulic conductivity, which describes groundwater movement through a saturated medium, has a constant value for certain type of soil (Jabro, 1992). In order to obtain better simulation results of soil water content and groundwater table level, the saturated hydraulic conductivity is given different values in the vertical and horizontal directions (i.e., horizontal groundwater hydraulic conductivity K_s, and vertical saturated hydraulic conductivity K_{sm}).

Saturated soil water content θ_s is the ratio of water volume content filling the soil pores to the total soil volume, so it is equivalent to porosity. The determination of the saturated soil water content is important in agricultural and ecological applications (Alharthi and Lange, 1987). ψ_s (saturated matric potential) and b (slope of the retention curve) are considered as soil parameters describing the hydraulic properties and are derived by fitting a power function to the moisture retention data (Cosby et al., 1984; Chen and Dudhia, 2001).

In accordance with the 12 global soil classification in the Biosphere-atmosphere Transfer Scheme (BATS) model (Dickinson et al., 1986) and the latitude and longitude of the study area, the properties of the sixth category (sandy clay loam) are: θ_s = 0.48, ψ_s = −200 mm, K_{sm} = 0.54 m/d, and b = 6.0. In addition, n_e and K_s are assigned 0.25 and 2.5 m/d respectively according to the work of Di et al. (2011). Both the look-up table and parameter calibration values show obvious uncertainties in the parameter values. The ranges of θ_s, ψ_s, K_{sm} and b are obtained from the values for the adjacent (i.e., fifth to seventh) soil texture in the BATS table. Because no guidance is available, the maximum values for the parameters of n_e and K_s are taken to be double their default values, as suggested for parameter ranges by Yang et al. (2012). (Note that the saturated matric potential ψ_s has the same value (−200 mm) for the adjacent soil texture in the BATS table, and therefore the parameter ψ_s is not adjusted). The ranges of the adjustable parameters are given in Table 1.

2.4. Experimental Design

Table 1. Model Parameters and Value Ranges

Number	Name	Default	Range	Description
P1	n_e	0.25	0.15–0.5	Specific yield of soil
P2	K_s	2.5	0.5–5.0	Horizontal groundwater hydraulic conductivity (m/d)
P3	K_{sm}	0.54	0.39–0.77	Hydraulic conductivity of the saturated soil column in the vertical direction (m/d)
P4	θ_s	0.48	0.45–0.51	Saturated soil water content
P5	b	6.0	5.5–6.8	Slope of retention curve

Table 2. Experimental Design for the Various Sensitivity Analysis (SA) Experiments

SA categories	SA method	Sampling method	Sample size
Qualitative SA	Local	QMC(Caflisch,1998)	10
	DT(Eirola et al., 2008)	QMC	50/200/500
	SOT(Breiman et al., 1984)	QMC	50/200/500
	MARS(Steinberg et al., 1999)	QMC	50/200/500
	GP(Rasmussen and Nickisch, 2010)	QMC	50/200/500
Quantitative SA	McKay(McKay et al., 1998)	QMC	2000
	Sobol’(Sobol’, 2001)	Sobol’(Sobol’, 2001)	1200/2400/3600
Response surface model based quantitative SA	RS_Sobol’ (Storlie and Helton, 2008)	Sobol’	60
		QMC	60/2000

To obtain more reasonable SA results, the three types of parameter SA methods are applied to the quasi-2D groundwater model with five adjustable parameters to analyze parametric sensitivity to the simulation results for groundwater table elevations. According to the characteristic of the various SA methods, the different sample sizes are used. The experimental design used are shown in Table 2. The local method analyzes only the gradients between the initial and terminal points of each independent variable one at a time, and hence 10 sample points are determined by five adjustable parameters. For the global qualitative SA methods, the uniform sampling method is QMC, and the uniform sample size is designed as 500. The smaller sample size is used to verify the efficiency of the qualitative SA methods. For the two quantitative SA methods, more samples are used to obtain reasonable parameter SA results. Therefore, 2,000 samples taken by the QMC sampling method and 2,400 samples obtained by the Sobol' sampling method are applied to the McKay and Sobol' quantitative SA methods, respectively. Note that the Sobol' sampling method is required only for the Sobol' SA method, and therefore the sample size is chosen as 2 × (5 + 1) × 200 (i.e., equal to 2,400). In addition, 1,200 and 3,600 samples are used to evaluate the efficiency of the two quantitative SA methods.

For the response surface model-based quantitative SA, the RS_Sobol' SA method is used. Both the QMC and Sobol' sampling methods are suitable for the RS_Sobol' SA method. The reason for this is that the sample points obtained by the two sampling methods are responsible only for constructing the MARS response surface model, after which Sobol' sampling and the corresponding Sobol' SA are automatically conducted on the MARS response surface model. Here, the initial sample size for the RS_Sobol' SA methods is chosen as 60 to demonstrate the advantage of the MARS response surface model for quantitative SA. However, the number of Sobol' samples on the MARS response surface is chosen as 12,000 to obtain a more accurate quantitative SA score from the response surface model. Note also that the accuracy of the MARS response surface model as a replacement for the original physical model may have a significant impact on RS_Sobol' SA results, especially for the specific parameter sensitivity scores. Therefore, it is essential to assess the reasonableness of MARS response surface model before implementing a quantitative SA on it.

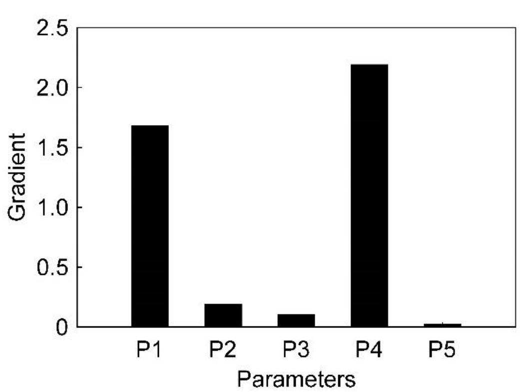

Figure 3. Parametric sensitivity obtained by the local sensitivity analysis (SA) method.

3. Results

3.1. Qualitative Parameter Screening Using Local SA Method

The sensitivity of the local SA method is measured by solving the gradient of an individual parameter between its initial and end points, while keeping other parameters at their default values. The parameter sensitivities of the quasi-2D groundwater model using the local SA method are shown in Figure 3. The ordinate represents the gradient of a single parameter between its initial and end points. It is found that the sensitivity ranking order of the five parameters using the local method is P4, P1, P2, P3, and P5. The degrees of sensitivity of P4 (saturated soil water content) and P1 (specific yield of soil) are evidently higher than those of the other parameters. However, the reasonableness of the results needs to be verified further using other SA methods.

3.2. Qualitative Parameter Screening Using the Four Global SA Methods

Besides the local qualitative SA method, global qualitative SA methods DT, SOT, MARS, and GP are applied in parameter SA experiments on the quasi-2D groundwater model. Multiple qualitative SA methods are used with the intention not only of obtaining more reasonable parameter sensitivity rankings, but also of evaluating these SA methods. For the four global qualitative SA experiments, the uniform sampling method is QMC, and the initial uniform sample size is set to be 500. To illustrate the parameter sensitivities more clearly, the sensitivity scores of the five parameters for each SA method are normalized by dividing the sensitivity score of each parameter by the maximum sensitivity score of all parameters. Therefore, the normalized score for the most sensitive parameter is set to 100, and the normalized score for the least sensitive parameter is set to 0. The sensitivity metric of each SA method is described in the Appendices. Finally, the normalized sensitivity scores of the five parameters for DT, SOT, MARS and GP methods with sample size 500 are shown in the form of radar graphs (see Figure 4a).

It is found that the parameter sensitivity rankings are consistent for the four SA methods, which identify P2 as the most sensitive parameter, P1 as the second most sensitive parameter, and the other three parameters as less sensitive. The normalized scores of the less sensitive parameters (P4, P3, and P5) for the DT method are slightly different from those for the other three methods, as shown in Figure 4a. A possible reason is that the DT fail to find the completely optimal parameter subset as the current sensitivity metric (Li et al., 2013).

Note that the parametric sensitivity rankings for all four qualitative methods are apparently inconsistent with those from the local SA method (see Figure 3), especially for the choice of the single most sensitive parameter. The most sensitive parameter found using the local SA method is P4, whereas for all four qualitative SA methods it is P2. There is a distinct bias in the choice of the most sensitive parameter, and the consequences can be very serious because this directly affects the assessment of model parameter importance and may even lead to wrong conclusions on model response shape and extreme values. Comparing with the parameter SA results from the five qualitative SA methods, only the order of parametric sensitivity rankings for local SA method is clearly inconsistent with those for the other four SA methods, especially for the single most sensitive parameter. Therefore, the SA results for the local method are considered to be inaccurate.

To carry out further validation of parameter sensitivity results and assess the efficiency of these SA methods, additional 50 and 200 QMC parameter samples are applied to the groundwater model to obtain its responses for conducting the new global SA experiments with the DT, SOT, MARS, and GP methods (see Figure 4b, c).

The SA experiments with sample size 50 are found to have parametric sensitivity rankings consistent with those with sample size 500, except for the SOT and MARS SA methods, which do not accurately distinguish the sensitivity rankings of P4 and P1 for sample size 50, but detect an obvious difference for sample size 500. However, the parameter sensitivity scores for sample size 200 are consistent with the results with sample size 500, demonstrating that a sample size of 200 is enough to obtain accurate qualitative SA results for all four SA methods. In addition, the consistent sensitivity rankings of the four SA methods from Figures 4a, c demonstrate once again the feasibility of these SA methods.

3.3. Quantitative Parameter SA Using McKay and Sobol' Methods

It is apparent in Figure 4 that the sensitivity score of a specific parameter such as P1 varies with the different qualitative SA methods used, mainly because of the coarse selection characteristics of qualitative SA, which could be improved by using quantitative parameter SA. Therefore, two quantitative SA experiments are conducted to obtain the accurate parametric sensitivity scores for the quasi-2D groundwater model. The first experiment involves conducting the McKay SA method on the model for a sample size of 2,000 (labeled McKay_2000). The second experiment involves conducting the Sobol' SA method on the model for a sample size of 2,400 (labeled Sobol'_2400).

To provide a better demonstration of the quantitative SA results, both the main effect for a certain parameter and the two-way interaction effect between two parameters are shown as contribution percentages. The sum of all parameter effects, including main effects, two-way interaction effects between two parameters, and higher-way interaction effects between more than two parameters, is defined as 100%. The higher the contribution percentage of the main effect for a certain parameter, the more sensitive is the parameter. The higher the contribution percentage of a two-way interaction effect between two parameters, the more sensitive is the joint effect between the two pa-

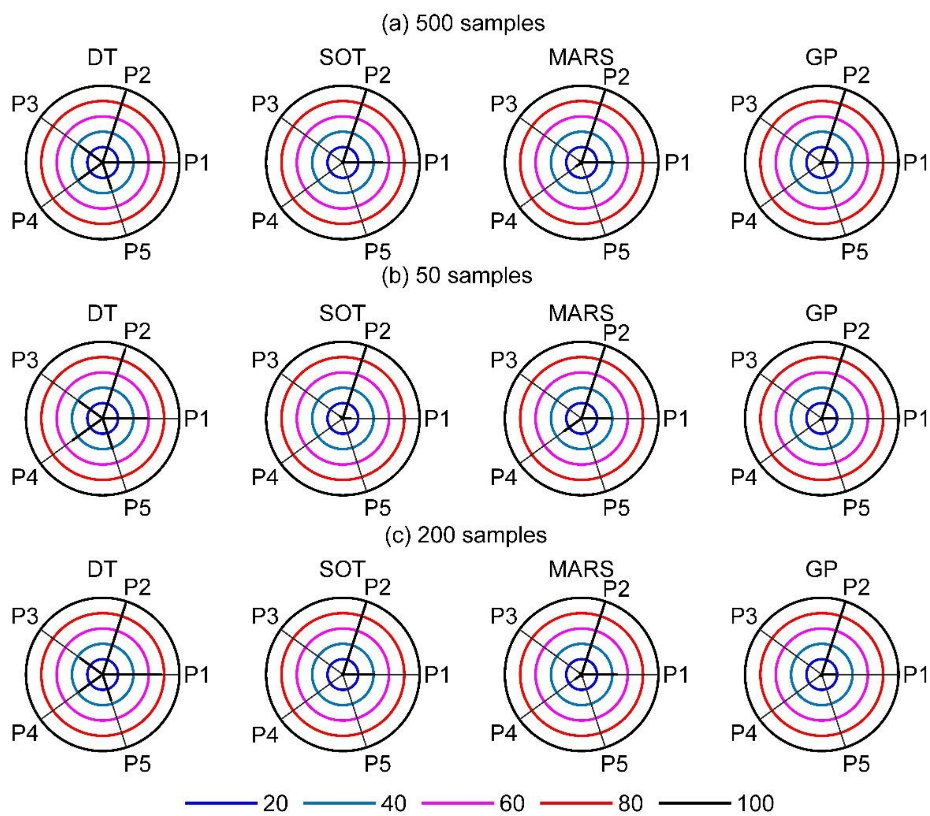

Figure 4. Normalized sensitivity scores of parameters based on global qualitative SA methods with different sample sizes.

rameters.

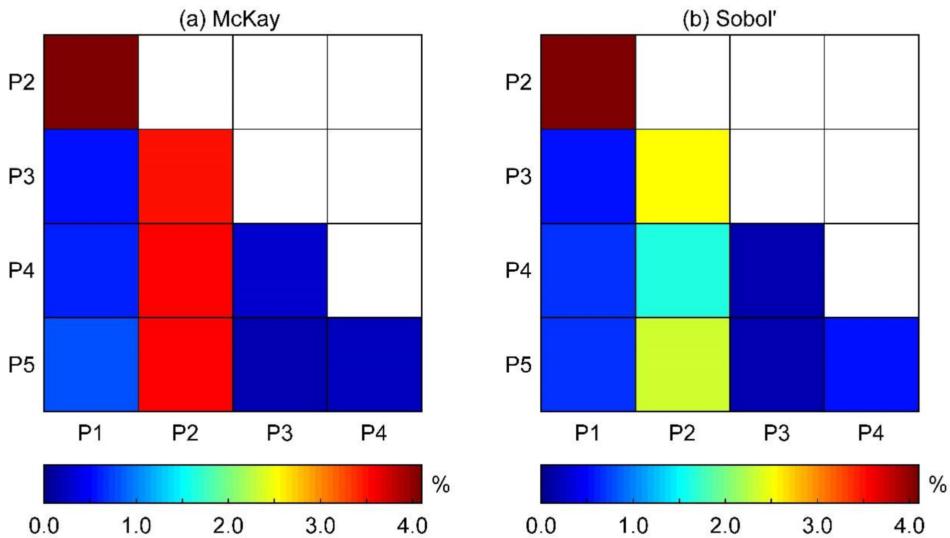

Figure 6. Comparison of the contribution percentages of two-way interaction effects for McKay_2000 and Sobol’_2400.

A comparison of the contribution percentages of the main effects of the five parameters for the McKay_2000 and Sobol’_2400 is shown in Figure 5. Obviously, P2 and P1 have higher sensitivity scores than the other parameters, which is consistent with the conclusions regarding parameter sensitivity rankings obtained by the four global qualitative SA methods (DT, SOT, MARS, and GP), but conflicts with the results of local SA, demonstrating the robustness of the conclusions on parametric sensitivity rankings and the reasonableness of the four qualitative SA methods. Moreover, the difference between the two sensitivity scores for a specific parameter, whether a sensitive or insensitive parameter, is less when the McKay SA method is replaced by the Sobol’ SA method, indicating that quantitative SA methods could obtain accurate sensitivity scores in these experiments.

This observation is distinctly different from the results of qualitative SA. In addition, the sensitivity scores for insensitive parameters using the SOT, MARS, and GP methods are also found to be closer to those obtained using the McKay and Sobol' methods than those obtained using the DT method.

Compared with the qualitative SA methods, the quantitative SA methods provide not only the main effects of parameters, but also the interaction effects, including two-way effects between two parameters and higher-way effects between more than two parameters. A comparison of the contribution percentages of the two-way interaction effects between two parameters for the McKay_2000 and Sobol’_2400 is shown in Figure 6. The strongest two-way interaction effect is found to be that between P1 and P2 for the McKay method (contribution 4.1%) and Sobol’ SA method (contribution 5.3%), demonstrating that the two-way interaction effects are basically consistent for the two methods. Moreover, the maximum contribution percentage of all two-way interaction effects (i.e., 5.3%) is far lower than the main effects of P1 (9.7%) and P2 (84.4%), which demonstrates that the model response to the five parameters relies mainly on parameter main effects, excluding interaction effects between parameters. A further three interaction effects is found (between P2 and P3, between P2 and P4, and between P2 and P5) following the highest-contribution interaction effect for the two SA methods, with contribution percentages less than 3.6%. It is found that the obvious two-way interactions are related to P2. This is mainly because P2 (i.e., the horizontal conductivity of groundwater) has stronger correlation with the objective function of groundwater table (i.e., MAE) than other parameters, which has been proved in Figure 5. Therefore, for all interactions of two parameters, the effect of the interactions including P2 on the variance of the simulated groundwater table is stronger than other interactions excluding P2. Similarly, the maximum interaction occurs between P2 and P1. The ranges of the difference in contribution percentages for the three interaction effects between McKay and Sobol’ methods vary from 0.9% to 1.9%. These low bias values also demonstrate that quantitative SA methods can obtain accurate sensitivity scores regarding two-way interaction effects in addition to the main effects of single parameters.

To validate the quantitative sensitivity results and assess the efficiency of quantitative SA methods, another two Sobol’ SA experiments are conducted. In both experiments, the other conditions are identical to Sobol’_2400 except for the sample sizes, with one being 1,200 and the other 3,600. A comparison of the contribution percentages of the main effects of five parameters for the Sobol’ SA methods using 1,200, 2,400, and 3,600 Sobol’ sample points is shown in Figure 7. Clearly, the Sobol’ SA method using 1,200 sample points (labeled Sobol’_1200) does not obtain sensitivity scores consistent with the Sobol’_2400 and Sobol’_3600. Moreover, the parameter sensitivity rankings of the Sobol’_1200 are evidently inconsistent with those of Sobol’_2400, and Sobol’_3600. Therefore, the 1,200 sample points are judged insufficient to obtain reasonable parameter sensitivity results for the quantitative Sobol’ SA method. Variations in parameter sensitivity scores for the Sobol’ SA method are smaller when the sample sizes increase from 2,400 to 3,600, demonstrating that the Sobol’ SA results using 2,400 sample points are robust and that the smallest sample size for Sobol’ SA should be 2,400.

3.4. Quantitative Parameter SA Using RS_Sobol’ Method

3.4.1. Advantage of a MARS Response Surface Model for Quantitative Parameter SA

A response surface model is constructed by determining a relationship between a set of discrete sample points consisting of parameter inputs and the corresponding outputs (or output errors). If a reasonable regression method is used, the response surface model better reflects the patterns of variation between the input parameter values and the model outputs compared with discrete sample points. Taking 60 Sobol’ sample points, for instance, two Sobol’ SA experiments are conducted: one directly conducting Sobol’ SA on the model for a sample size of 60 (labeled Sobol’_60), and the other conducting Sobol’ SA on the MARS response surface model built by the 60 sample points (labeled RS_Sobol’_60). A comparison of the contribution percentages of the main effects of the five parameters for the two Sobol’ SA experiments is shown in Figure 8. In addition, the results of the previous Sobol’_2400 (see Figure 5) are also shown to verify the results of the two Sobol’ SA experiments with 60 sample points.

Finally, the parametric sensitivity rankings for RS_Sobol’_60 are found to be closer to those for Sobol’_2400, whereas Sobol’_60 has not distinguished the sensitivity order of the parameters, as discussed. The results demonstrate that a small number of sample points may be sufficient to obtain reasonable quantitative parameter SA results when used in conjunction with a response surface model. Note also that the 60 sample points on the physical model are used to construct the statistical response model, but 12,000 points are sampled on the response surface model to conduct Sobol’ SA. For a given number of parameter inputs, the solution speed of the statistical response surface model is far faster than for the original physical model, meaning that the run time for tens of thousands of runs of the

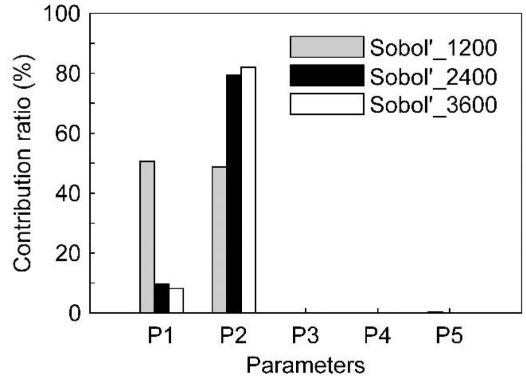

Figure 7. Comparison of the contribution percentages of the main effects of five parameters for the Sobol’ SA methods using 1200, 2400, and 3600 Sobol’ sample points.

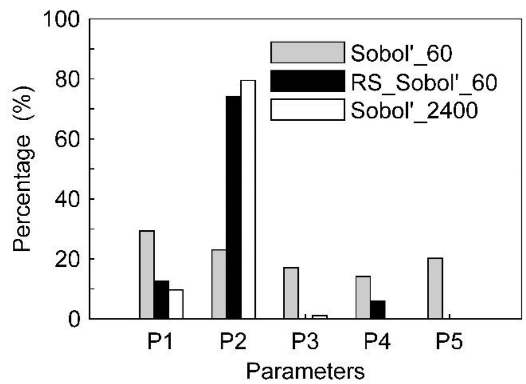

Figure 8. Comparison of the contribution percentages of the main effects of five parameters for the three Sobol’ SA. Gray bars = Sobol’_60; white bars = Sobol’_2400; black bars = RS_Sobol’_60.

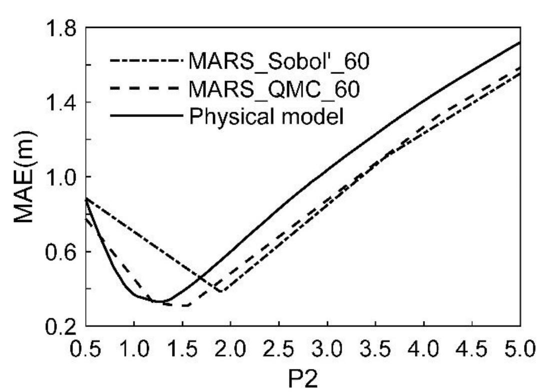

Figure 9. Variance of mean absolute error (MAE) of simulated groundwater table elevations with respect to P2 on two MARS response surface models and the physical model. The two MARS response surface models are constructed using 60 Sobol’ and 60 QMC sample points (i.e., MARS_Sobol’_60 and MARS_QMC_60).

response surface model are negligible.

3.4.2. Evaluation of the MARS Response Surface Model

Before conducting the RS_Sobol' SA experiments, the representativeness of the response surface model with respect to the original physical model should be assessed because it directly affects its quantitative SA results. Here, the representativeness of the MARS response surface model is assessed from the two aspects of sampling methods and sample sizes. The first assessment analyzes the effect of the two sampling methods (i.e., Sobol' and QMC) with the same number of sample points on the accuracy of the MARS response surface model. Here, the uniform sample size is 60. Once the more suitable sampling method (i.e., QMC or Sobol') is determined, the next step is to determine the effect of the different numbers of sample points (i.e., 60 and 2,000) from the fixed sampling method on the accuracy of the MARS response surface model.

After adjusting the suitable number of basic functions, the two MARS response surface models are built using 60 Sobol' sample points and 60 QMC sample points respectively. Compared with the corresponding sample outputs, the root means square errors (RMSE) of the simulated sample outputs using the respective MARS response surface model are 5.441 and 2.848 cm, respectively. This demonstrates that the MARS response surface model built using the QMC sample points is closer to the true physical model than that built using Sobol' sample points. As shown in Figure 5, P2 is the most sensitive parameter, and therefore the responses of P2 in the different models are extracted to analyze the difference between the two response surface models. With other parameters hold constant at their default values, the variances of the MAE of the simu-

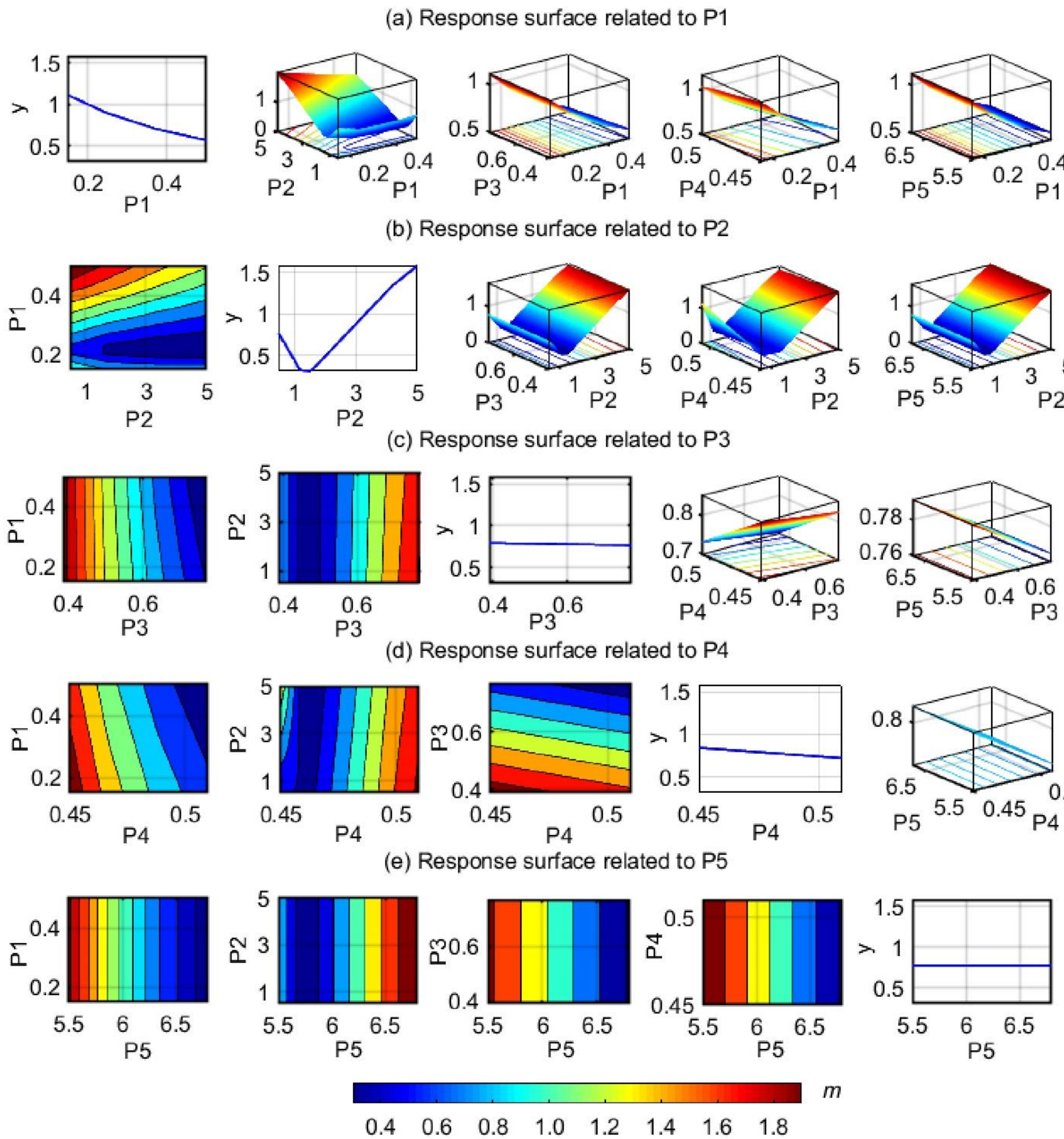

Figure 10. MARS response surface model built by 60 QMC sample points.

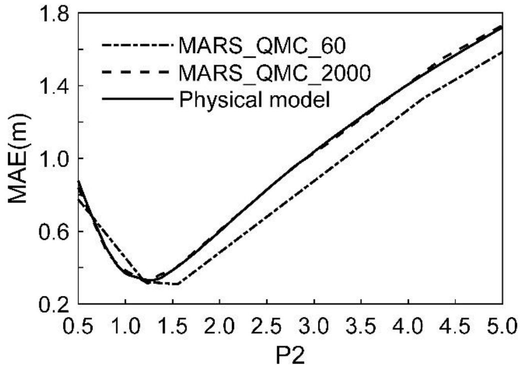

Figure 11. Variances of the mean absolute error (MAE) of simulated groundwater table elevations with respect to P2 for two MARS response surface models and the physical model. The two MARS response surface models are constructed using 60 and 2000 QMC sample points (i.e., MARS_QMC_ 60 and MARS_QMC_2000).

lated groundwater table elevations with respect to P2 in the three response models, including the two MARS models built using QMC and Sobol' sample points respectively and the real physical model, are shown in Figure 9. The results show that the MARS response surface model built from 60 QMC sample points (labeled MARS_QMC_60) is closer to the physical model than that built from 60 Sobol' sample points (labeled MARS_Sobol'_60). This result can be explained by noting that the QMC sampling method has a better space-filling capability than the Sobol' sampling method. Therefore, constructing the response surface model using QMC sample points rather than Sobol' sample points is recommended.

As a better statistical response surface model, MARS_ QMC_60 is illustrated in Figure 10. In each subplot, only the variable parameters (one or two parameters) are shown, whereas other parameters are held constant at their default values. From five subplots reflecting the response to a single parameter along the diagonal, it is clear that P2 and P1 have significant effects on model response *y*, P4 has a slight effect on response *y*, and P3 and P5 have basically no effect on response *y*. These findings are consistent with the conclusions of parameter sensitivity rankings over global qualitative and quantitative SA methods (see Figures 4a and 5).

After the QMC sampling method is decided upon, the representativeness of the response surface model with the two different sample sizes is also investigated: One MARS response surface model is built using 60 QMC sample points (i.e., MARS_QMC_60), and the other is built using 2000 QMC sample points (MARS_QMC_2000). For the two MARS response surface model, the variances of the MAE of the simulated groundwater table elevations with respect to P2 are shown in Figure 11. The response surface model is found to become closer to the physical model as sample size increases. Especially, when the sample size increases to 2,000, there is basically no difference between the MARS response surface model and physical model. However, larger sample sizes greatly increase the number of model runs with the different parameter values included. It is found that MARS_QMC_60 has a similar

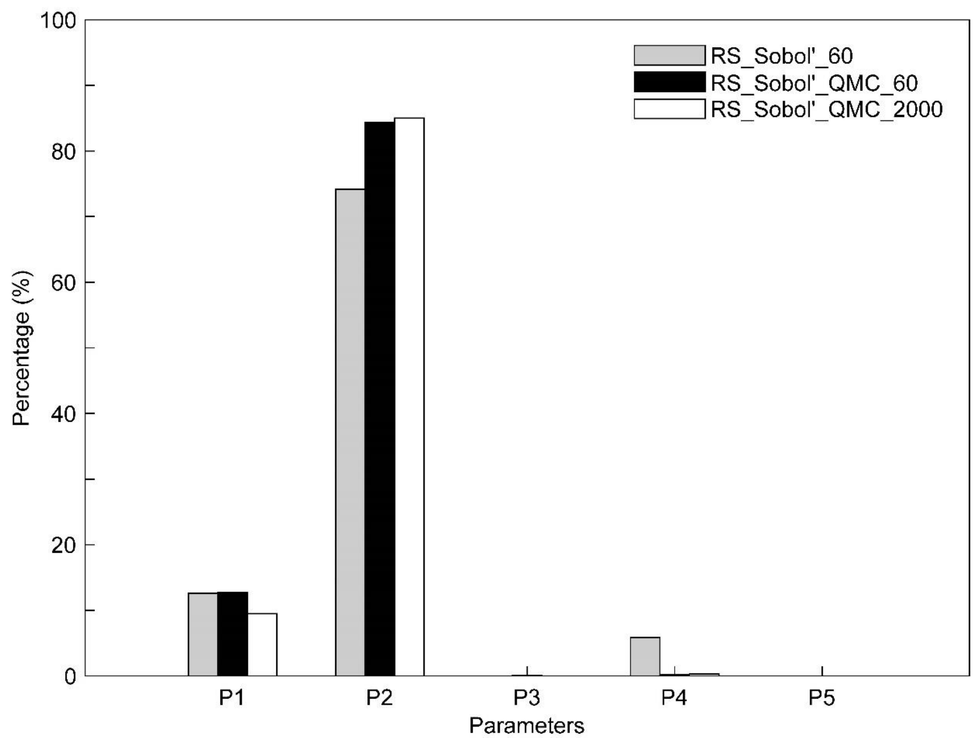

Figure 12. Comparison of the contribution percentages of the main effects of five parameters for three RS_Sobol' SA (RS_Sobol'_60, RS_Sobol'_QMC_60, and RS_Sobol'_QMC_2000).

variance pattern with real physical model, therefore, 60 sample points may be enough for model parameter SA by the RS_Sobol' method.

3.4.3. Comparison of RS_Sobol' SA with the Different QMC Sample Sizes Used to Build the MARS Response Surface Models

The MARS response surface model built using QMC sample points resembles the physical model more closely than that built using Sobol' sample points, as discussed. However, whether the corresponding MARS model has a consistent advantage over the quantitative Sobol' SA results must also be determined. Three RS_Sobol' SA experiments are conducted using RS_Sobol'_60 together with RS_Sobol' SA with 60 QMC sample points (RS_Sobol'_QMC_60) and 2,000 QMC sample points (RS_Sobol'_QMC_2000) to build the MARS response surface models. A comparison of the three RS_Sobol' SA results is shown in Figure 12. Compared to RS_Sobol'_

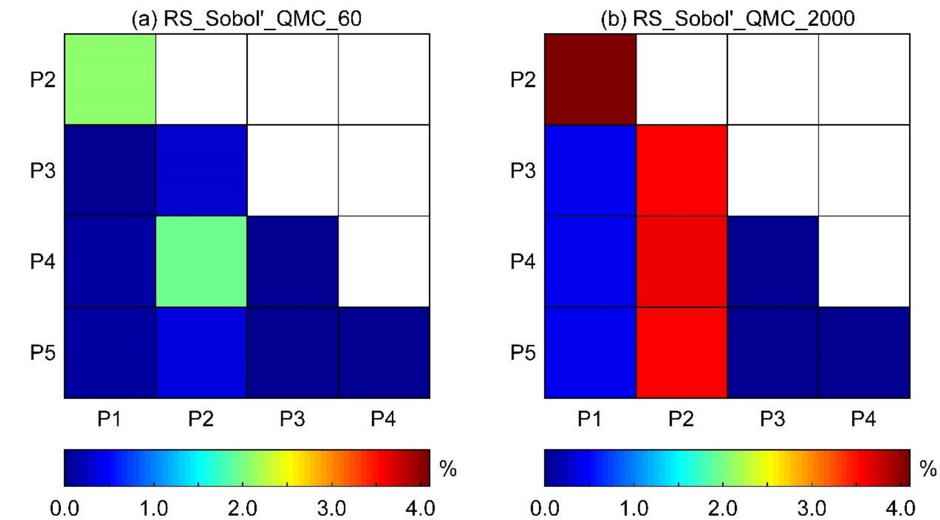

Figure 13. Comparison of contribution percentages of two-way interaction effects between two parameters for RS_Sobol'_QMC_60 and RS_Sobol'_QMC_2000.

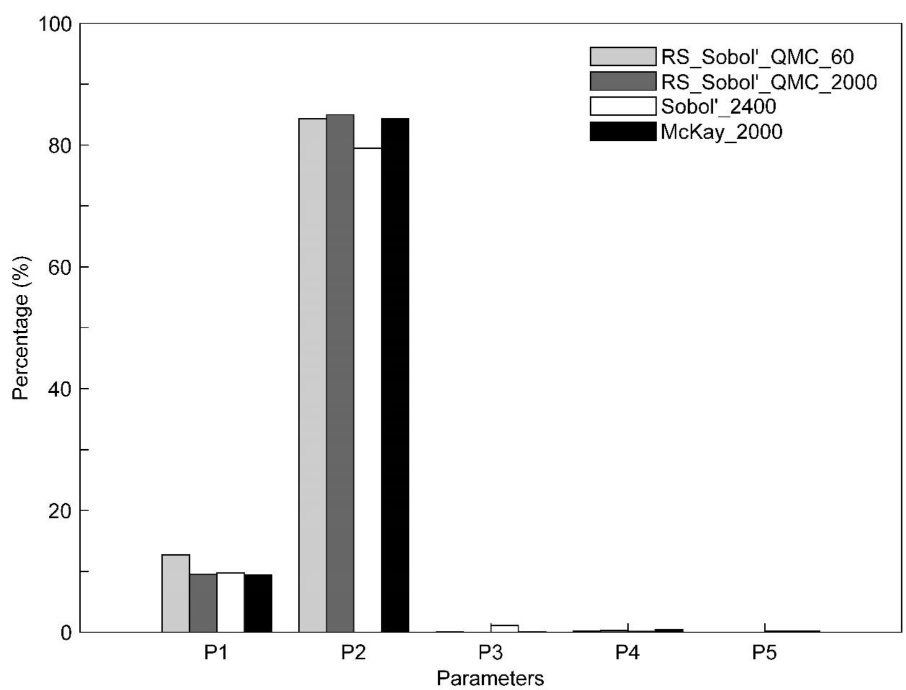

Figure 14. Comparison of the contribution percentages of the five parameter main effects for the four quantitative SA (RS_Sobol'_QMC_60, RS_Sobol'_QMC_2000, Sobol'_2400, and McKay_2000).

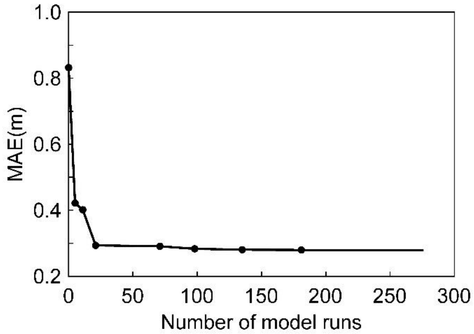

Figure 15. Convergence results for the model after optimizing the two most sensitive parameters.

QMC_2000, RS_Sobol'_QMC_60 has more consistent contribution percentages of the main effects of five parameters than RS_Sobol'_60, although RS_Sobol'_60 provided accurate parameter sensitivity rankings. These results demonstrate that the more suitable substitution response surface model would provide more accurate parameter sensitivity scores.

Besides the main effect for each single parameter, the two-way interaction effects between two parameters for RS_Sobol' _QMC_60 and RS_Sobol'_QMC_2000 are compared. A comparison of the contribution percentages of two-way interaction effects for RS_Sobol'_QMC_60 and RS_Sobol'_QMC_2000 is shown in Figure 13. From which, some conclusions are drawn as for the comparison of McKay and Sobol' quantitative SA methods. The most sensitive two-way interaction effects occur between P2 and P1, and the leading four two-way interaction effects are related to P2. In addition, the maximum error values of all two-way interaction effects between RS_Sobol'_ QMC_60 and RS_Sobol'_QMC_2000 are less than 2%, demonstrating that RS_Sobol'_QMC_60 obtain relatively consistent sensitivity scores with RS_ Sobol'_QMC_2000 for two-way interaction effects. Overall, it is apparent from Figures 12 and 13 that 60 QMC sample points could obtain basically consistent parameter sensitivity scores with 2,000 QMC sample

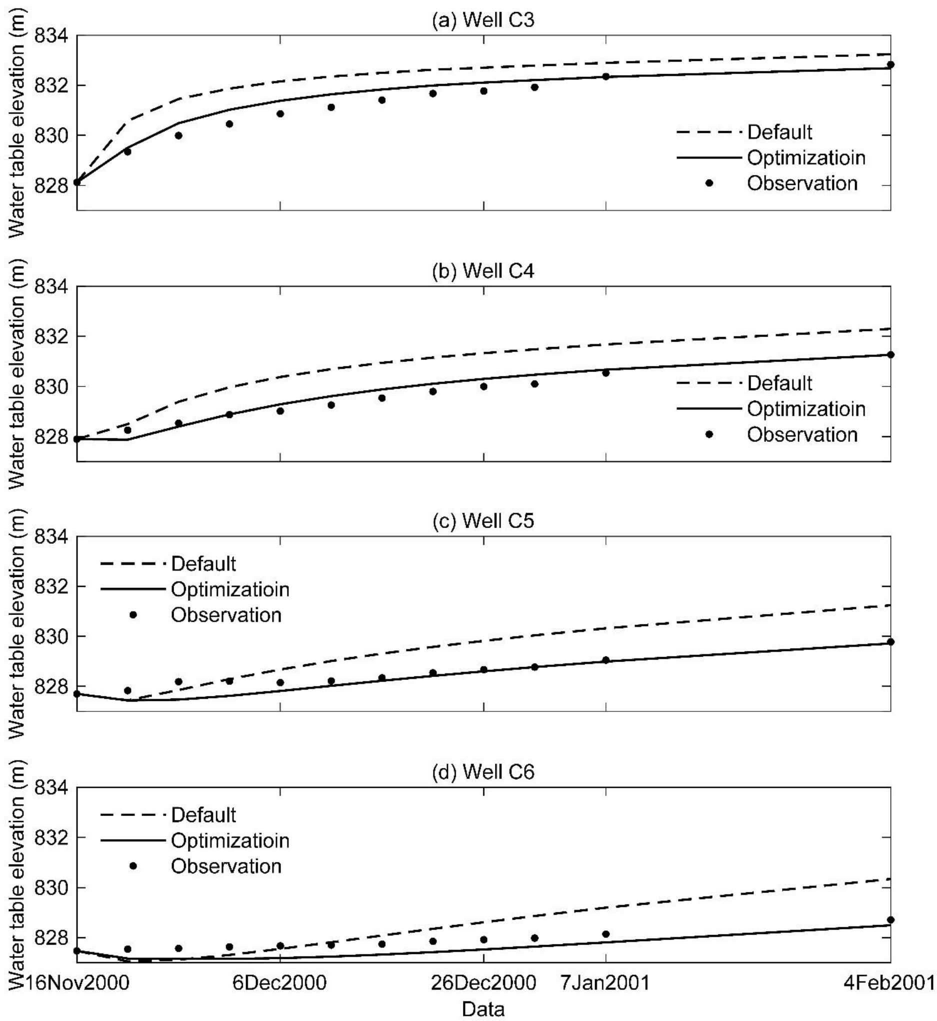

Figure 16. Comparison of the observed and two simulated water table elevations obtained by the quasi-2D groundwater table model with default and optimal parameters for wells C3–C6.

points using the RS_Sobol' SA method.

Although RS_Sobol'_QMC_60 has basically consistent sensitivity scores with RS_Sobol'_QMC_2000, including the contribution percentages of parameter main effects and two-way interaction effects, the difference in parameter sensitivity scores between RS_Sobol'_QMC_2000 and the other quantitative SA (i.e., Sobol'_2400, and McKay_2000) is not assessed. Figure 14 shows a comparison of the contribution percentages of the main effects of five parameters for the four quantitative SA experiments (RS_Sobol'_QMC_60, RS_Sobol' _QMC_ 2000, Sobol'_2400, and McKay_2000). The results showe that RS_Sobol'_QMC_2000 has contribution percentages for the five parameter main effects that are consistent with the other quantitative SA methods, especially for McKay_ 2000, demonstrating that the results of parameter main effects from RS_Sobol'_QMC_2000 are accurate and furthermore that the corresponding results from RS_Sobol'_QMC_60 are credible. In addition, by a comparison of Figures. 6 and 13, it is also found that the contribution percentages of two-way interaction effects for RS_Sobol'_QMC_2000 are basically consistent with those for McKay_2000, demonstrating that the results of parameter two-way interaction effects obtained by RS_Sobol'_ QMC_2000 are also accurate and furthermore that the corresponding results from RS_Sobol'_QMC_60 are credible. Overall, due to fewer errors in the SA results using RS_Sobol'_ QMC_60 and RS_Sobol'_QMC_2000. It is recommended that RS_Sobol' SA method using 60 sample points (i.e., RS_ Sobol'_QMC_60) is used to obtain accurate parameter sensitivity scores for the quasi-2D groundwater model.

3.5. Identification of Two Sensitive Parameters

The sensitive parameters for the model are selected using previous SA methods. The results demonstrate that P1 and P2 are sensitive to the simulation of water table elevations. Based on the results of SA analyses, identifying the suitable sensitive parameter values facilitates the improvement of model simulation results and thus reduces model uncertainty. Next, the Shuffled Complex Evolution method (Duan et al., 1992) is used to optimize the two sensitive parameters of the model, while other parameters are held at their default values. The simulation period is the same as for SA analysis experiments. The optimization results are shown in Figure 15. After 180 searches in the parameter space, the optimal simulation results of the water table elevation for the quasi-2D groundwater model are found. The corresponding MAE of simulated water table elevations reduces from 0.832 m using the default parameters, to 0.28 m with the optimal parameters, an improvement of approximately 66%. This implies that the identification of the sensitive parameters significantly reduces model simulation uncertainty, and that the optimization of the sensitive parameters of the model is very efficient. The optimal values for P1 and P2 are 0.494 and 1.66, respectively.

Based on the observation data of groundwater table for four wells (C3 ~ C6), the simulation results of the quasi-2D groundwater model with the default and optimal parameters are compared. The variances of the simulated and observed groundwater table elevations for four wells (C3 ~ C6) are shown in Figure 16. It is found that the simulation results of the model with the default parameters are significantly higher than the observations for wells C3 ~ C6, which implies the model should be calibrated to close to the observed values. After optimizing the two sensitive parameters, the simulation results of the model with the optimal parameter values are basically consistent with observations. It demonstrates that the optimization for the sensitive parameters of the model is very effective.

3.6 Physical Interpretation and Verification of the Parameter Sensitivity Results

The reasonableness of the sensitive parameters obtained by various SA methods should also be verified by explanations of parametric physical meanings. Parameter P1, the specific yield of soil, is one of the most sensitive parameters for simulation of groundwater table elevation. When a certain amount of river water is poured into unsaturated soil of a river bank, the smaller the value of specific yield of soil (i.e., P1 value), the higher the groundwater table rises. This occurs because the increase in groundwater table elevation is mainly the results of filling the pores in unsaturated soil. Parameter P2, which is related to horizontal hydraulic conductivity, is another highly sensitive parameter for simulation of groundwater table elevation. In stream water conveyance, the hydraulic potential difference between river water and an unsaturated soil column is greater than that between unsaturated soil columns. Darcy's law states that a larger horizontal hydraulic conductivity (i.e., P2) with a constant hydraulic potential difference will inject more river water into the unsaturated soil column. However, there is no vertical downward flow in the vertical groundwater section due to the equal potentials in the vertical direction, and therefore the added water mainly serves to elevate the groundwater table.

4. Conclusions and Discussions

In this study, three types of parametric SA methods are systematically assessed for their effectiveness and efficiency: qualitative SA, quantitative SA, and the response surface model-based quantitative SA. According to the requirements of SA methods, the suitable and more uniform QMC sampling method is used to enhance the representativeness of perturbed parameter samples and thus SA efficiency. The assessments are conducted on a quasi-2D groundwater model for groundwater table elevation simulation at the Yingsu section in the lower reaches of the Tarim River in China during the second period of river water release (i.e., from 16 November, 2000 to 4 February, 2001).

Five common methods of qualitative SA are assessed and some general conclusions are drawn:

(1) The local method may accumulate errors in sensitive parameters, because the several parameter samples used do not represent the whole parameter space.

(2) With 500 samples, the parametric sensitivity rankings

of four SA methods, DT, SOT, MARS and DT, are consistent, demonstrating that all four qualitative SA methods are effective. Finally, P2 (horizontal groundwater hydraulic conductivity) and P1 (specific yield of soil) are identified as the most sensitive parameters in simulating groundwater table elevation. The parameter sensitivity rankings for four SA methods are consistent using 200 samples but fail using 50 samples, demonstrating that a sample size of 200 is enough to obtain accurate sensitive parameters.

(3) For the GP method, 50 samples produce consistent parametric sensitivity scores with 500 samples, which differs from the other three SA methods. This demonstrates that GP is a highly efficient qualitative SA method.

For quantitative SA, two variance-based methods are assessed, and the conclusions are drawn:

(1) With about 2,000 samples, both the McKay and Sobol' quantitative SA methods obtain consistent parameter sensitivity rankings with four qualitative SA methods, but also basically equal parameter sensitivity scores for whatever main effect or two-way interaction effect, demonstrating that they are as effective as the four qualitative SA methods.

(2) The variation of each parameter sensitivity score for the Sobol' quantitative SA method is smaller when the sample size increases from 2,400 to 3,600. However, scores with 1,200 samples are evidently different from those for 2,400 or 3,600 samples, and therefore a sample size of 2,400 is enough to obtain accurate sensitivity scores. This implies that quantitative SA methods require more samples than qualitative SA methods for obtaining reasonable sensitivity parameters.

For the response surface model-based quantitative SA, the RS_Sobol' SA method is assessed. Theoretically the method is highly efficient because it combines the strengths of qualitative and quantitative SA methods; however, the extent to which the response surface model represents the original model is very critical to the conducting of a reasonable quantitative SA. Therefore, the MARS response surface model should be assessed before using the Sobol' SA method on it. The main conclusions are:

(1) Using a small number of sample points, the quantitative SA method, in association with a response surface model, obtains reasonable sensitivity parameters.

(2) The representativeness of the MARS response surface model with respect to the original physical model is assessed from the two aspects of sampling methods and sample sizes. It is demonstrated that the MARS response model using 60 QMC samples produces a variance pattern similar to that of the real physical model.

(3) For the RS_Sobol' SA method, the 60 QMC samples obtain basically consistent parameter sensitivity scores with 2,000 QMC samples for both main effect and two-way interaction effect. Additionally, the parameter sensitivity scores are consistent with those obtained by common quantitative SA methods (e.g., Sobol' and Mckay) using at least 2,000 samples. For the RS_Sobol' SA method, 60 samples are found to be enough to obtain accurate sensitivity scores comparable with the quantitative SA method using 2,000 samples, demonstrating that it is effective and efficient.

Overall, our assessments for the three types of parametric SA methods are very practical. It examines the effectiveness of these SA methods by comparing parameter sensitivity results and demonstrates the efficiency of these SA methods by determining the minimum sample size required. Finally, the response surface model-based quantitative SA method is demonstrated to be both effective and highly efficient. This provides a basis to enable other researchers to conduct more efficient parametric SA methods on physical models, to quickly screen the sensitive parameters of the model. This is also useful for enhancing parameter optimization speed, especially for complex physical models requiring with extremely large computational demand.

This work focuses on the assessment of SA methods, and therefore the case study is located in a small region (i.e., the river bank in an arid region). If the quasi-2D groundwater model is applied to the larger region (e.g., the whole of China), it not only increases computational time for a model run compared to current simulation, but also requires separate experiments on the parameter SA and optimization of sensitive model parameters for the different climatic regions to obtain accurate simulations of groundwater table elevation. In this case, the most highly efficient SA method is very important in enhancing the efficiency of parameter optimization by quickly screening out a small number of sensitive parameters to be optimized using as few model runs as possible. Subsequent work will incorporate the RS_Sobol' quantitative SA method and the highly efficient parameter optimization method into the model to simulate the variation of groundwater table in larger regions such as the whole of China or a global region.

Acknowledgments. This work is supported by the Ministry of Science and Technology of China (Grant No. IUMKY201603), Intergovernmental Key International S & T Innovation Cooperation Program (Grant No. 2016YFE0102400), and the China Special Fund for Meteorological Scientific Research (Grant No. GYHY201506002, CRA40: 40-year CMA global atmospheric reanalysis), the National Natural Science Foundation of China (Grant Nos. 41305052, 41375139), and the Fundamental Research Funds for the Central Universities-Beijing Normal University Research Fund (Grant No. 12500-310421103). Special thanks to the group of Prof. Yaning Chen at Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, for providing the groundwater table elevation validation

References

Alharthi, A. and Lange, J. (1987). Soil water saturation: Dielectric determination. *Water Resour. Res.,* 23(4), 591-595. https://doi.org/

10.1029/WR023i004p00591
Arnold, J., Allen, P., and Bernhardt, G. (1993). A comprehensive surface-groundwater flow model. *J. Hydrol.,* 142(1-4), 47-69. https://doi.org/10.1016/0022-1694(93)90004-S
Bhadeshia, H.K.D.H. (1999). Neural networks in materials science. *ISIJ Int.,* 39(10), 966-979. https://doi.org/10.2355/isijinternational.39.966
Breiman, L., Friedman, J., Stone, C., and Olshen, R. (1984). *Classification and Regression Trees*. Chapman and Hall/CRC Press, Florida.
Caflisch, R.E. (1998). Monte Carlo and quasi-Monte Carlo methods. *Acta Numer.,* 7(2), 1-49. https://doi.org/10.1017/S0962492900002804
Chen, Y., Chen, Y., Xu, C., Ye, Z., Li, Z., Zhu, C., and Ma, X. (2010). Effects of ecological water conveyance on groundwater dynamics and riparian vegetation in the lower reaches of Tarim River, China. *Hydrol. Process.,* 24(2), 170-177. https://doi.org/10.1002/hyp.7429
Chen, F. and Dudhia, J. (2001). Coupling an Advanced Land Surface Hydrology Model with the Penn State NCAR MM5 Modeling System. Part I: Model Implementation and Sensitivity. *Mon. Weather Rev.,* 129(4), 569-585. https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2
Clapp, R.B. and Hornberger, G.M. (1978). Empirical equations for some soil hydraulic properties. *Water Resour. Res*., 14(4), 601-604. https://doi.org/10.1029/WR014i004p00601
Collins, D.C. and Avissar, R. (1994). An evaluation with the Fourier amplitude sensitivity test (fast) of which land-surface parameters are of greatest importance in atmospheric modeling. *J. Clim.,* 7(5), 681-703.
Confalonieri, R., Bellocchi, G., Bregaglio, S., Donatelli, M., and Acutis, M. (2010). Comparison of sensitivity analysis techniques: a case study with the rice model warm. *Ecol. Model.,* 221(16), 1897-1906. https://doi.org/10.1016/j.ecolmodel.2010.04.021
Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.,* 20(3), 273-297. https://doi.org/10.1007/BF009 94018
Cosby, B.J., Hornberger, G.M., Clapp, R.B., and Ginn, T.R. (1984). A Statistical Exploration of the Relationships of Soil Moisture Characteristics to the Physical Properties of Soils. *Water Resour. Res.,* 20(6), 682-690. https://doi.org/10.1029/WR020i006p00682
Di, Z., Xie, Z., Yuan, X., Tian, X., Luo, Z., and Chen, Y. (2011). Prediction of water table depths under soil water-groundwater interaction and stream water conveyance. *Sci. China Earth Sci.,* 54(3), 420-430. https://doi.org/10.1007/s11430-010-4050-8
Di, Z., Duan, Q, Gong, W., Ye, A., and Miao, C. (2017). Parametric sensitivity analysis of precipitation and temperature based on multi-uncertainty quantification methods in the weather research and forecasting model. *Sci. China Earth Sci.,* 60(5), 876-898. https://doi.org/10.1007/s11430-016-9021-6
Dickinson, R., Henderson-Sellers, A., Kennedy, P., and Wilson, M. (1986). *Biosphere/Atmosphere Transfer Scheme (BATS) for the NCAR Community Climate Model.* National Center for Atmosphere Research, Boulder.
Duan, Q., Sorooshian, S., and Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.,* 28(4), 1015-1031. https://doi.org/10.1029/91WR02985
Duan, Q., Gupta, H., Sorooshian, S., Rousseau, A., and Turcotte, R. (2003). *Calibration of Watershed Models.* Washington, DC, American Geophysical Union. https://doi.org/10.1029/WS006
Durand, V., Léonardi, V., de Marsily, G., and Lachassagne, P. (2017). Quantification of the specific yield in a two-layer hard-rock aquifer model. *J. Hydrol.,* 551, 328-339. https://doi.org/10.1016/j.jhydrol.2017.05.013
Eirola, E., Liitiäinen, E., Lendasse, A., Corona, F., and Verleysen, M. (2008). Using the Delta test for variable selection. *European Symposium on Artificial Neural Networks*, Bruges, Belgium, April 23-25.
Faure, H. (1982). Discrépance de suites associées à un système de numération (en dimensions). *Acta. Arith.,*4, 337-351. (in French). https://doi.org/10.4064/aa-41-4-337-351
Feng, Q., Endo, K., and Cheng, G. (2001). Towards sustainable development of the environmentally degraded arid rivers of China–A case study from Tarim River. *Environ. Geol.,* 41(1-2), 229-238. doi: 10.1007/s002540100.
Gan, Y., Duan, Q., Gong, W., Tong, C., Sun, Y., Chu, W., Ye, A., Miao, C., and Di, Z. (2014). A comprehensive evaluation of various sensitivity analysis methods: a case study with a hydrological model. *Environ. Modell. Software.,* 51(1), 269-285. https://doi.org/10.1016/j.envsoft.2013.09.031
Gong, W., Duan, Q., Li, J., Wang, C., Di, Z., Ye, A., Miao, C., and Dai, Y. (2016). An intercomparison of sampling methods for uncertainty quantification of environmental dynamic models. *J. Environ. Inf.,* 28(1), 11-24. https://doi.org/10.3808/jei.201500310
Griewank, A.O.K. and Walther, A. (2000). *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation,* Society for Industrial and Applied Mathematics, Philadelphia, P.A.
Halton, J.H. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numer. Math.,* 2(1), 84-90. https://doi.org/10.1007/BF01386213
Homma, T. and Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliab. Eng. Syst. Saf.,* 52(1), 1-17. https://doi.org/10.1016/0951-8320(96)00002-6
Hou, Z., Huang, M., Leung, L.R., Lin, G., and Ricciuto, D.M. (2012). Sensitivity of surface flux simulations to hydrologic parameters based on an uncertainty quantification framework applied to the Community Land Model. *J. Geophys. Res (D Atmos.)*, 117, D15108. https://doi.org/10.1029/2012JD017521
Jabro, J.D. (1992). Estimation of Saturated Hydraulic Conductivity of Soils From Particle Size Distribution and Bulk Density Data. *Trans. ASAE,* 1992, 35(2), 557-560. https://doi.org/10.13031/2013.28633
Li, J., Duan, Q., Gong, W., Ye, A., Dai, Y., Miao, C., Di, Z., Tong, C., and Sun, Y. (2013). Assessing parameter importance of the common land model based on qualitative and quantitative sensitivity analysis. *Hydrol. Earth. Syst. Sci.,* 10(8), 3279-3293. https://doi.org/10.5194/hess-17-3279-2013
Liang, X., Lettenmaier, D., Wood, E., and Burges, S. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res. (DAtmos.),* 99(D7), 14415-14428. https://doi.org/10.1029/94JD00483
McKay, M.D., Morrison, J.D., and Upton, S.C. (1998). Evaluating prediction uncertainty in simulation models. *Comput. Phys. Commun.,* 117(1-2), 44-51. https://doi.org/10.1016/S0010-4655(98)00155-6
McKay, M.D., Beckman, R.J., and Conover, W.J. (2000). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics,* 42(1), 55-61. https://doi.org/10.2307/1271432
Morris, M. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics,* 33(2), 161-174. https://doi.org/10.2307/1269043
Navarra, A. and Simoncini, V. (2010). Canonical correlation analysis, *A Guide to Empirical Orthogonal Functions for Climate Data Analysis,* Netherlands, Springer, 107-121. https://doi.org/10.1007/978-90-481-3702-2_7
Neumann, M.B. (2012). Comparison of sensitivity analysis methods for pollutant degradation modelling: a case study from drinking water treatment. *Sci. Total Environ.*, 433(1), 530-537. https://doi.org/10.1016/j.scitotenv.2012.06.026
Niederreiter, H. (1988). Low-discrepancy and low-dispersion sequences. *J. Number Theory,* 30(1), 51-70. https://doi.org/10.1016/0022-314X(88)90025-X
Ortega, J., and Rheinboldt, W. (1970). Iterative solution of nonlinear equations in several variables, *Computer Science and Applied Math-*

ematics, New York, Academic Press.

Patil, S.R., and Frey, H.C. (2004). Comparison of Sensitivity Analysis Methods Based on Applications to a Food Safety Risk Assessment Model. *Risk Anal.,* 24(3), 573-585. https://doi.org/10.1111/j.0272-4332.2004.00460.x

Rasmussen, C., and Nickisch, H. (2010). Gaussian processes for machine learning (GPML) toolbox. *J. Mach. Learn. Res.,* 11(6), 3011-3015.

Richards, L.A. (1931). Capillary conduction of liquids through porous mediums. *Physics,* 1(5), 318-333. https://doi.org/10.1063/1.1745010

Steinberg, D., Colla, P., and Martin, K. (1999). *MARS User Guide.* San Diego C.A., Salford Systems.

Storlie, C.B. and Helton, J.C. (2008). Multiple predictor smoothing methods for sensitivity analysis. *Reliab. Eng. Syst. Saf.,* 93(1), 55-77. https://doi.org/10.1016/j.ress.2006.10.013

Sobol', I.M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simulat*., 55(1-3), 271-280. https://doi.org/10.1016/S0378-4754(00)00270-6

Sun, X.Y., Newham, L.T.H., Croke, B.F.W., and Norton, J.P. (2012). Review: three complementary methods for sensitivity analysis of a water quality model. *Environ. Model. Software,* 37, 19-29. https://doi.org/10.1016/j.envsoft.2012.04.010

Taheri Tizro, A., Voudouris, K., and Basami, Y. (2012). Estimation of porosity and specific yield by application of geoelectrical method-A case study in western Iran. *J. Hydrol.,* 454-455(4), 160-172. https://doi.org/10.1016/j.jhydrol.2012.06.009

Wang, C., Duan, Q., Gong, W., Ye, A., Di, Z., and Miao, C. (2014). An evaluation of adaptive surrogate modeling based optimization with two benchmark problems. *Environ. Model. Software,* 60(76), 167-179. https://doi.org/10.1016/j.envsoft.2014.05.026

Yang, B., Qian, Y., Lin, G, Leung, R., and Zhang Y. (2012). Some issues in uncertainty quantification and parameter tuning: a case study of convective parameterization scheme in the WRF regional climate model. *Atmos. Chem. Phys.,* 12(5), 2409-2427. https://doi.org/10.5194/acp-12-2409-2012