An efficient integrated approach for global sensitivity analysis of hydrological model parameters

Che-sheng Zhan [a],[*],[1], Xiao-meng Song [b],[**],[1], Jun Xia [a], Charles Tong [c]

[a] Key Laboratory of Water Cycle & Related Land Surface Processes, Institute of Geographical Sciences and Natural Resources Research, CAS, Beijing 100101, China
[b] State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing 210029, China
[c] Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, CA 94551-0808, USA

ARTICLE INFO

ABSTRACT

Efficient sensitivity analysis, particularly for the global sensitivity analysis (GSA) to identify the most important or sensitive parameters, is crucial for understanding complex hydrological models, e.g., distributed hydrological models. In this paper, we propose an efficient integrated approach that integrates a qualitative screening method (the Morris method) with a quantitative analysis method based on the statistical emulator (variance-based method with the response surface method, named the RSMSobol' method) to reduce the computational burden of GSA for time-consuming models. Using the Huaihe River Basin of China as a case study, the proposed approach is used to analyze the parameter sensitivity of distributed time-variant gain model (DTVGM). First, the Morris screening method is used to qualitatively identify the parameter sensitivity. Subsequently, the statistical emulator using the multivariate adaptive regression spline (MARS) method is chosen as an appropriate surrogate model to quantify the sensitivity indices of the DTVGM. The results reveal that the soil moisture parameter WM is the most sensitive of all the responses of interest. The parameters Kaw and g_1 are relatively important for the water balance coefficient (WB) and Nash–Sutcliffe coefficient (NS), while the routing parameter $RoughRss$ is very sensitive for the Nash–Sutcliffe coefficient (NS) and correlation coefficient (RC) response of interest. The results also demonstrate that the proposed approach is much faster than the brute-force approach and is an effective and efficient method due to its low CPU cost and adequate degree of accuracy.

1. Introduction

Distributed hydrological models play a key role in studying hydrology and water resources and are also particularly useful tools for investigating many important issues in the planning, design, operation and management of water resources (Muleta and Nicklow, 2005). Parameter identification, model calibration and uncertainty quantification are important steps in the modeling process. These steps must be considered to ensure that the results are credible and that valuable information is obtained (Campolongo et al., 2007; Jakeman et al., 2006). Most hydrological models are highly complex and are characterized by a set of parameters that may not be exactly known or directly measurable. Therefore, model parameter estimation must be performed by calibration in most model applications, which can reduce the parameter uncertainty in the simulation results (Cibin et al., 2010). However, when the number of parameters is large, the calibration processes may be computationally intensive, and the computational cost may become prohibitive. A lack of knowledge about parameter sensitivities may result in time wasted on insensitive parameters (Bahremand and De Smedt, 2008). Therefore, focusing on sensitive parameters can reduce uncertainty and lead to a better understanding of the model and more satisfactory simulation (Lenhart et al., 2002). At present, sensitivity analysis (SA) is helpful to identify the important and requisite factors or parameters and rank parameters that have significant impact on specific model outputs of interest (Saltelli et al., 2000; Tarantola and Saltelli, 2003; Sieber and Uhlenbrook, 2005). In addition, SA provides useful information regarding the behavior of the simulation model, including the identification of relevant model inputs and the information on model construction (Confalonieri, 2010). In general,

* Corresponding author.
** Corresponding author. Tel./fax: +86 25 8582 8535.
E-mail addresses: zhancs2006@gmail.com (C.-s. Zhan), xmsong.cn@gmail.com (X.-m. Song).
[1] Che-sheng Zhan and Xiao-meng Song contributed equally to this work and should be regarded as co-first and corresponding authors.

sensitivity analysis is conducted for a variety of reasons. For examples, to determine which input parameters contribute most to output variability, additional research is required to increase knowledge of parameter behavior to reduce output uncertainty, to determine which groups of parameters interact with each other if parameter interactions exist, to determine which parameters are insensitive and can be held constant or eliminated from the final model, and to identify the optimal regions within the parameter space in subsequent calibration studies.

Uncertainty analysis (UA) generally refers to the determination of the uncertainty that derives from uncertainty in model factors (Helton et al., 2006), and SA refers to the determination of the contributions of individual and different sources of uncertain inputs to the uncertainty in the output of a model (Saltelli et al., 2008). SA methods are generally classified as either local or global SA (Saltelli et al., 2000; Muleta and Nicklow, 2005; van Griensven et al., 2006). Local SA (LSA) methods compute or approximate the local response of the model outputs by varying input factors or parameters individually with other factors or parameters at some nominal settings, known as the "baseline" or "nominal value" point, in the hyperspace of the input factors (Spruill et al., 2000; Holvoet et al., 2005; Cibin et al., 2010; Saltelli and Annoni, 2010). By contrast, global sensitivity analysis (GSA) evaluates the effects of input variations on the outputs in the entire allowable ranges of the input space (Confalonieri et al., 2010; Tong, 2010). GSA has become widely used in hydrological applications in recent years (Crosetto and Tarantola, 2001; van Griensven et al., 2006; Cibin et al., 2010; Ren et al., 2010) because it accounts for the effects of interactions between different parameters, particularly the nonlinear relationship between parameters and state variables (Saltelli et al., 2000; Makler-Pick et al., 2011). Saltelli et al. (2000, 2004) defined GSA methods by two properties (Tong, 2007b, 2010): the inclusion of influence of scales and shapes of the probability density functions for all inputs and the sensitivity estimates of individual inputs that are evaluated while varying all other inputs.

GSAs offer a comprehensive approach to model analysis because they evaluate the effect of one factor while varying all other factors, efficiently exploring the multidimensional input space (Campolongo et al., 1999, 2011). A wide range of GSA methods are available (Saltelli et al., 2000, 2005, 2006, 2008; Helton et al., 2006; Campolongo et al., 2011) and range from qualitative screening methods (Morris, 1991; Campolongo et al., 1999, 2007, 2011; Saltelli et al., 2009) to quantitative techniques based on variance decomposition (Cukier et al., 1978; Sobol', 1993, 2001; Homma and Saltelli, 1996; Saltelli et al., 1999, 2010; Oakley and O'Hagan, 2004; Xu and Gertner, 2011). The Fourier amplitude sensitivity test (FAST) (Cukier et al., 1978) and Sobol' methods (Sobol', 1993) are the most popular and widely investigated variance decomposition-based methods (Homma and Saltelli, 1996; Saltelli and Bolado, 1998; Ratto et al., 2001; Francos et al., 2003; Cariboni et al., 2007; Cibin et al., 2010). However, the FAST method does not efficiently address higher-order interaction terms (Saltelli and Bolado, 1998; Cibin et al., 2010). By contrast, the Sobol' method can estimate the interactions between the parameters and the total sensitivity index of individual parameters (Sobol', 1993, 2001). Although the Sobol' method has been applied in many fields of science and engineering, its application in hydrology has been very limited (Pappenberger et al., 2006, 2008; Tang et al., 2007a,b; Cloke et al., 2008; Cibin et al., 2010). A shortcoming of GSA methods is their high computational demands (Hamby, 1994; Moore and Ray, 1999; Ascough et al., 2005; Makler-Pick et al., 2011). Therefore, in this paper, we use a response surface model (RSM) to construct a statistical simulator for the distributed hydrological model. Furthermore, an uncertainty quantification toolkit called PSUADE (Problem Solving environment for Uncertainty Analysis and Design Exploration, see the Appendix) is used to generate the emulators to quantify the parameter sensitivities.

The remainder of this paper is organized as follows: Section 2 contains a brief description of sensitivity analysis methods, such as the Morris screening method, response surface method and RSMSobol' method and describes the fundamentals of the distributed time-variant gain model (DTVGM). A case study of the Huaihe River Basin with the available data, model parameters and evaluated criteria are described in Section 3. Subsequently, Section 4 illustrates and discusses the sensitivity of the DTVGM parameters based on the statistical emulator. Some conclusions of the study are discussed in Section 5.

2. Material and methods

2.1. Integrated approach for efficient sensitivity analysis

An efficient integrated approach is proposed to analyze the sensitivity of hydrological model parameters in four steps: 1) constructing a complete description of the input parameters, 2) performing a down-select screening analysis on all uncertainty parameters, 3) constructing an approximate model using the response surfaces (also known as surrogate functions and emulators) for a complex hydrological model, and 4) performing quantitative sensitivity analysis via variance decomposition techniques. The details are as follows:

2.1.1. Morris screening method

The Morris method (also called elementary effect method) has been proposed as a screening method to identify a subset of inputs that have the greatest influence on the outputs (Morris, 1991). It is a simple but effective way of screening a few important input factors among the many that can be contained in a model (Saltelli et al., 2008), which is based on replicated and randomized "one-at-a-time" (OAT) design, and the detail introduction of the OAT design can see the work of Morris (1991).

An elementary effect is defined as follows. Consider a model with n independent inputs X_i, $i = 1, 2, ..., n$, which varies in the n-dimensional unit cube across p selected levels (Saltelli et al., 2008). For a given value of X, the elementary effect of the ith input factor is defined as

$$d_i(X) = \frac{f(X_1, ..., X_{i-1}, X_i + \varDelta, X_{i+1}, ..., X_n) - f(X_1, ..., X_{i-1}, X_i, ..., X_n)}{\varDelta} \quad (1)$$

where $\varDelta$ is a value in $\{1/(p-1), 2/(p-1), ..., 1 - 1/(p-1)\}$, p is the number of levels, and $X = (x_1, ..., x_{i-1}, x_i, ..., x_n)$ is a random sample in the parameter space so that the transformed point $(x_1, ..., x_{i-1}, x_i + \varDelta, ..., x_n)$ is still within the parameter space.

Morris proposed two sensitivity measures to analyze the data: μ which estimates the overall effect of each input on the output, and σ which estimates the higher order effects such as nonlinearity and interactions between inputs (Tong and Graziani, 2008). To estimate these measures, Morris (1991) suggests sampling R elementary effects for each input by randomly sampling R point $X^{(1)}$, $X^{(2)}$, ..., $X^{(R)}$ to ensure that there are enough regions in the design space. Campolongo et al. (2007) proposed an improved measure, μ^* in place of μ, with the following formulas:

$$\mu_i^* = \frac{1}{R} \sum_{j=1}^{R} \left| d_i(X^{(j)}) \right| \quad (2)$$

$$\sigma_i = \sqrt{\frac{1}{R-1} \sum_{j=1}^{R} \left[d_i(X^{(j)}) - \frac{1}{R} \sum_{j=1}^{R} d_i(X^{(j)}) \right]^2} \quad (3)$$

If μ_i^* is substantially different from zero, then input i has an important "overall" influence on the output. A large σ_i implies that input i has a nonlinear effect on the output or that there are interactions between input i and the other inputs (Tong, 2008).

2.1.2. Response surface analysis

A response surface model (RSM), also known as a meta-model or surrogate model, is a collection of statistical and mathematical techniques that are useful for developing, improving, and optimizing processes (Meyers and Montgomery, 2002). The choice of RSM for a given computational model depends on the knowledge of the computational model itself. The software PSUADE provides a number of response surface methods, ranging from parametric regression methods to nonparametric methods such as Friedman's multivariate adaptive regression splines

(MARSs) method. In this study, the MARS approximation method is used to create the surrogate model for an original hydrological model (DTVGM).

The MARS model is based on a nonparametric modeling approach that requires mild assumptions about the form of the relationship between the predictor and dependent variables (Friedman, 1991). As a consequence, the MARS model can be used to characterize relationships between explanatory and response variables that are difficult, if not impossible, to reveal by other regression methods (Balshi et al., 2009). MARS modeling partitions the parameter hyperspace of explanatory variables into disjoint hyper-regions. Within each of these hyper-regions, a linear relationship is used to characterize the impact of explanatory variables on the response. The point at which the slope changes among hyper-regions is called a knot, and thus the crux of the MARS technology is to automatically and adaptively choose the number and the locations of the knots. Once the knots have been determined, basis functions representing either single variable transformations or multivariable interactions are constructed for each region. The basis functions and the corresponding coefficients are then examined to extract main effect information and detect interactions. Friedman (1991) has given a more detailed and technical study of the MARS model.

2.1.3. RSMSobol' method

The RSMSobol' method, which is based on the response surface method and the Sobol' method, is proposed to analyze first-order, second-order and total sensitivity indices. The implementation of the RSMSobol' method involves the following steps: (1) Determine what output is optimized, what input parameters should be adjusted, and what the ranges of the input parameters should be; (2) Choose an appropriate design of experiments (i.e., quasi-random sequence sampling method, also known as LP-τ, or LPTAU (Sobol', 1967)) for generating the parameters samples; (3) Run the original distributed hydrological model (i.e., DTVGM) to obtain the response objective functions for these samples; (4) Choose an approximation emulator (i.e., the MARS method) to generate the RSM; (5) Use the RSM generate new parameters samples based on an appropriate sampling design (i.e., Sobol' sequence sampling method); (6) Run the Sobol' method using these samples from the RSM to obtain the sensitivity indices of input parameters. The following provides some theoretical description of the Sobol' method:

The variance-based methods use a variance ratio to estimate the importance of input factors. These methods are based on the partitioning of the total variance of model output $V(Y)$ using the following equation:

$$V(Y) = \sum_{i=1}^{n} V_i + \sum_{i \leq j \leq n}^{n} V_{ij} + \cdots + \sum_{i \leq \ldots n}^{n} V_{1\ldots n} \qquad (4)$$

where V_i represent the first order effect for each factor X_i ($V_i = V[E(Y|X_i)]$) and V_{ij} ($V_{ij} = V[E(Y|X_i, X_j)] - V_i - V_j$) to $V_{1\ldots n}$ the interactions among n factors.

The first-order sensitivity index S_i can be calculated by

$$S_i = V_i/V(Y) = V[E(Y|X_i)]/V(Y) \qquad (5)$$

And the second-order sensitivity index S_{ij} can be calculated by

$$S_{ij} = \frac{V_{ij}}{V(Y)} = \frac{V[E(Y|X_i, X_j)] - V_i - V_j}{V(Y)} \qquad (6)$$

In general, the total sensitivity index can be defined as (Homma and Saltelli, 1996; Saltelli and Tarantola, 2002; Saltelli and Annoni, 2010)

$$S_{Ti} = E(V(Y|X_{\sim i}))/V(Y) \qquad (7)$$

where the subscript $\sim i$ refers to all of the inputs except input i. If the inputs are correlated, the variance decomposition is no longer valid. However, Eq. (7) is still a valid measure of total sensitivity (Tong and Graziani, 2008). In this case, we can compute the total sensitivities by evaluating the integrals corresponding to the partial variances of all the inputs based on the response surface models in PSUADE software.

In addition, we also use the McKay (1995) proposed method (used as direct method in the paper, and the direct McKay method was used for validation only, to compare with the results of the RSMSobol' method) to evaluate the first-order sensitivity indices for the DTVGM parameters in the PSUADE platform (can see the Appendix B McKay proposed method). Note that the choice of sampling method has implications for what type of sensitivity analysis the user is able to perform. The sampling method should be compatible with the McKay approach of first-order sensitivity analysis in the PSUADE, that is, the Latin hypercube samples (LHS) are used to estimate the first-order sensitivity indices using the direct method. According to the work of Sobol' (1967), the LP-τ sampling design is designed for numerical integration computations and generally performs better than the LHS. We will enhance the implementation of the approach of McKay for the direct method in the PSUADE using the quasi-random sequence sampling method in a future work.

2.2. DTVGM

The DTVGM, which was developed based on the time-variant gain model (TVGM) (Xia et al., 1997; Xia, 2002), is used in this paper. The DTVGM has been successfully applied to many river basins in China (Xia et al., 2005; Wang, 2005; Wang et al., 2009; Ye, 2007; Li et al., 2009, 2010; Song et al., 2012c). The modeling system includes multiple components of hydro-information analysis and modeling such as input data processing, runoff generation on each sub-basin, and flow routing between adjacent sub-basins. The model characteristics include the following: (1) the ability to describe time-space variations of rainfall and evapotranspiration based on DEM and spatial digital information, and (2) the ability to combine the runoff generation process and flow routing process together by soil moisture content and to perform the hydrological simulation based on sub-basins.

2.2.1. Water balance model

Runoff generation occurs at each unit element with three layers in the vertical direction, involving surface runoff, interflow runoff and groundwater runoff. The water balance equation in this model is expressed as follows:

$$P_i + AW_i = AW_{i+1} + RS_i + ET_{ai} + RI_i + RG_i \qquad (8)$$

where P is the precipitation [mm], AW is the soil moisture content [mm], ET_a is the actual evapotranspiration [mm], and RS, RI and RG are surface runoff, interflow runoff and groundwater runoff [mm], respectively. The subscripts i and $i + 1$ represent variables at time steps i and $i + 1$.

2.2.2. Actual evapotranspiration

The actual evapotranspiration is calculated by:

$$\frac{ET_a}{ET_p} = f\left(\frac{AW}{AWM}, \frac{P}{ET_p}\right) \qquad (9)$$

$$\frac{ET_a}{ET_p} = \left[(1 - Kaw)KET + Kaw\frac{AW}{AWM}\right] \qquad (10)$$

where ET_p is the potential evapotranspiration [mm], AWM is the saturated soil moisture content [mm], Kaw is a coefficient between 0 and 1, KET is a linear or nonlinear function of P/ET_p, and $f()$ is a linear or nonlinear function.

The soil moisture content can be calculated considering the thickness of soil and the volume ratio of soil moisture content as follows:

$$AW = Thick*W \qquad (11)$$

$$AWM = Thick*WM \qquad (12)$$

where $Thick$ is the thickness of the soil [mm] and W and WM are the volume ratio of the soil moisture content and saturated soil moisture content, respectively.

2.2.3. Runoff method

There are three layers in the DTVGM: the vegetation layer, surface soil layer and deep soil layer. The three runoff components are the surface runoff on the land surface, interflow runoff from the surface soil layer and base flow (groundwater runoff) from the deep soil layer (Ye et al., 2010), as shown in Fig. 1.

The surface runoff (RS_t) generated in a sub-basin is calculated by

$$RS_i = g_1 \cdot \left(\frac{AW_{u,i}}{WM}\right)^{g_2} \cdot P_i \qquad (13)$$

where g_1 and g_2 are the time-variant gain coefficients, g_1 is the runoff coefficient when the soil moisture content is equal to the saturated soil moisture, and g_2 is the impact coefficient of the soil moisture content. The subscript u represents the veneer of the soil.

The interflow runoff (RI) is calculated with the assumption that the interflow runoff is proportional to the soil moisture content, while the veneer soil moisture content is larger than the field moisture capacity. Therefore, RI can be calculated by a linear storage–outflow relationship (Lee, 2007; Wang et al., 2009):

$$RI_i = \frac{AW_{u,i} + AW_{u,i+1}}{2} \cdot K_r \qquad (14)$$

where K_r is the flow coefficient of soil water and is related to soil properties and terrain feature, such as soil particle size, soil layer thickness, soil clearance, and grade of slope.

The groundwater runoff (RG) is derived by the equation:

$$RG_i = AW_{d,i} \cdot K_{rg} \qquad (15)$$

where K_{rg} is groundwater runoff coefficient, and the subscript d is the deep layer of soil. The deep subsoil water can be computed by
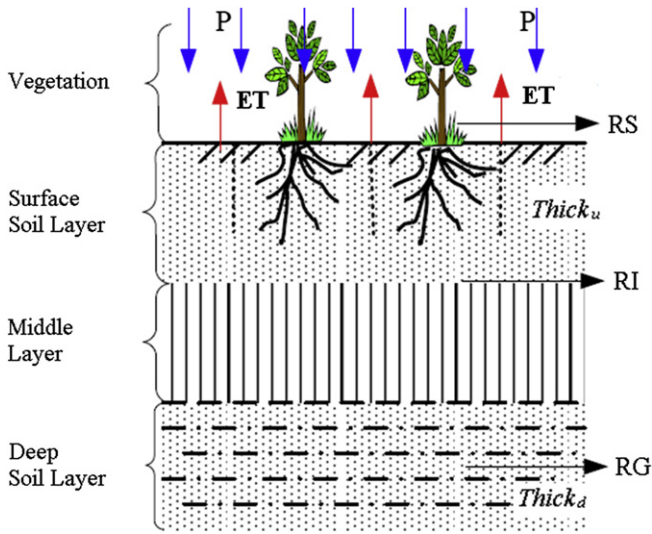
Fig. 1. The runoff generation process in the DTVGM (Ye et al., 2010).

$$AW_{d,i} = AW_{d,i-1} + fc \cdot \Delta t \tag{16}$$

where fc is the infiltration rate from surface soil to deep subsoil [mm/h] and the Δt is the computational time interval [h].

The total runoff is as follows:

$$R = RS + RI + RG \tag{17}$$

2.2.4. Routing model

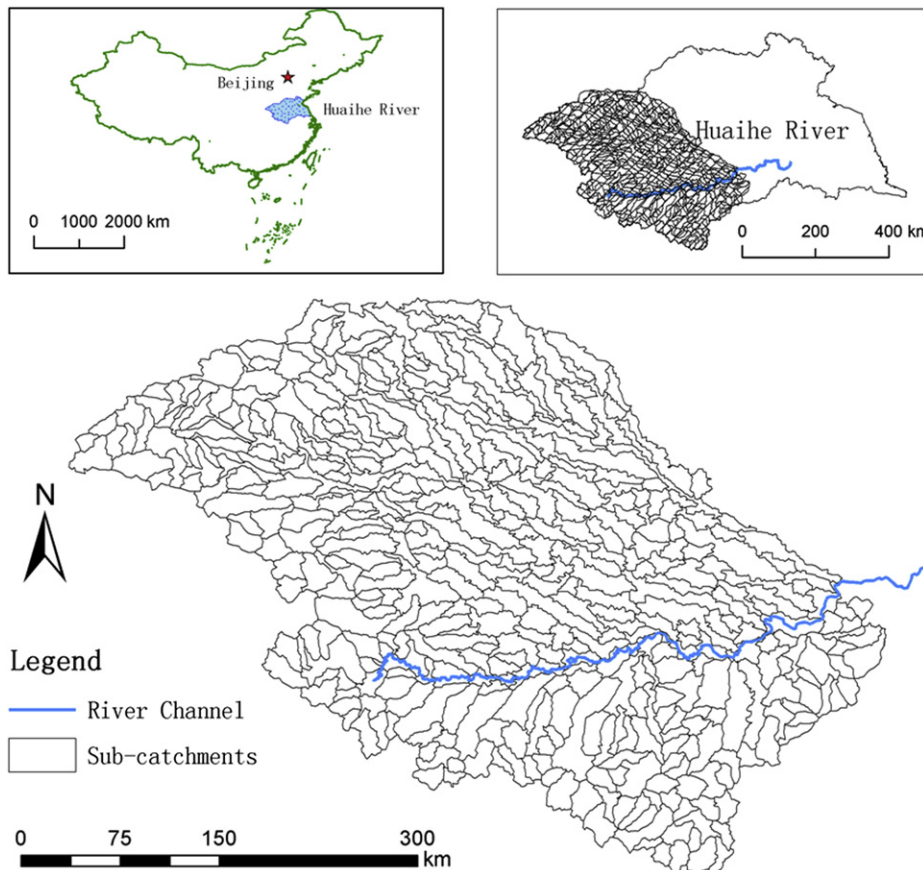The kinematic wave method is used to calculate the routing runoff. The friction term in the momentum equation is ignored, assuming that the friction slope is equal to slope (S_0) and the river flow is unsteady open channel gradual change flow (Ye, 2007). The continuity equation is as follows:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = q \tag{18}$$

where A is river cross-sectional area of the river [m^2], t is time [s], Q is discharge [m^3 s^{-1}], x is flow path [m] and q is later inflow [m^2 s^{-1}]. In addition, Q can be calculated by

$$Q = A \cdot h^{2/3} \cdot S_0^{1/2} / n \tag{19}$$

where n is the Manning roughness coefficient (notated as *RoughRss* in Table 1 to distinguish from the number of input n) and h is the average depth of the cross section [m].

3. Study area and data

3.1. Study area

The Huaihe River basin (see Fig. 2) is one of the seven largest rivers in China, and flows from west to east, neighboring with the Yellow River in the north and the Yangtze River in the south. Its upper reaches are located in Henan, its middle reaches are located in the Anhui province, and its lower reaches are located in the Jiangsu province. Its trunk is approximately 1000 km long with 120 main tributaries and a catchment area of approximately 185,700 km^2. The Huaihe River basin is mostly plain with a very complicated water system, including a large number of tributaries, many inter-provincial rivers, and numerous artificial river water controls.

The average annual precipitation in the basin is approximately 900 mm, of which 70–80% occurs in summer. Therefore, there is

Fig. 2. Study area and location of the Huaihe basin in China.

a considerable variability in the river flow. The average annual flow is 853 m^3/s, with a flooding discharge of greater than 11,000 m^3/s; the discharge drops to nearly zero in the dry season. There are four main flood control gates on the Huaihe River, approximately 4300 sluices and more than 5000 reservoirs.

In this paper, the daily precipitation, evaporation, and discharge data from the 174 weather stations and 32 hydrological stations from 1991 to 2000 are used to construct a DTVGM for the Huaihe River basin. The Huaihe basin is divided into 441 sub-basins using the DEM data, and the gradient plus inverse-distance-squared method is used for interpolation in the study (Ye, 2007).

3.2. Model parameters

The most important and challenging task in this work occurs in the first step, i.e., the prior knowledge of parameter information, because the proper prescription of the ranges and shapes of the input distributions can dramatically alter the outcome of the analysis (Tong and Graziani, 2008), and oftentimes these ranges are obtained by carefully analyzing data from physical experiments. For large-scale multi-physics processes applications, physical experiments for all of the multi-physics processes may not be feasible. For this reason, the ranges and shapes of the input distributions in this work are obtained by carefully analyzing data from physical experiments and from the prior work of Ye (2007), Li et al. (2009, 2010) and Wang et al. (2010). The parameter ranges of each input and control factor are listed in Table 1. Because the parameter probability density functions of most of the model parameters are unknown (Freer et al., 1996; Manache and Melching, 2008), and some parameters are correlated with others, we instead use uniform distribution and no correlation for all parameters from the first step study.

3.3. Evaluation criteria

The performance of the developed models can be evaluated using several statistical tests that describe the errors associated

Table 1
List of parameters and ranges for the DTVGM.

Parameter	Description	Ranges	Unit
g_1	Coefficient of time-variant gain factor, related to surface runoff generation	[0.05, 0.50]	–
g_2	Coefficient of time-variant gain factor, related to soil moisture content	[1.0, 5.0]	–
K_r	Storage–outflow coefficient related to interflow runoff generation	[0.01, 1.0]	–
K_{rg}	Storage–outflow coefficient related to groundwater runoff generation	[0.50, 1.0]	–
fc	Infiltration rate from surface soil to deep subsoil	[0.10, 5.0]	mm/h
Kaw	Coefficient for calculating actual evapotranspiration	[0.01, 1.0]	–
$RoughRss$	Roughness coefficient of Manning's formula	[0.001, 0.1]	–
Wmi	Minimum soil moisture storage	[0.01, 0.40]	mm/mm
WM	Upper layer saturated soil moisture storage	[0.40, 1.0]	mm/mm
WMd	Deep layer saturated soil moisture storage	[0.01, 1.0]	mm/mm
$Thick_u$	Thickness of the upper soil layer	[400, 800]	mm
$Thick_d$	Thickness of the deep soil layer	[400, 800]	mm
AW	Soil initial moisture	[0.01, 0.20]	mm/mm
AWd	Subsoil initial moisture	[0.01, 0.20]	mm/mm

Note: In Section 4.3, when the normal distribution was used to analyze the first-order sensitivity index, the mean (the first number in the parentheses) and standard deviation (the second number in the parentheses) of the six parameters are the following: g_1 (0.3, 1.5), g_2 (1, 0.8), Kaw (0.1, 2), $RoughRss$ (0.01, 5), WM (0.5, 2), and $Thick_u$ (600, 20), respectively.

with the model (Song et al., 2012a). After each of the model structures is calibrated using the calibration/testing data set, the performance can then be evaluated in terms of these statistical measures of goodness of fit. In this study, three objective functions and evaluation criteria, such as the water balance coefficient (WB), Nash–Sutcliffe coefficient (NS) and correlation coefficient (RC) are used to evaluate the model efficiency.

$$\text{WB} = \frac{\sum_{i=1}^{n} Q_{s,i}}{\sum_{i=1}^{n} Q_{o,i}} \tag{20}$$

$$\text{NS} = 1 - \frac{\sum_{i=1}^{n} \left(Q_{o,i} - Q_{s,i}\right)^2}{\sum_{i=1}^{n} \left(Q_{o,i} - \overline{Q}_o\right)^2} \tag{21}$$

$$\text{RC} = \frac{\sum_{i=1}^{n} \left(Q_{o,i} - \overline{Q}_o\right)\left(Q_{s,i} - \overline{Q}_s\right)}{\sqrt{\sum_{i=1}^{n} \left(Q_{o,i} - \overline{Q}_o\right)^2 \sum_{i=1}^{n} \left(Q_{s,i} - \overline{Q}_s\right)^2}} \tag{22}$$

where Q_s and Q_o are the simulated value and observed value for the runoff, and $\overline{Q}$ is the mean value for the corresponding runoff.

In this work, we use the selected parameters as inputs and the objective functions as outputs for PSUADE to perform uncertainty quantification.

4. Results and discussion

4.1. Morris screening

The Morris screening analysis is used to identify qualitatively important parameters with respect to the responses of interest. The Morris method must define the number of levels p, with a value of p normally within the range of [4, 10] (Yang, 2011). Morris (1991) and Saltelli (1999) elaborated the selection of the level number p and the sample size, which is closely related to the computational cost (Tong and Graziani, 2008). The paper gives the assumption that all the inputs are uniformly distributed in the range of the parameters, and thus the number of replications R for Morris screening is set as 40, the level $p = 10$, and the sample size $N = 600$.

The Morris screening measures can be shown graphically by screening plots, which have as the x- and y-axes the modified means μ^* and standard deviations σ, respectively, as shown in Fig. 3. Each input is represented on the screening plot by a point with the coordinate (μ_i^*, σ_i). A relatively large magnitude (modified mean values μ^*) for a given input indicates its importance. Fig. 3 shows that, Kaw and $RoughRss$ are the most sensitive parameters and g_1, g_2, WM and $Thick_u$ are the second most sensitive, while the others are relatively less sensitive for all the objective function responses of interest (WB, NS and RC). The rankings for the different objective function responses are slightly different, e.g., Kaw is the most sensitive parameter for the water balance coefficient WB and $RoughRss$ is the most important for the Nash–Sutcliffe coefficient NS and correlation coefficient RC. These results agree well with the conclusions from the previous studies (e.g., Li et al., 2009; Wang et al., 2010) in which the GLUE method and regional sensitivity analysis (multi-parameter sensitivity analysis, MPSA) are applied to the sensitivity analysis of the parameters g_1, g_2, Kaw, WM, K_r, Wmi, and AW. However, this paper selects 14 parameters for sensitivity analysis with a view to multi-objectives and GSA. In general, the Morris screening method has been used as an effective screening measure to identify the few important factors in models with many factors, and it shares many of the positive qualities of the variance-based techniques, with the advantage of a lower computational cost (Campolongo et al., 2007). For this work, the model run
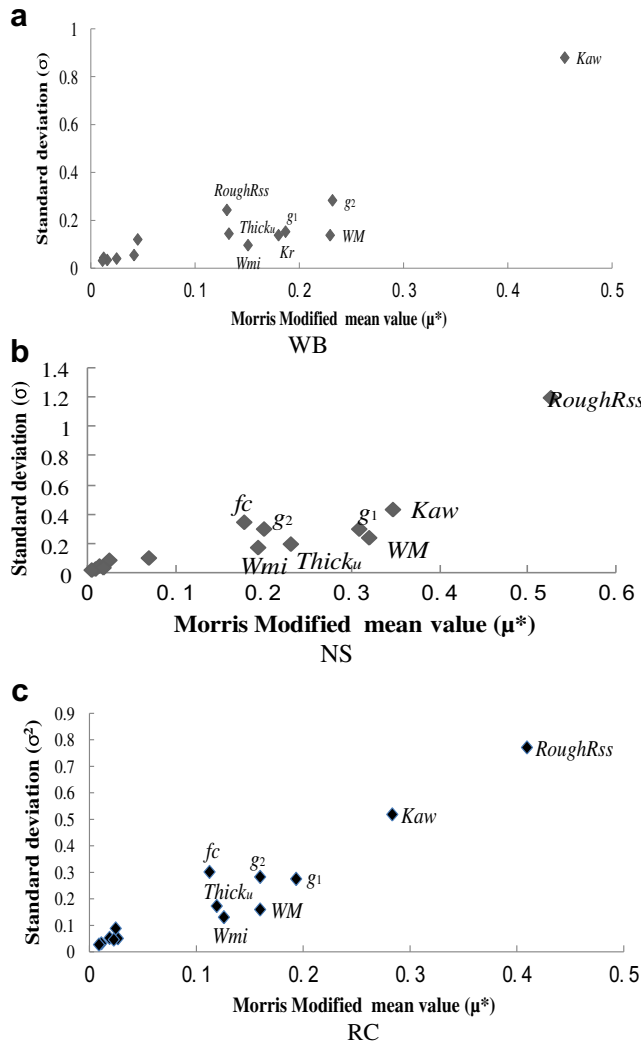
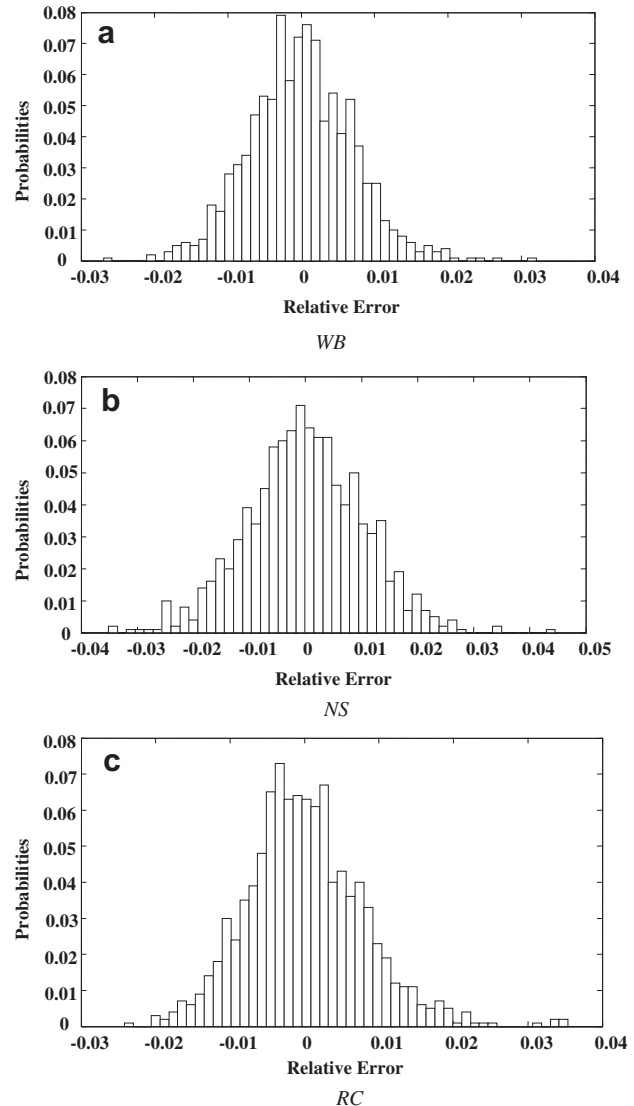**Fig. 3.** Scatter plots for 3 responses of interest.

Fig. 4. Histogram of prediction errors for the 3 responses of interest.

required approximately 5 h for $N = 600$, while the run times may be approximately 2 months for $N = 200,000$ based on the GLUE method (Li et al., 2009) and 3 days for $N = 10,000$ based on MPSA (Wang et al., 2010) if their methods are used in this study.

The scatter plots, in addition to giving detailed information on the standard deviation of the elementary effects, also aid in identifying anomalies as well as separating nonlinearities from interactions. Fig. 4 shows that the parameter *Kaw* is most important with respect to the objective function response WB, and the largest standard deviation implies that their non-linear relationship is more obvious or the interaction of *Kaw* with other parameters is more significant. Similarly, the same results are observed for *RoughRss* with respect to the responses NS and RC. In addition, g_1, g_2, *WM*, $Thick_u$, K_r and *fc* are relatively important and demonstrate relatively strong nonlinearities and/or interactions.

Generally, the main limitations of the Morris screening method include the following: it cannot provide an accurate quantitative estimation of how much a factor contributes to the output variability, and it cannot distinguish the non-linearity of a factor from the interaction with other factors, i.e., the nonlinearities and parameter interactions are confounded (Yang, 2011). Campolongo and Braddock (1999) proposed the extended Morris method to detect interactions. However, the number of model evaluations is $O(n^2)$, where n is the number of model input factors (Tong and

Graziani, 2008). The Morris screening method in PSUADE has been enhanced to provide a qualitative assessment of parameter nonlinearities without additional cost by taking advantage of Morris's multiple visits to individual inputs levels. When the number of replications R is 40, then it is likely that each of the p levels is used more than once. Then, a rough assessment of interactions can be made by comparing the elementary effects on the same levels (Tong and Graziani, 2008).

The Morris screening experiments, which are used to identify the inputs that have little effect on the output variability, are computationally affordable techniques for computationally expensive models that often have a large number of uncertain input parameters. These non-sensitive inputs can therefore be fixed in subsequent analyses without significantly modifying the prediction characteristics of the model (Lilburne and Tarantola, 2009). Therefore, the six parameters (i.e., g_1, g_2, *Kaw*, *RoughRss*, *WM*, and $Thick_u$) are selected for the subsequent quantitative sensitivity analysis.

4.2. Response surface analysis

As stated above, the response surface method is an effective approach to reduce the computational cost of model evaluation.

Some prerequisite should be prepared to construct a reliable and accurate response surface model for general applications, such as a suitable sampling method, a good interpolation scheme, and effective validation approaches (Tong and Graziani, 2008). The process consists of four basic steps: experimental design, model selection, model fitting, and model validation (Hsieh, 2007). Given the choice of a sampling method, a natural yet difficult question is which simulation run is appropriate. A small sample may not yield accurate statistical data, while a large sample will unnecessarily consume computing resources (Tong, 2006). Moreover, the choice of the sampling design is a difficult decision. Sobol' and Shukhman (1995) conclude that none of the sampling design techniques is superior to the others. However, they also demonstrate that the main advantage of a quasi-random sequence sampling design approach is its potential for improving the convergence characteristic. Therefore, the quasi-random sequence sampling design technique (Sobol', 1967; Statnikow and Matusov, 2002) is used in the paper. The MARS model has some dominant characteristics: it requires little data preparation, is suitable for handling fairly large datasets with a tendency toward good bias-variance trade-off, and is flexible enough to model non-linearity and variable interactions. Therefore, the MARS method is used as a function approximation to generate the response surface model (surrogate model). In the MARS method based on the PSUADE platform, the number of basis functions is 2500 and the number of knots is 500. A mix sampling design was used to construct the response surface because Latin hypercube/quasi-random sequence samples are good space-filling designs, while factorial/fractional factorial are good for covering the corners and edges (Tong, 2008). First, 2000 samples for the six most important parameters (g_1, g_2, Kaw, $RoughRss$, WM, and $Thick_u$) based on Morris screening are generated to construct and validate the response surface model, in which 1000 samples are used as the training set to construct model and the other 1000 samples are used to validate it. The total computational cost is approximately 15 h for 2000 model runs, and the time for constructing the response surface (nearly 5–10 min) is far shorter than that of the model runs.

Another vital and indispensable step is to validate and test the response surface. The two approaches (cross-validation based on the first 1000 samples and experimental evaluation based on the other 1000 samples) are applied to check the responses for goodness of fit by examining the interpolation error and prediction error. The k-fold, leave-one-out and repeated random sub-sampling validation are the common types of cross-validation. However, leave-one-out cross-validation is usually very computationally expensive because of the large number of times that the training process is repeated, and the disadvantage of repeated random sub-sampling validation is that some observations may never be selected in the validation subsample, whereas others may be selected more than once (in other words, validation subsets may overlap). Consequently, the k-fold cross-validation method is used in this work. We select the number of groups $k = 500$, that is, we divide the sample into 500 groups and hold out one group at a time and compute the prediction error statistics. The maximum relative errors are 0.032 for WB response of interest, 0.045 for NS response of interest and 0.035 for RC. The relative root mean square error (RMSE) is 0.00714, 0.191 and 0.00792 for the three responses, respectively. The error histograms for all of the responses are shown in Fig. 4. The results demonstrate that all of the responses of interest give acceptable interpolation errors, i.e., the sample is adequate to represent the model input–output relationship.

In addition, the other 1000 quasi-random sequence samples are used as an experiment test to further validate the response surface models. Fig. 5 depicts a scatter plot between the response surface model simulated objective function values and the corresponding DTVGM-generated values. Note that the objective function values (WB, NS and RC) in these figures correspond to the values from the MARS model. The data points do not deviate greatly from the 1:1 line (solid line in the plot) for the three objective functions, which is evident from the high R^2 value of 0.995, 0.971, and 0.982 for WB, NS and RC, respectively. The plots exhibit a very intense scatter, implying that the response surface models or surrogate models mimic the performance of DTVGM fairly well.

For surrogate models with sufficient prediction capabilities, the bias due to the use of the surrogate models instead of the true model is negligible (Marrel et al., 2009). Once the sample is deemed
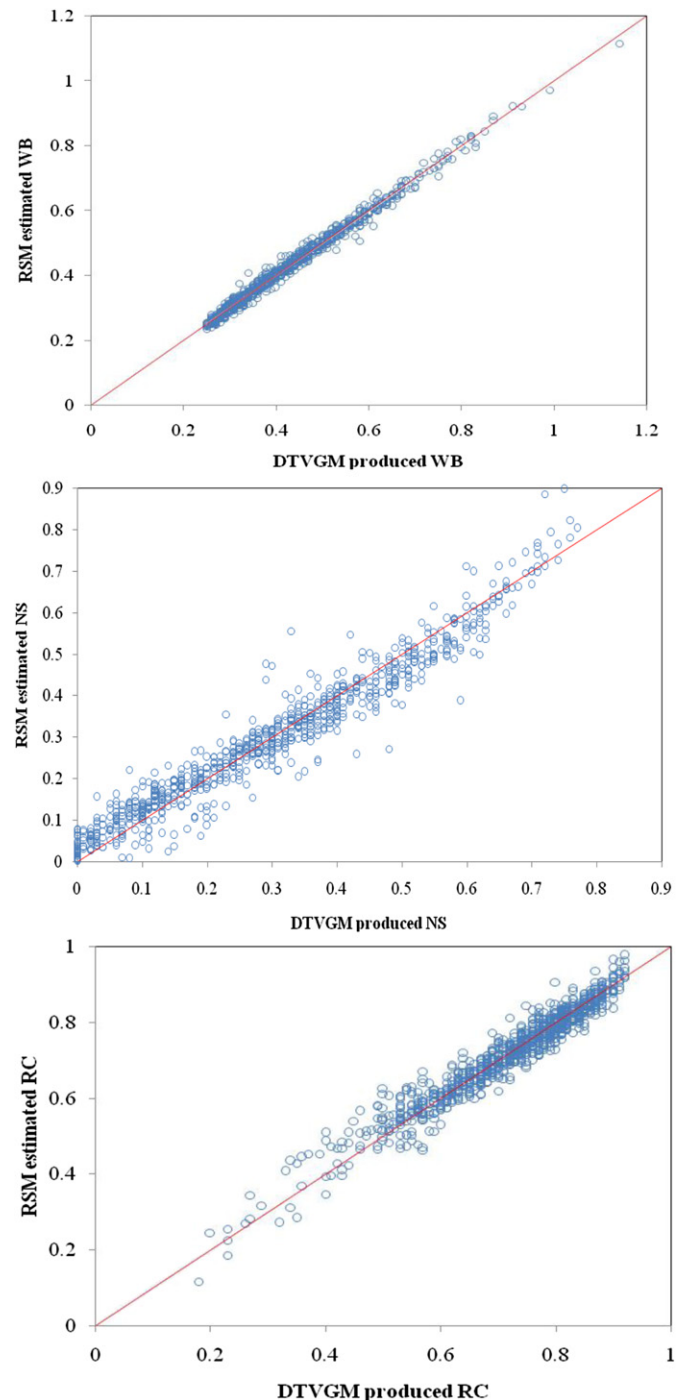

Fig. 5. Comparison between the response surface model and DTVGM.

satisfactory, subsequent analysis can rely on this response surface model which is inexpensive to evaluate. Space constraints permit some representative response surface models to be provided. The three parameters g_1, Kaw and WM are selected to construct the response surface with respect to the three output objective function responses. The response surface plots for the three responses of interest are as shown in Fig. 6.

4.3. Results of sensitivity analysis

The proposed approach (RSMSobol' method) was used to estimate the first, second and total order sensitivity indices of the DTVGM parameters. The surrogate models used are inexpensive and tractable to evaluate, and the first-order sensitivity indices for the 6 parameters of the DTVGM were computed by sampling 100,000 times from the response surface based on the quasi-random sequence sampling design method. The normalized results of the first-order sensitivity index are shown in Table 2, and indicate that the soil moisture parameter WM is the most sensitive, while Kaw, g_1 and $Thick_u$ are relatively sensitive parameters and the parameter g_2 and $RoughRss$ are less sensitive for the first two objective function responses of interest. For the RC response of interest, the concentration parameter $RoughRss$ is the most sensitive, and the parameters WM, Kaw and g_1 are of second-most importance. Table 2 also shows that the sum of the first-order sensitivity indices for all the objective function responses of interest is less than 1, i.e., there are some interactions between the parameters.

Table 2 shows the results, which are in agreement with the results reported by Li et al. (2009) and Wang et al. (2010). Li et al. (2009) stated the ranking of the parameters sensitivity with the conclusion that g_1, WM and g_2 are the first-class ranking, Kaw and K_r are the second-class ranking, and Wmi is non-sensitive based on the GLUE approach with a total of 200,000 runs of a monthly scale DTVGM. However, Li et al. (2009) merely provided the qualitative ranking results for the six parameters with an extensive computational cost. In addition, Wang et al. (2010) also analyzed the five parameters for a monthly scale DTVGM using the multi-parameter sensitivity analysis (MPSA) method, and reported that the parameter Kaw for estimating actual evapotranspiration is the most sensitive. For a daily DTVGM, the computational cost of each run is larger than that of the monthly DTVGM due to the huge data and complex routing calculation. Moreover, the proposed approach can efficiently estimate the quantitative sensitivity indices for a complex hydrological model.

The sensitivity outcomes usually depend on both the model structure (e.g., time scale) and the parameter distributions (Elsawwaf et al., 2010). Undoubtedly, challenges remain in how to address the dependence of sensitivity on parameter distributions. For this reason, the normal distribution is applied to compute the first-order sensitivity indices to analyze the effect of the parameters distributions. In this paper, the mean μ and standard deviation σ of the input parameters are selected subjectively based on the previous modeling experience, which may be not well-suited to the real system processes. Fig. 7 shows that the first-order sensitivity indices based on the normal distribution differ from that based on the uniform distribution, while the ranking results are accordant with that of the uniform distribution. Fig. 7 also demonstrates that the parameters WM and Kaw are the most important for all of the responses. Unfortunately, the real distribution type of the input parameters in the regions cannot be known. For the other distributions (e.g., normal distribution, lognormal distribution and t-distribution, etc.), it is difficult to provide the exact value of the mean and standard deviation, which have significant effects on the output objective function responses. In future work, we will explore

the variety of sensitivity indices with respect to non-uniform distribution and ranges of input parameters. For this reason, the subsequent sensitivity analysis (second order and total order) is estimated based on the assumption of uniform distribution in the ranges of input parameters.

The second-order sensitivity indices of the parameters are also estimated based on the RSMSobol' method. The 100,000 sample points based on different objective function responses are used to quantify the sensitivity indices, and the cost time is nearly to 23 h, including 21 h for 2600 runs of the original model, approximately 1 h for constructing the response surface model and approximately 1 h for estimating the second-order sensitivity indices. The second-order sensitivity indices between parameters for different response surface models are shown in Table 3. The proposed method of sensitivity analysis obviously emphasizes the interaction between the parameters. For instance, the soil moisture parameter WM is highly correlated to the parameter $Thick_u$ with a correlation of 0.121 for the objective function WB. In addition, the two-way interactions between WM and other sensitive parameters (e.g., the evapotranspiration parameter Kaw with second-order sensitivity index 0.077 for WB and the routing parameter $RoughRss$ with 0.099, 0.088 and 0.074 for WB, NS and RC, respectively) are relatively larger. The other interactions are quite small for all of the objective function responses. The results illustrate that the proposed method is effective for identifying the interacting parameters in the DTVGM.

To address the correlated inputs or parameters, the total order sensitivity indices are also applied to evaluate the total effect of the model parameters on the different objective function responses (Song et al., 2012c). The results, as shown in Table 4, indicate that the soil moisture parameter WM is most sensitive for all responses of interest. The ranking results show that the saturated soil moisture WM, actual evapotranspiration coefficient Kaw and the runoff coefficient g_1 have significant effects on the WB response of interest, while WM and the Manning's roughness coefficient $RoughRss$ are most important for the NS and RC responses of interest.

As mentioned above, the DTVGM combines the runoff generation process and the flow routing process in the soil moisture content (e.g., WM). The study by Ye (2007) implies that the saturated soil moisture parameter WM influences the runoff generation process and discharge hydrographical type. It is easy to determine that the saturated soil moisture parameter WM affects almost all of the processes outlined in Section 2.2. For the daily-scale DTVGM, the runoff routing process is also important, that is, the routing method parameter $RoughRss$ is relatively sensitive. In particular, the Nash–Sutcliffe coefficient and relation coefficient objective functions, which are used to validate the goodness-of-fit between the observed and simulated discharge, have a significant effect on the output responses. The ranking results also indicate that the two parameters are the most important and sensitive in the daily-scale DTVGM. The parameters g_1 (related to surface runoff generation) and Kaw (coefficient for calculating actual evapotranspiration) are also important for the water balance response because the water balance strongly depends on the runoff (e.g., surface runoff, interflow runoff, and subsurface runoff, in which the surface runoff comprises the greatest proportion of the runoff) and actual evapotranspiration according to Eq. (9). The parameters $Thick_u$ and g_2 are also important. In this study, the ranges of parameters are estimated from the previous study (Ye, 2007) and may be related to the sensitivity of the parameters $Thick_u$ and g_2.

4.4. Validation of RSMSobol' method

The direct method of McKay is also used to estimate the first-order sensitivity indices of all the 14 parameters to validate the
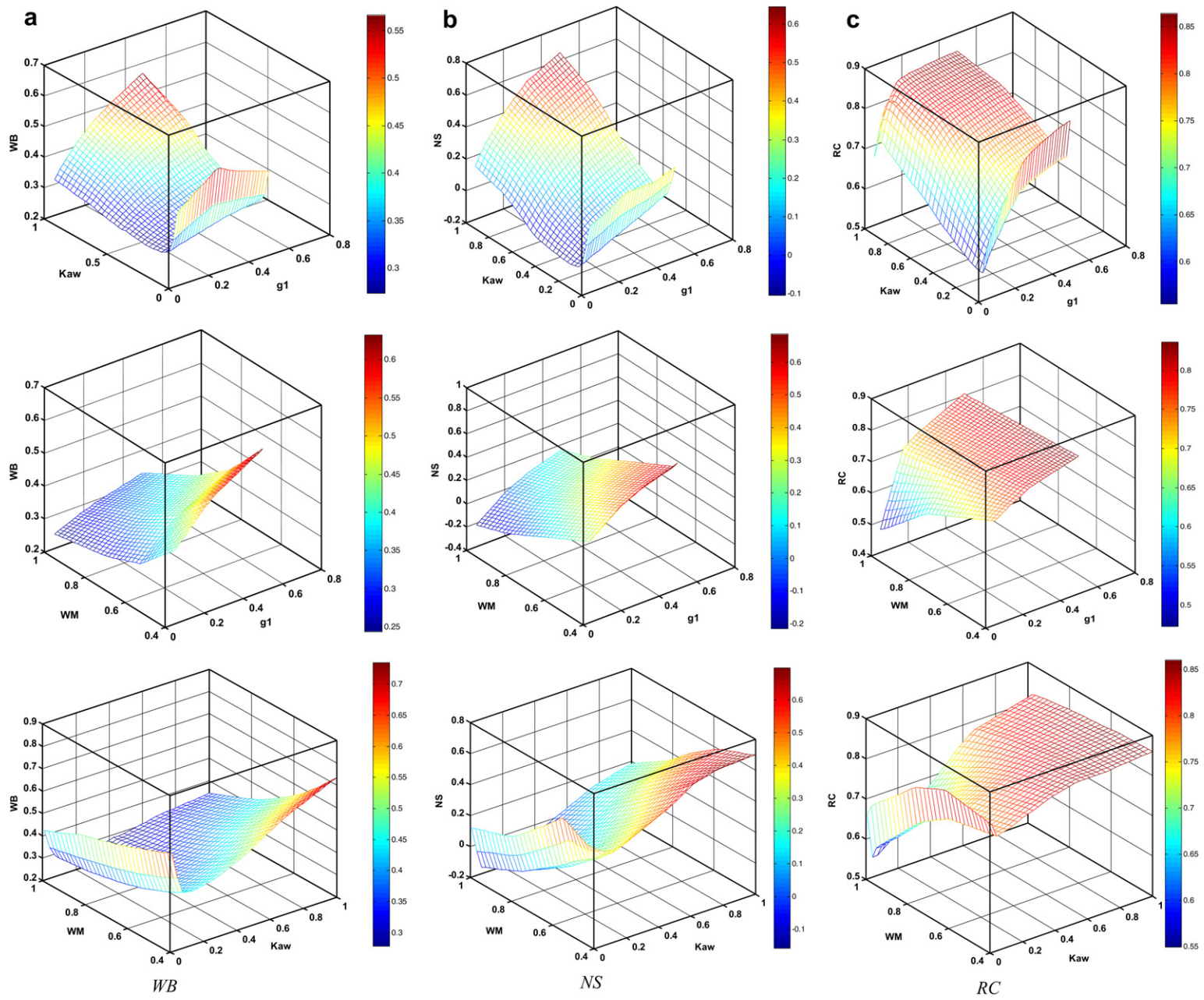
Fig. 6. Response surface plots for the 3 responses of interest with: (left) g_1 and *Kaw*, (mid) g_1 and *WM*, and (right) *Kaw* and *WM*.

Table 2
The first-order sensitivity index for 3 responses of interest.

Parameter	WB		NS		RC	
	RSMSobol1[a]	Main effect[b]	RSMSobol1[a]	Main effect[b]	RSMSobol1[a]	Main effect[b]
g_1	0.157	0.153	0.163	0.133	0.143	0.127
g_2	0.041	0.052	0.024	0.025	0.081	0.048
K_r	–	0.011	–	0.003	–	0.003
K_{rg}	–	0.000	–	0.000	–	0.000
fc	–	0.029	–	0.008	–	0.005
Kaw	0.180	0.056	0.145	0.047	0.122	0.041
$RoughRss$	0.016	0.046	0.057	0.205	0.250	0.350
Wmi	–	0.075	–	0.024	–	0.046
WM	0.288	0.311	0.245	0.146	0.113	0.099
WMd	–	0.023	–	0.020	–	0.009
$Thick_u$	0.139	0.094	0.126	0.042	0.070	0.013
$Thick_d$	–	0.000	–	0.000	–	0.000
AW	–	0.000	–	0.000	–	0.000
Awd	–	0.000	–	0.000	–	0.000
Sum	0.821	0.850	0.760	0.653	0.779	0.741

[a] RSMSobol1: it performs variance-based main effect analysis on a response surface model (with the LP-τ sampling design).
[b] Main effect: it computes first-order Sobol' indices for the replicated Latin hypercube samples (the parameters assumed to be uniform distribution in the work) due to the direct method.

proposed approach. Usually, the two methods should adopt the same sample size. Because the brute-force approach entails huge computational cost, the number of model evaluation is set to 10,000, and the sampling design technique is replicated LHS sampling design because the direct method in PSUADE should be in accordance with the LHS sampling design.

The running time of the RSMSobol' method is approximately 30—40 min including 20 min for estimating the first-order sensitivity index and approximately 10—20 min for constructing and validating the response surface model. The 2600 runs of the model cost approximately 21 h, that is, the total cost is approximately 22 h, while the main effect analysis (direct method) with 10,000 runs for the DTVGM costs 84 h (i.e. nearly 4 days, most of the time cost is incurred in running the DTVGM). Obviously, if we estimate the first-order sensitivity index using the brute-force approach based on the 100,000 samples, the computational cost will be greater (potentially greater than 1 month). In addition, the results of the direct method for all 14 parameters are nearly identical to those of the response surface method for the six screening parameters, implying that the proposed approach is an effective method and can be applied in the complex hydrological model.

In this study, the proposed approach, based on the statistical emulator (or surrogate model), is an effective and efficient technique for analyzing the parameters sensitivities of a complex hydrological model. The total computational cost is approximately

21 h for this proposed approach for 2600 runs of the DTVGM (600 runs for Morris screening, 1000 runs for constructing the response surface and 1000 runs for testing the response surface) and 100,000 runs of the surrogate model, while the total computational cost of the brute-force approach, which was used to compute the main effect using the direct McKay approach with 10,000 runs of the DTVGM, is almost 84 h. Thus, the proposed approach was approximately 4 times faster than the classical approach (or brute-force approach). Assuredly if the interactions and total effects are also estimated by using the brute-force approach, the computational cost would undoubtedly be more expensive.

4.5. Effect of sensitivity analysis in model calibration

The classification of factors (Ratto et al., 2001) consists of three types: (I) factors with a high main effect, (II) factors with a small main effect but high total effect, and (III) factors with a small main and total effect. Type I factors affect model output singularly without a dependence on interactions, while type II factors influence the output mainly through interactions. Type III factors have a negligible effect on the model output and can be regarded as a constant. From the difference between the main effect and the total effect, we observe an interesting phenomenon, i.e., the parameter RoughRss is type III for the WB response, type II for the NS response and type I for the RC response. Thus, these parameters should be calibrated and optimized carefully, particularly for multi-objective calibration. If the non-sensitive parameters are selected for optimization, the calibration of the hydrological model will be more complex and entail a greater computational cost because it is heavily dependent on the number of parameters.

The sensitivity analysis was performed in order to discriminate between insensitive and sensitive model parameters. The information thus obtained can be used as guidelines in a data campaign in the sense that sensitive parameters must be determine (measured) accurately. Therefore, as known to all, parameter identification (or sensitivity analysis) and model calibration are important steps in model building process, to ensure that the results are credible (Song et al., 2012c). For a given model, when the number of parameters is large, the calibration process will be complex with intensively computational cost. However, only few important or sensitive parameters contribute most to the modeling results (Bahremand and De Smedt, 2008). So, it is a critical precondition and requirement for efficient parameter calibration
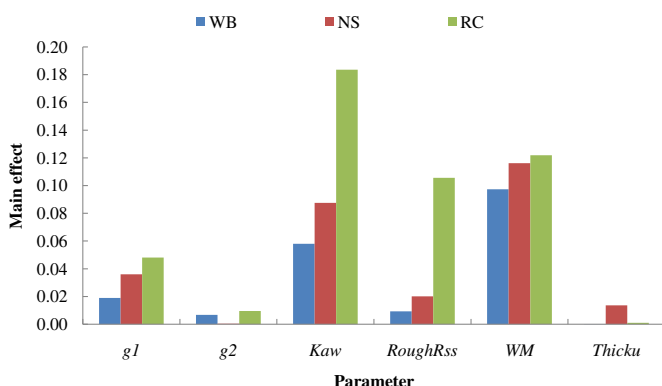

Fig. 7. First-order sensitivity indices for 3 responses based on normal distribution.

Table 3
Second-order sensitivity indices for response surface models.

	g_2			Kaw			$RoughRss$			WM			$Thick_u$		
	WB	NS	RC	WB	NS	RC	WB	NS	RC	WB	NS	RC	WB	NS	RC
g_1	0.048	0.011	0.023	0.048	0.025	0.017	0.002	0.027	0.007	0.031	0.033	0.014	0.022	0.015	0.021
g_2				0.004	0.001	0.03	0.006	0.023	0.041	0.062	0.017	0.006	0.011	0	0.007
Kaw							0.013	0.002	0.015	0.077	0.039	0.006	0.036	0.025	0.031
$RoughRss$										0.099	0.088	0.074	0.015	0.031	0.019
WM													0.121	0.042	0.019

and optimization (Song et al., 2012c). In addition, SA provides useful information regarding the behavior of the simulation model, which includes the identification of relevant model inputs and the information on model construction (Confalonieri et al., 2010). Ratto et al. (2001) also indicated that the quantitative sensitivity analysis can account for conditioning on observations and provide useful information about the model internal structure because the influence of model structure on model output due to the input parameters to a certain extent.

Furthermore, model calibration and optimization has been performed based on the results of sensitivity analysis to estimate the important parameters (see Song et al., 2012b). For those non-sensitive parameters, it is better to use fixed values for the calibration to increase the tractability of the model calibration at lower computational cost. Meanwhile, most models of reasonable complexity include some unidentifiable parameters (e.g., parameters for which sufficient information does not exist in the streamflow observations or other response variables of interest) and that therefore cannot be identified in the calibration process. If these unidentifiable parameters are allowed to vary freely during calibration, the resulting values will have little meaning (since one value is presumably no better than another) (van Werkhoven et al., 2009). Therefore, the SA results may also aid the identification of the model parameters, and make parameter space be effectively searched by current optimization algorithms by reducing the number of model parameters.

Saltelli and Annoni (2010) indicated that sensitivity analysis is particularly useful in pinpointing which assumptions are appropriate candidates for additional data collection to narrow the degree of uncertainty in the results. In addition, we also find that the response surface method as a surrogate model can be used for various purposes: (1) sensitivity analysis, by helping to highlight the most important input factors of the mapping, (2) model simplification, by finding a surrogate model containing a subset of the input factors that account for most of the variability Y, (3) model calibration, in which the surrogate model is used to find directly the optimal parameterization for fulfillment of the given calibration. In a way, it is useful to search the optimal parameter value from the response surface relationship between the model parameters and the output objective functions, and it can be used to optimize the parameters sets in a complex model, which has been widely used in the engineering problems and would be a good practice for hydrological modeling system.

Table 4
Total sensitivity indices for the 3 responses of interest.

Parameter	WB		NS		RC	
	S_{Ti}	Rank	S_{Ti}	Rank	S_{Ti}	Rank
g_1	0.216	3	0.198	4	0.212	3
g_2	0.071	5	0.043	6	0.129	6
Kaw	0.248	2	0.188	5	0.192	4
$RoughRss$	0.024	6	0.241	2	0.357	1
WM	0.365	1	0.386	1	0.234	2
$Thick_u$	0.184	4	0.231	3	0.178	5

5. Conclusions

In this study, an efficient integrated approach is proposed to conduct global sensitivity analysis for distributed hydrological models. This approach integrates the qualitative screening method (Morris method) with the quantitative analysis method (RSMSobol' method based on a statistical emulator) and is implemented with less computational cost to the DTVGM. The results can be summarized as follows:

The Morris screening method is particularly useful for computationally expensive models or multi-parameter models (tens or hundreds of parameters). It can screen out non-sensitive or unimportant parameters with a few runs of the model. The six sensitive parameters, g_1, g_2, Kaw, $RoughRss$, WM, and $Thick_u$ were selected from the 14 input parameters for quantitative GSA.

The response surface method is a statistical emulator or meta-model approach to simulate the original DTVGM and can greatly reduce the number of model evaluation runs in the sensitivity analysis. The MARS method, a nonparametric modeling approach, can characterize nonlinear relationships between the objective function responses and input variables in hydrological models. Compared to other response surface methods, the MARS method requires little data preparation, is suitable for handling fairly large datasets with a tendency toward a good bias-variance trade-off, and is flexible enough to model non-linearity and variable interactions.

The RSMSobol' method consists of a combination of the response surface method and the variance-based method. It is an efficient approach with less computational cost than the brute-force approach. The RSMSobol' method provides a quantitative measure of the sensitivity of the output variables to the different parameters as well as complementary information such as the impact of small changes in a specific parameter on a specific output variable and how several parameters interact with each other to produce changes to an output variable.

While the application of the GSA is useful for improving model calibration by reducing the number of parameters for optimization and lowering the computational cost in the calibration, the possibility of achieving different results should be considered if a non-uniform distribution is used. The assumption of uniform ranges is not accurate in most conditions. In addition, although the normal distribution was considered in this work, the means and standard deviations of the input parameters were selected subjectively by the author due to individual practice for the DTVGM, which may be not well-suited to real system processes. Further analysis is required to explore the possible impact of parameter non-uniform distribution on the results of the GSA.

Sensitivity analysis can single out and rank the parameters which are most sensitive, thereby providing a method with which to prioritize our research efforts to improve parameter estimation for the model. Furthermore, the sensitivity analysis results provide a unique opportunity to deepen and extend our understanding of the importance of functional interrelations within the water cycle system. And thus, the study will provide a strong basis for

implementing efficient parameter optimization for large and complex hydrological models.

Acknowledgments

information exchange to be performed via a user-written handler. There is no need to change the simulation source code to accommodate the analysis, and it is sometimes called a "non-intrusive" interface. The results can be analyzed using PSUADE's rich collection of response surface and analysis tools. PSUADE also creates "Matlab" files for data visualization. Taking the sensitivity analysis as an example, the flowchart for running PSUADE is as shown in Fig. A1 (Song et al., 2011).
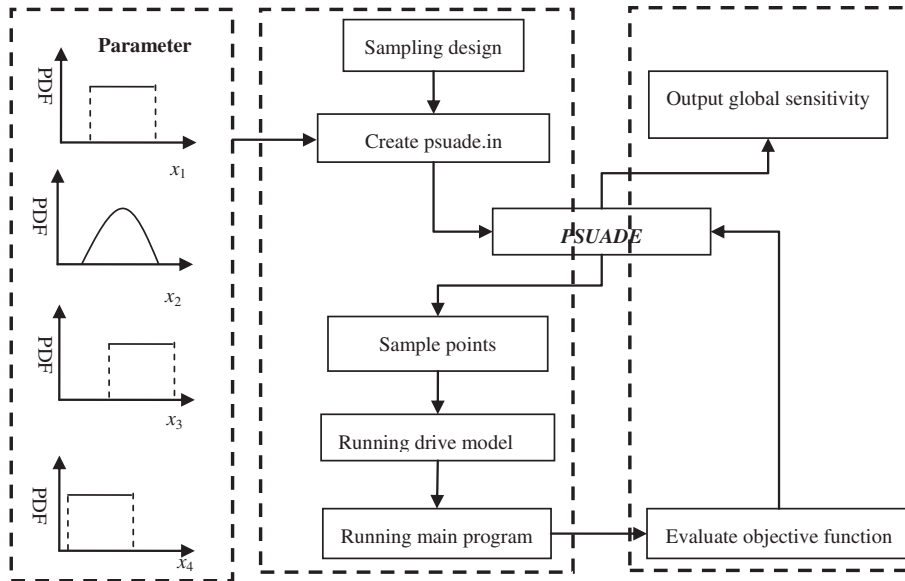

Fig. A1. Operation framework in PSUADE for sensitivity analysis. Note: PDF is short for probability density function.

Appendix A. PSUADE

PSUADE (https://computation.llnl.gov/casc/uncertainty_quantification/) contains a rich set of tools for performing UA, GSA, design optimization, and model calibration, etc. It also has a suite of sampling methods such as Monte Carlo, full and fractional factorial, Latin hypercube, orthogonal arrays, and one-at-a-time methods, etc. Furthermore, uniform, normal and lognormal distributions are available for inputs (Tong and Graziani, 2008). PSUADE also provides a rich set of statistical analysis tools such as basic statistics, correlation analysis, and variance decomposition using main effect analysis, among others (Tong, 2007a). In particular, PSUADE supports a global sensitivity methodology for models with large number of parameters and complex constraints and has been used for multi-physical processes models (e.g. Hsieh, 2007; Tong, 2007b, 2008; Tong and Graziani, 2008).

PSUADE supports a user-friendly interface via input and output filters. Once a sampling design has been created, it systematically feeds the design points into parameter files and calls a user handler, which absorbs the sample data and inserts the data into the model input files. The user handler then requests computational resources for the evaluation, waits for its completion, and extracts the output data from the model output files. The protocol allows all

Appendix B. Main effect analysis by McKay

McKay (1995) proposed an efficient estimation method based on the use of a single replicated Latin hypercube sampling design for all n inputs. The essence of this analysis is the statistical measure called variance of conditional expectation. We let $E(Y)$ and $V(Y)$ be the prediction mean and variance of an output variable Y, and

$$V(Y) = V(E(Y|X_i)) + E(V(Y|X_i)) \tag{B1}$$

where X_i is the ith input. Here the first term on the right hand side is the variance of the conditional expectation of Y, conditioned on X_i. It is also denoted as $VCE(X_i)$. The second term is an error or residual term. VCE measures the variability in the conditional expected value of Y as the input X_i takes on different values, while the residual term represents the variability in Y not accounted for by the input X_i.

Using a replicated Latin Hypercube design, each X_i takes on distinct values X_{ij}, $j = 1,\ldots, S$ where S is the number of levels. The expectation conditioned on $X_i = x_{ij}$ is

$$\overline{Y}_j = \frac{1}{R} \sum_{k=1}^{R} Y_{kj} \tag{B2}$$

where Y_{kj} is the output corresponding on the kth replication and with $X_i = x_{ij}$. And the variance is approximated by

$$V(Y) = \frac{1}{SR} \sum_{j=1}^{S} \sum_{k=1}^{R} \left(Y_{kj} - \overline{Y} \right)^2 \tag{B3}$$

The variance of conditional expectation on X_i is approximated by

$$V(Y|X_i = x_{ij}) = \frac{1}{R} \sum_{k=1}^{R} \left(Y_{kj} - \overline{Y}_j \right)^2 \tag{B4}$$

and hence

$$E(V(Y|X_i)) = \frac{1}{SR} \sum_{j=1}^{S} \sum_{k=1}^{R} \left(Y_{kj} - \overline{Y}_j \right)^2 \tag{B5}$$

According to the work of McKay (1995), the variance of a sampling distribution can be derived from the expectation of a sum of squares about the sample mean, i.e., it can be calculated by

$$\begin{aligned} E\left(\sum_{j=1}^{S} (\overline{Y}_j - \overline{Y})^2 \right) &\approx SV(\overline{Y}_j) = S\left\{ V\left[E(\overline{Y}_j|X_i = x_{ij}) \right] \right. \\ &\quad + E\left[V(\overline{Y}_j|X_i = x_{ij}) \right] \right\} \\ &= S\left\{ V\left[E(Y|X_i = x_{ij}) \right] \right. \\ &\quad \left. + \frac{1}{R} E\left[V(Y|X_i = x_{ij}) \right] \right\} \end{aligned} \tag{B6}$$

Then the $VCE(X_i)$ can be calculated by

$$VCE(X_i) = \frac{1}{S} \sum_{j=1}^{S} (\overline{Y}_j - \overline{Y})^2 - \frac{1}{SR^2} \sum_{j=1}^{S} \sum_{k=1}^{R} \left(Y_{kj} - \overline{Y}_j \right)^2 \tag{B7}$$

Finally, we have the main effect or first-order sensitivity index S_i by

$$S_i = V_i/V(Y) = VCE(X_i)/V(Y) \tag{B8}$$

References

Ascough II, J.C., Green, T.R., Ma, L., Ahjua, L.R., 2005. Key criteria and selection of sensitivity analysis methods applied to natural resource models. In: Modsim 2005 International Congress on Modeling and Simulation. Modeling and Simulation Society of Australia and New Zealand, Melbourne, Australia, pp. 2463–2469.

Bahremand, A., De Smedt, F., 2008. Distributed hydrological modeling and sensitivity analysis in Torysa watershed, Slovakia. Water Resources Management 22, 393–408.

Balshi, M.S., McGuire, A.D., Duffy, P., Flannigan, M., Walsh, J., Melillo, J., 2009. Assessing the response of area burned to changing climate in western boreal North America using a multivariate adaptive regression splines (MARS) approach. Global Change Biology 15 (3), 578–600.

Campolongo, F., Braddock, R., 1999. The use of graph theory in the sensitivity analysis of the model output: a new screening method. Reliability Engineering & System Safety 64 (1), 1–12.

Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. Environmental Modelling & Software 22 (10), 1509–1518.

Campolongo, F., Saltelli, A., Cariboni, J., 2011. From screening to quantitative sensitivity analysis. A unified approach. Computer Physics Communications 182, 928–988.

Campolongo, F., Tarantola, S., Saltelli, A., 1999. Tackling quantitatively large dimensionality problems. Computer Physics Communications 117, 75–85.

Cariboni, J., Gatelli, D., Liska, R., Saltelli, A., 2007. The role of sensitivity analysis in ecologic modeling. Ecological Modelling 203, 167–182.

Cibin, R., Sudheer, K.P., Chaubey, I., 2010. Sensitivity and identifiability of stream flow generation parameters of the SWAT model. Hydrological Processes 24, 1133–1148.

Cloke, H.L., Pappenberger, F., Renaud, J.P., 2008. Multi-method global sensitivity analysis (MMGSA) for modelling floodplain. Hydrological Processes 22 (11), 1660–1674.

Confalonieri, R., 2010. Monte Carlo based sensitivity analysis of two crop simulators and considerations on model balance. European Journal of Agronomy 33 (2), 89–93.

Confalonieri, R., Bellocchi, G., Bregaglio, S., Donatelli, M., Acutis, M., 2010. Comparison of sensitivity analysis techniques: a case study with the rice model WARM. Ecological Modelling 221 (16), 1897–1906.

Crosetto, M., Tarantola, S., 2001. Uncertainty and sensitivity analysis: tools for GIS-based model implementation. International Journal of Geographical Information Science 15 (5), 415–438.

Cukier, R.I., Levine, H.B., Shuler, K.E., 1978. Nonlinear sensitivity analysis of multi-parameter model systems. Journal of Computational Physics 26, 1–42.

Elsawwaf, M., Willems, P., Feyen, J., 2010. Assessment of the sensitivity and prediction uncertainty of evaporation models applied to Nasser Lake, Egypt. Journal of Hydrology 395, 10–22.

Francos, A., Elorza, F.J., Bouraoui, F., Bidoglio, G., Galbiati, L., 2003. Sensitivity analysis of distributed environmental simulation models: understanding the model behavior in hydrological studies at the catchment scale. Reliability Engineering & System Safety 79 (2), 205–218.

Freer, J., Beven, K.J., Ambroise, B., 1996. Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. Water Resources Research 32 (7), 2162–2173.

Friedman, J.H., 1991. Multivariate adaptive regression splines. The Annals of Statistics 19 (1), 1–67.

Hamby, D.M., 1994. A review of techniques for parameter sensitivity analysis of environmental models. Environmental Monitoring and Assessment 32, 135–154.

Helton, J.C., Johnson, J.D., Sallaberry, C.J., Storlie, C.B., 2006. Survey of sampling-based methods for uncertainty and sensitivity analysis. Reliability Engineering & System Safety 91 (10–11), 1175–1209.

Holvoet, K., van Griensven, A., Seuntjens, P., Vanrolleghem, P.A., 2005. Sensitivity analysis for hydrology and pesticide supply towards the river in SWAT. Physics and Chemistry of the Earth 30 (8–10), 518–526.

Homma, T., Saltelli, A., 1996. Importance measures in global sensitivity analysis of model output. Reliability Engineering & System Safety 52 (1), 1–17.

Hsieh, H., 2007. Application of the PSUADE Tool for Sensitivity Analysis of an Engineering Simulation. Lawrence Livermore National Laboratory Report, UCRL-TR-237205.

Jakeman, A.J., Letcher, R.A., Norton, J.P., 2006. Ten iterative steps in development and evaluation of environmental models. Environmental Modelling & Software 21 (5), 602–614.

Lee, D.H., 2007. Testing a conceptual hillslope recession model based on the storage–discharge relationship with the Richards equation. Hydrological Processes 21, 3155–3161.

Lenhart, T., Eckhardt, K., Fohrer, N., Frede, H.G., 2002. Comparison of two different approaches of sensitivity analysis. Physics and Chemistry of the Earth 27, 645–654.

Li, L., Xia, J., Xu, C.Y., Chu, J.T., Wang, R., 2009. Analyse the sources of equifinality in hydrological model using GLUE methodology. In: Hydroinformatics in Hydrology, Hydrogeology and Water Resources, Proceedings of Symposium JS.4 at the Joint IAHS & IAH Convention, Hyderabad, India, September 2009, vol. 331. IAHS Publ., pp. 130–138.

Li, L., Xia, J., Xu, C.Y., Singh, V.P., 2010. Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models. Journal of Hydrology 390 (3–4), 210–221.

Lilburne, L., Tarantola, S., 2009. Sensitivity analysis of spatial models. International Journal of Geographical Information Science 23 (2), 151–168.

Makler-Pick, V., Gal, G., Gorfine, M., Hipsey, M.R., Carmel, Y., 2011. Sensitivity analysis for complex ecological models – a new approach. Environmental Modelling & Software 26, 124–134.

Manache, G., Melching, C.S., 2008. Identification of reliable regression- and correlation-based sensitivity measures for importance ranking of water-quality model parameters. Environmental Modelling & Software 23, 549–562.

Marrel, A., Looss, B., Laurent, B., Roustant, O., 2009. Calculations of Sobol indices for the Gaussian process metamodel. Reliability Engineering & System Safety 94, 742–751.

McKay, M.D., 1995. Evaluating Prediction Uncertainty. Los Alamos National Laboratory Technical Report NUREG/CR-6311, LA-12915-MS.

Meyers, R.H., Montgomery, D.C., 2002. Response Surface Methodology: Process and Product Optimization Using Designed Experiments. Wiley-Interscience.

Moore, L.M., Ray, B.K., 1999. Statistical methods for sensitivity and performance analysis in computer experiments. In: Proceedings of the 1999 Winter Simulation Conference, Arizona, United States, pp. 486–491.

Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics 33 (2), 161–174.

Muleta, M.K., Nicklow, J.W., 2005. Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. Journal of Hydrology 306 (1–4), 127–145.

Oakley, J.E., O'Hagan, A., 2004. Probabilistic sensitivity analysis of complex models: a Bayesian approach. Journal of the Royal Statistical Society Series B 66, 751–769.

Pappenberger, F., Iorgulescu, I., Beven, K.J., 2006. Sensitivity analysis based on regional splits and regression trees (SARS-RT). Environmental Modelling & Software 21 (7), 976–990.

Pappenberger, F., Beven, K.J., Ratto, M., Matgen, P., 2008. Multi-method global sensitivity analysis of flood inundation models. Advances in Water Resources 31 (1), 1–14.

Ratto, M., Tarantola, S., Saltelli, A., 2001. Sensitivity analysis in model calibration: GSA-GLUE approach. Computer Physics Communications 136, 212–224.

Ren, Q.W., Chen, Y.P., Shu, X.J., 2010. Global sensitivity analysis of Xinanjiang model parameters based on extend FAST method. Acta Scientiarum Naturalium Unversitatis Sunyatseni 49 (3), 127–134 (in Chinese).

Saltelli, A., 1999. Sensitivity analysis: could better methods be used? Journal of Geophysical Research – Atmospheres 104 (D3), 3789–3793.

Saltelli, A., Annoni, P., 2010. How to avoid a perfunctory sensitivity analysis. Environmental Modelling & Software 25, 1508–1517.

Saltelli, A., Bolado, R., 1998. An alternative way to compute Fourier amplitude sensitivity test (FAST). Computational Statistics & Data Analysis 26 (4), 445–460.

Saltelli, A., Tarantola, S., 2002. On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal. Journal of American Statistical Association 97, 702–709.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Computer Physics Communication 181 (2), 259–270.

Saltelli, A., Campolongo, F., Cariboni, J., 2009. Screening important inputs in models with strong interaction properties. Reliability Engineering & System Safety 94 (7), 1149–1155.

Saltelli, A., Chan, K., Scott, M., 2000. Sensitivity Analysis. In: Probability and Statistics Series. John Wiley & Sons, Ltd., West Sussex, England.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global Sensitivity Analysis: the Primer. John Wiley & Sons, Ltd., West Sussex, England.

Saltelli, A., Ratto, M., Tarantola, S., Campolongo, F., 2005. Sensitivity analysis for chemical models. Chemical Reviews 105 (7), 2811–2828.

Saltelli, A., Ratto, M., Tarantola, S., Campolongo, F., 2006. Sensitivity analysis practices: strategies for model-based inference. Reliability Engineering & System Safety 91 (10–11), 1109–1125.

Saltelli, A., Tarantola, S., Chan, K., 1999. Quantitative model-independent method for global sensitivity analysis of model output. Technometrics 41 (1), 39–56.

Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., 2004. Sensitivity Analysis in Practice: a Guide to Assessing Scientific Models. John Wiley & Sons, Ltd., England.

Sieber, A., Uhlenbrook, S., 2005. Sensitivity analysis of a distributed catchment model to verify the model structure. Journal of Hydrology 310, 216–235.

Sobol', I.M., 1967. On the distribution of points in a cube and the approximate evaluation of integrals. USSR Computational Mathematics and Mathematical Physics 7, 86–112.

Sobol', I.M., 1993. Sensitivity estimates for nonlinear mathematical models. Mathematical Modeling and Computational Experiment 1, 407–414.

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

Sobol', I.M., Shukhman, B.V., 1995. Integration with quasi-random sequences: numerical experience. International Journal of Modern Physics C 6 (2), 263–275.

Song, X.M., Kong, F.Z., Zhan, C.S., Han, J.W., 2012a. A hybrid optimization rainfall–runoff simulation based on Xinanjiang model and artificial neural network. Journal of Hydrologic Engineering 17 (9), 1033–1041. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000548.

Song, X.M., Zhan, C.S., Xia, J., 2012b. Integration of statistical emulator approach with SCE-UA method in parameter optimization of hydrological model. Chinese Science Bulletin 57 (26), 3393–3407. http://dx.doi.org/10.1007/s11434-012-5305-x.

Song, X.M., Zhan, C.S., Xia, J., Kong, F.Z., 2012c. An efficient global sensitivity analysis approach for distributed hydrological model. Journal of Geographical Sciences 22 (2), 209–222.

Song, X.M., Zhan, C.S., Kong, F.Z., Xia, J., 2011. Advances in the study of uncertainty quantification of large-scale hydrological modeling system. Journal of Geographical Sciences 21 (5), 801–819.

Spruill, C.A., Workman, S.R., Taraba, J.L., 2000. Simulation of daily and monthly stream discharge from small watersheds using the SWAT model. Transactions of the ASAE 43 (6), 1431–1439.

Statnikow, R.B., Matusov, J.B., 2002. Multicriteria Analysis in Engineering. Kluwer Academic Publishers, Netherlands.

Tang, Y., Reed, P.M., Wagener, T., van Werkhoven, K., 2007a. Advancing the identification and evaluation of distributed rainfall–runoff models using global sensitivity analysis. Water Resources Research 43 (6), W06415.

Tang, Y., Reed, P.M., van Werkhoven, K., Wagener, T., 2007b. Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. Hydrology and Earth System Sciences 11, 793–817.

Tarantola, S., Saltelli, A., 2003. SAMO 2001: methodological advances and innovative applications of sensitivity analysis. Reliability Engineering & System Safety 79 (2), 121–122.

Tong, C., 2006. Refinement strategies for stratified sampling methods. Reliability Engineering & System Safety 91 (10–11), 1257–1265.

Tong, C., 2007a. The PSUADE Software Package Version 1.0. LLNL code release UCRL-CODE-235523.

Tong, C., 2007b. Toward a More Robust Variance-based Global Sensitivity Analysis of Model Outputs. Lawrence Livermore National Laboratory Report, UCRL-TR-235561.

Tong, C., 2008. Quantifying Uncertainty of a Soil-foundation Structure-interaction System Under Seismic Excitation. Lawrence Livermore National Laboratory Report, UCRL-TR-402883.

Tong, C., Graziani, F., 2008. A practical global sensitivity analysis methodology for multi-physics applications. Computational Methods in Transport: Verification and Validation. Lecture Notes in Computational Science and Engineering, vol. 62, pp. 277–299.

Tong, C., 2010. Self-validated variance-based methods for sensitivity analysis of model outputs. Reliability Engineering & System Safety 95 (3), 301–309.

van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., Srinivasan, R., 2006. A global sensitivity tool for the parameters of multi-variable catchment models. Journal of Hydrology 324 (1–4), 10–23.

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

Wang, G.S., 2005. Theory and method of distributed time-variant gain model. Ph.D. thesis, China Academic Science, Beijing (in Chinese).

Wang, G.S., Xia, J., Chen, J., 2009. Quantification of effects of climate variations and human activities on runoff by a monthly water balance model: a case study of the Chaobai River basin in northern China. Water Resources Research 45, W00A11. http://dx.doi.org/10.1029/2007WR006768.

Wang, G.S., Xia, J., Chen, J.F., 2010. A multi-parameter sensitivity and uncertainty analysis method to evaluate relative importance of parameters and model performance. Geographical Research 29 (2), 263–270 (in Chinese).

Xia, J., 2002. A system approach to real time hydrological forecasts in watersheds. Water International 27 (1), 87–97.

Xia, J., O'Connor, K.M., Kachroo, R.H., Liang, G.C., 1997. A non-linear perturbation model considering catchment wetness and its application in river flow forecasting. Journal of Hydrology 200 (1–4), 164–178.

Xia, J., Wang, G.S., Tan, G., Ye, A.Z., Huang, G.H., 2005. Development of distributed time-variant gain model for nonlinear hydrological systems. Science in China Series D, Earth Sciences 48 (6), 713–723.

Xu, C.G., Gertner, G., 2011. Understanding and comparisons of different sampling approaches for the Fourier amplitudes sensitivity test (FAST). Computational Statistics & Data Analysis 55 (1), 184–198.

Yang, J., 2011. Convergence and uncertainty analysis in Monte-Carlo based sensitivity analysis. Environmental Modelling & Software 26 (4), 444–457.

Ye, A.Z., 2007. Study on catchment water cycle simulation in changing environments. Ph.D. thesis, Wuhan University, Wuhan (in Chinese).

Ye, A.Z., Duan, Q.Y., Zeng, H.J., Li, L., Wang, C.Y., 2010. A distributed time-variant gain hydrological model based on remote sensing. Journal of Resources and Ecology 1 (3), 222–230.