Sensitivity Analysis Methods

Yanjun Gan and Qingyun Duan

Contents

1 Introduction 2
2 Methodologies and Applications 3
2.1 Gradient-Based Methods 4
2.2 Variance-Based Methods 11
2.3 Regression-Based Methods 19
3 Which SA Methods to Use? 28
4 Summary 31
References 31

Abstract

Sensitivity analysis (SA) is an important tool for assessing and reducing uncertainties in computer-based models. This chapter presents a comprehensive review of some commonly used SA methods, including gradient-based, variance-based, and regression-based methods. Features and applicability of those methods are described and illustrated with some examples. Merits and limitations of different methods are explained, and the criteria of choosing appropriate SA methods for different applications are suggested.

Y. Gan (✉)
State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing, China
e-mail: yjgan@cma.gov.cn

Q. Duan
State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing, China
e-mail: qyduan@bnu.edu.cn

Q. Duan et al. (eds.), *Handbook of Hydrometeorological Ensemble Forecasting*,
https://doi.org/10.1007/978-3-642-40457-3_65-1

Keywords
Uncertainty quantification · Sensitivity analysis · Uncertainty analysis · Parameter estimation · Design of experiment · Sampling · Parameter screening · Variance decomposition

1 Introduction

Computer-based models are used to predict real-world processes and are important tools for facilitating understanding of complex, real-world phenomena or solving challenging engineering design problems. They have become indispensable in many fields of science and engineering, from finance to life sciences, from quantum physics to earth sciences and environmental engineering (Gan et al. 2014). Different sources of uncertainties, such as forcing data, observational data, and model structure and parameters, exert great influences on model performance (Renard et al. 2010; Walker et al. 2003). Consequently, output uncertainties should be assessed and attributed to different sources to increase our understanding and confidence of the model based predictions.

Uncertainty analysis (UA) and sensitivity analysis (SA) are two of the fundamental steps in assessing and reducing model uncertainties, a discipline known as "uncertainty quantification (UQ)," with the former a forward propagation, whereas the latter an inverse assessment of model uncertainties (Jakeman et al. 2006). UA focuses on quantifying uncertainty in model outputs, while SA refers to the study of how the uncertainty in the model outputs can be apportioned to different input uncertainty sources (Cariboni et al. 2007; Saltelli et al. 2008). In general, output uncertainty can be numerically represented by statistical measures such as means, standard deviations, skewness, kurtosis, and confidence intervals or pictorially described by probability density functions, cumulative distribution functions, and box plots. Although UA should be run in tandem with SA, it is beyond our scope to give a full review of the UA methods, and hence we refer the readers to Uusitalo et al. (2015) for an overview of UA methods. Yet most commonly, SA is focused on evaluating the influences of model parameters (also called experimental factors) on model outputs (also called responses) (Rakovec et al. 2014). This is because parameters govern many aspects of a model and are of great uncertainty, even if the model structure is correct and the data errors are assumed negligible. However, focusing on SA of model parameters does not mean, for example, the model structure or data are not important, but should help understand the roles played by them.

SA has been employed by many researchers to evaluate the influence of each parameter on model performance and then screen out insensitive parameters from analyses (Borgonovo et al. 2012; Campolongo et al. 2007). It yields key insights into model parameter behaviors and would help reduce parameter dimensionality for subsequent analyses such as parameter estimation (PE), a process for calibrating model simulations to historical observations by tuning influential parameters (Duan

et al. 2006). On the other hand, a full understanding of parameter behaviors would facilitate model verification and validation throughout the course of model development and refinement (Frey and Patil 2002; Sieber and Uhlenbrook 2005). In addition, SA can identify critical regions of parameter space that can aid in model calibration.

Numerous SA approaches have been developed over the years, and they can be classified in a variety of ways. Usually, they are categorized into two groups as local and global methods according to their action ranges (Saltelli et al. 2008). Local SA methods explore the changes of model response by varying one parameter at a time while keeping other parameters constant, using partial derivatives or finite differences at a fixed parameter location as the measure of parametric sensitivity. Though simple and intuitive, local SA methods measure only local sensitivity whose value is obviously location dependent. Consequently, they are applicable only for linear and monotonic problems. On the other hand, global SA methods examine the changes of model response by varying all or a subset of the parameters simultaneously over the entire parameter space, allowing them to provide robust measures in the presence of nonlinearity and interactions among the parameters (Wainwright et al. 2014). Other classifications include methodological categories as mathematical, statistical, and graphical methods by Frey and Patil (2002) and capability categories as qualitative and quantitative methods by Saltelli et al. (1999). Qualitative SA methods aim to screen out a subset of non-influential parameters using a small number of model evaluations, whereas quantitative SA methods aim to measure each parameter's contribution to the response variance, a process that requires a large number of model evaluations (Campolongo et al. 2011; Cariboni et al. 2007). There are many different software packages which include a variety of different SA methods. An excellent review of available software packages that could be adopted for SA, as well as UA and PE, is available in Matott et al. (2009) and Wang et al. (2016).

The selection of the appropriate SA methods for specific problem is not a trivial issue in practice but a potentially tricky task to those who have a minimal amount of experience in mathematical and statistical theories. We later review a series of commonly used SA methods by applying them to a few illustrative examples and discuss their strengths and limitations. The objectives are to present a systematic introduction and illustrative application of those methods, as well as to provide guidance on choosing the appropriate techniques for specific applications.

The remainder of this chapter is arranged as follows: Methodologies of some commonly used SA methods are presented in Sect. 2. An overview of literature on comparison of different SA methods is given in Sect. 3. Finally, we discuss the criteria for selecting appropriate SA methods for specific applications in Sect. 4.

2 Methodologies and Applications

A comprehensive review of different SA methods is given in this section. The SA methods are divided into three categories according to their mathematical approaches used to compute the sensitivity indices: (1) gradient-based,

(2) variance-based, and (3) regression-based. The features of different SA categories are summarized in Table 1. The methodologies of those SA methods are presented below with some illustrative examples.

2.1 Gradient-Based Methods

Gradient-based methods compute the sensitivity indices based on the change in response gradient to the variation of an input factor. If the response gradient induced by a varying factor is larger than that of other factors, it indicates that the varying factor is more sensitive than the other factors. There are numerous gradient-based SA methods. Some of them are reviewed below.

2.1.1 One-At-a-Time

The one-at-a-time (OAT) method (Daniel 1958) is perhaps the most fundamental and intuitive SA method, which assesses parameter sensitivity by sequentially perturbing one parameter at a time while keeping the other parameters at their baseline values. A schematic diagram of three-parameter OAT design is given in Fig. 1.

Assume that we have an n-dimensional parameter space, sensitivity index of the OAT method for parameter X_i ($i = 1, 2, \cdots, n$) is

$$S_i = \frac{Y(X_1, \cdots, X_{i-1}, X_i \pm \Delta X_i, X_{i+1}, \cdots, X_n) - Y(\mathbf{X})}{\Delta X_i} \tag{1}$$

where ΔX_i is the increment for the ith parameter. The OAT method requires only $n + 1$ experiments for an n-dimensional problem, i.e., an experiment for the base point plus n experiments for small perturbation in each of the n parameters.

The OAT method is also known as local SA method since it explores only a local space around the base point. However, the OAT method has been extensively used because it is easy to implement, computationally inexpensive, and useful to provide a glimpse at the model behavior (Saltelli 1999). By reviewing 33 SA-related papers published in *Science* between 1997 and 2003, Saltelli et al. (2006) found that the OAT method has been improperly applied by many researchers, even though this method is actually only justified for linear models.

2.1.2 Fractional Factorial Screening

The fractional factorial (FF) screening makes use of FF sampling (Box and Hunter 1961a, b) to design a small number of experiments for estimating parameter sensitivity. Assume that a model has n parameters with each of them having p levels. A full factorial design would require p^n experiments, while a FF design needs only $1/p^k$ fraction of the experiments of the full factorial design (i.e., p^{n-k} experiments), where k is the number of generators. For example, a 2^{5-2} FF design is 1/4 of a two-level five-parameter full factorial design (Table 2). If we denote the two levels of each parameter as "−" (the low level) and "+" (the high level), the 2^{5-2} FF design can be generated from a three-parameter (say A, B, and C) full factorial design and

Table 1 Features of different sensitivity analysis methods

Category	Method	Property	Model independent	Elementary effect	Main effect	Second-order interaction effect	Higher-order interaction effect	Total effect
Gradient-based	One-at-a-time (OAT)	Local	No (linearity and monotonicity)	√	×	×	×	×
	Fractional factorial screening (FF)	Qualitative global	No (linearity or monotonicity)	×	√	×	×	×
	Plackett-Burman screening (PB)	Qualitative global	No (linearity and additivity)	×	√	×	×	×
	Morris one-at-a-time (MOAT)	Qualitative global	Yes	√	×	(a)	(a)	√
Variance-based	Analysis of variance (ANOVA)	Qualitative global	Yes	×	√	√	√	×
	Fourier amplitude sensitivity test (FAST)	Quantitative global	Yes	×	√	×	×	×
	Extended FAST (EFAST)	Quantitative global	Yes	×	√	×	×	√
	McKay correlation ratios	Quantitative global	Yes	×	√	√	×	×
	Sobol' sensitivity indices	Quantitative global	Yes	×	√	√	√	√

(*continued*)

Table 1 (continued)

Category	Method	Property	Model independent	Elementary effect	Main effect	Second-order interaction effect	Higher-order interaction effect	Total effect
Regression-based	Linear regression (LR)	Qualitative global	No (linearity or monotonicity)	×	√	×	×	×
	Multivariate adaptive regression splines (MARS)	Qualitative/ quantitative global	Yes	(b)	(b)	(b)	(b)	(b)
	Delta test (DT)	Qualitative/ quantitative global	Yes	(b)	(b)	(b)	(b)	(b)
	Sum-of-tree (SOT)	Qualitative/ quantitative global	Yes	(b)	(b)	(b)	(b)	(b)
	Gaussian process (GP)	Qualitative/ quantitative global	Yes	(b)	(b)	(b)	(b)	(b)

(a) Specific order interactions cannot be detected, but all order interactions are measured together; (b) qualitative evaluation, parameter total effects; quantitative evaluation, dependents on the quantitative SA method that the surrogate model combined with

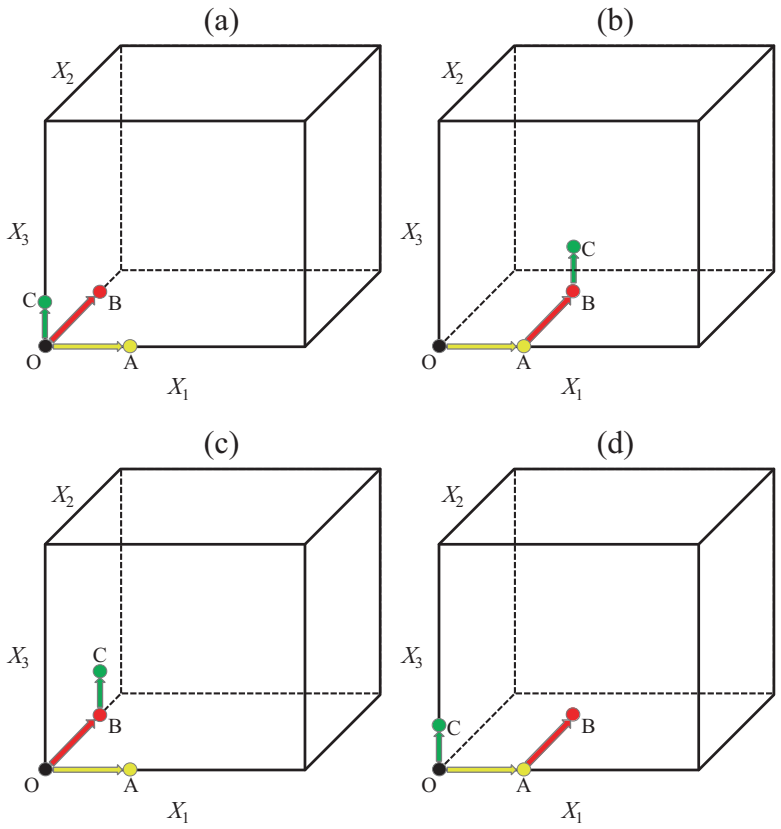

Fig. 1 Illustration of three-parameter one-at-a-time (OAT) design

Table 2 A 2^{5-2} fractional factorial design

Treatment combination	Factor effect					
	I	A	B	C	$D = AB$	$E = AC$
de	+	−	−	−	+	+
a	+	+	−	−	−	−
be	+	−	+	−	−	+
abd	+	+	+	−	+	−
cd	+	−	−	+	+	−
ace	+	+	−	+	−	+
bc	+	−	+	+	−	−
abcde	+	+	+	+	+	+

then choosing to confound the two remaining parameters D and E with interactions generated by $D = A \times B$ and $E = A \times C$. These two expressions are the generators of the 2^{5-2} FF design, with $D = A \times B$ means that the main effect of D is confounded

with the interactions involving A and B, while $E = A \times C$ means that the main effect of E is confounded with the interactions involving A and C.

The FF design shares the balance property of corresponding full factorial design, meaning that every level of a parameter appears the same number of times. However, some parameter effects of the FF design would inevitably be aliased with others since it uses only a fraction of the experiments of the full factorial design. A critical consideration when selecting a proper FF design is that the effects of primary interest are aliased only with higher-order interactions that are negligible.

For parameter screening purpose, two-level FF experiments are often designed to investigate parameter main effect. Suppose that parameter X_i $(i = 1, 2, \cdots, n)$ has two levels as "high" (denoted as $X_{\mathrm{i}}^{(+)}$) and "low" (denoted as $X_{\mathrm{i}}^{(-)}$), the main effect of X_i can be obtained by

$$S_i = \frac{\bar{Y}_i^{(+)} - \bar{Y}_i^{(-)}}{X_i^{(+)} - X_i^{(-)}} \tag{2}$$

where $\bar{Y}_{\mathrm{i}}^{(+)}$ and $\bar{Y}_{\mathrm{i}}^{(-)}$ are mean response values when X_i equals to high and low levels, respectively. FF screening is effective only if the parameter-response relationship is linear or monotonic.

Henderson-Sellers (1993) designed three sets of 32-run two-level FF experiments for assessing the relative importance of 23 ecotype parameters of the Biosphere-Atmosphere Transfer Scheme (BATS) (Dickinson et al. 1986) under three different climatic regimes. A detailed description of the FF screening method for parameter sensitivity analyses of environmental models was given by Henderson-Sellers and Henderson-Sellers (1996).

2.1.3 Plackett-Burman Screening

Take a two-level ten-parameter problem, for example, a 1/4 FF design would still require 256 ($=2^8$) experiments. The cost of FF design is sometimes prohibitive for time-consuming high-dimensional problems. Plackett-Burman (PB) design (Plackett and Burman 1946) provides an alternative when the FF design is impractical to implement. With a N-run PB design, one can run a screening experiment for up to $N - 1$ parameters, where N is a multiple of four. The design matrices for two-level n-parameter problem with a sample size up to 100 except 92 are given in Plackett and Burman (1946). Briefly, PB design can be generated by taking a one-dimensional matrix with "+" and "−" signs as the first column (or row), shifting it cyclically one place $N - 2$ times, and adding a final row of "−" signs to complete the design. A 12-run two-level PB design is shown in Table 3, and it can be used for screening experiment containing up to 11 parameters.

The sensitivity measure for PB screening is the same with FF screening. Main effects of PB screening are clear of each other but aliased with two-way interactions. Therefore, it is applicable when the two-way interactions are negligible. Besides, the parameter-response relationship should be linear and additive. Beres and Hawkins (2001) summarized the virtues of PB screening and gave a guide for performing

Table 3 A 12-run two-level Plackett-Burman design

Run	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
1	+	+	–	+	+	+	–	–	–	+	–
2	–	+	+	–	+	+	+	–	–	–	+
3	+	–	+	+	–	+	+	+	–	–	–
4	–	+	–	+	+	–	+	+	+	–	–
5	–	–	+	–	+	+	–	+	+	+	–
6	–	–	–	+	–	+	+	–	+	+	+
7	+	–	–	–	+	–	+	+	–	+	+
8	+	+	–	–	–	+	–	+	+	–	+
9	+	+	+	–	–	–	+	–	+	+	–
10	–	+	+	+	–	–	–	+	–	+	+
11	+	–	+	+	+	–	–	–	+	–	+
12	–	–	–	–	–	–	–	–	–	–	–

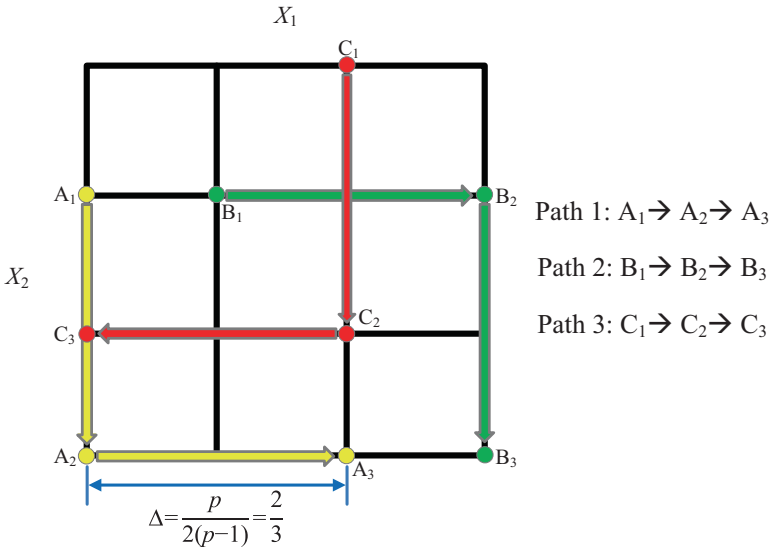

Fig. 2 Illustration of four-level two-parameter Morris one-at-a-time (MOAT) design, where A_1, B_1, and C_1 are random points and all other points are generated following OAT paths

it. Applications of PB screening for SA can also be found in Cryer and Havens (1999), Dion et al. (2011), and Grant et al. (2007).

2.1.4 Morris One-At-a-Time

The Morris one-at-a-time (MOAT) method (Morris 1991) was designed to overcome the deficiency of the OAT method, which is location dependent, by including multidimensional averaging of the local measures. The experimental plans consist of individually randomized OAT designs (Fig. 2). Theoretic basis of this method is

based on the elementary effect, which is representative of the change in a model response due to the change in a particular parameter.

Assume that we have an n-dimension p-level orthogonal parameter space, where each X_i may take on values from $\{0,\ 1/(p-1),\ 2/(p-1),\ \cdots,\ 1\}$. The elementary effect of the ith parameter is defined as

$$d_i = \frac{Y(X_1,\ \cdots,\ X_{i-1},\ X_i \pm \Delta,\ X_{i+1},\ \cdots,\ X_n) - Y(\mathbf{X})}{\Delta} \tag{3}$$

where the increment Δ usually is set to $p/[2(p-1)]$ and p is an even number. Overall and interaction effects of each parameter can then be approximated, respectively, by the mean and standard deviation of the elementary effects from r OAT paths as

$$\mu_i = \sum_{j=1}^{r} d_i(j)/r \tag{4}$$

and

$$\sigma_i = \sqrt{\sum_{j=1}^{r} [d_i(j) - \mu_i]^2/r} \tag{5}$$

The total number of experiments needed for a MOAT screening is $(n + 1)r$.

The MOAT screening method has been widely applied due to its efficiency for high-dimensional problems (Francos et al. 2003; Kleijnen 1997). On the other hand, improvements have also been made to this method. For example, van Griensven et al. (2006) replaced the Monte Carlo (MC) sampling with the Latin hypercube (LH) sampling for generating more uniform samples to improve the efficiency of the MOAT screening, which is known as LH-OAT screening. Campolongo et al. (2007) proposed a modified mean μ^*, which is an estimate of the mean of absolute elementary effects, to solve the problem of the compensating effect of opposite signs in elementary effects as

$$\mu_i^* = \sum_{j=1}^{r} |d_i(j)|/r \tag{6}$$

Example 2.1

Situation. Consider the g-function proposed by Sobol' (1993) as $f = \prod_{i=1}^{n} g_i(X_i)$, where $g_i(X_i) = (|4X_i - 2| + a_i)/(1 + a_i)$ depends on a nonnegative parameter a_i. Let $n = 15$, $a_1 = 9$, $a_2 = a_3 = 15$, $a_4 = a_5 = a_6 = 50$, $a_7 = a_8 = a_9 = a_{10} = 0$, $a_{11} = a_{12} = a_{13} = a_{14} = a_{15} = 70$, and $X_i \in [0,\ 1]$ with uniform distribution. Please find the influential parameters of this function.

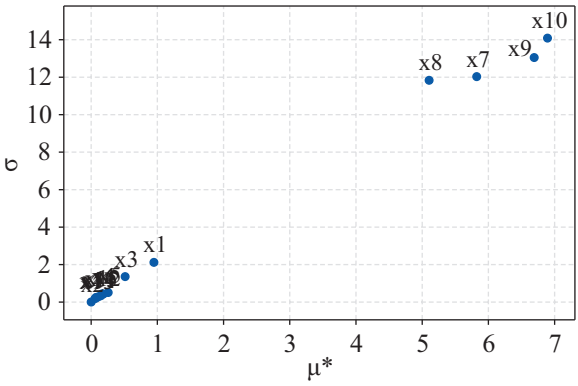

Fig. 3 MOAT screening for the 15-parameter *g*-function

Solution. The MOAT screening method is adopted to analyze parameter sensitivity of this 15-parameter problem. The range of each parameter is evenly divided into four levels. We then design 320 (=(15 + 1) × 20) experiments to screen out the insensitive parameters. The SA result is given in Fig. 3. It is easy to distinguish the insensitive parameters (i.e., X_1, X_2, X_3, X_4, X_5, X_6, X_{11}, X_{12}, X_{13}, X_{14}, and X_{15}) from the sensitive ones (X_7, X_8, X_9, and X_{10}) from this figure.

2.2 Variance-Based Methods

Variance-based methods make quantitative decomposition of the variance of model response into the contributions from individual parameters and their interactions. They are model independent and accurate but computationally expensive. Sampling techniques such as LH (McKay et al. 1979), quasi-MC (QMC) (Sobol' 1990), orthogonal array (OA) (Owen 1992), and orthogonal array-based Latin hypercube (OALH) (Tang 1993) have been widely used to generate uniformly distributed samples for variance-based SA.

2.2.1 Analysis of Variance

Analysis of variance (ANOVA) requires no assumptions for the relationship between model parameters and responses, but the responses should be normally distributed with same variance. Suppose we consider a problem with two independent factors (i.e., parameters) A and B. Among them, factor A has a levels (or treatments) as A_1, A_2, …, A_a, and factor B has b levels as B_1, B_2, …, B_b. Each level combination of factors A and B is repeated n times ($n \geq 2$). Each model response can be recorded as Y_{ijk}, where $i = 1, 2, \ldots, a$; $j = 1, 2, \ldots, b$; $k = 1, 2, \ldots, n$. Y_{ijk} is independent of each other, and $Y_{ijk} \sim N(\mu_{ij}, \sigma^2)$. The ANOVA model of this two-factor problem can be presented in terms of a linear statistical model as

$$\begin{cases} Y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \\ \mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} \end{cases} \tag{7}$$

where μ is the overall mean, α_i is the main effect of the ith level A_i, β_j is the main effect of the jth level B_j, γ_{ij} is the interaction effect of level combination $(A_i,\ B_j)$, μ_{ij} is the mean of level combination $(A_i,\ B_j)$, and $\varepsilon_{ijk} \sim N(0,\ \sigma^2)$ is the random error (or residual).

If we denote

$$\bar{Y} = \frac{Y_{\cdots}}{abn} = \frac{1}{abn}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} Y_{ijk} \tag{8}$$

$$\bar{Y}_{ij\cdot} = \frac{1}{n} Y_{ij\cdot} = \frac{1}{n}\sum_{k=1}^{n} Y_{ijk} \tag{9}$$

$$\bar{Y}_{i\cdot\cdot} = \frac{1}{bn} Y_{i\cdot\cdot} = \frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n} Y_{ijk} \tag{10}$$

$$\bar{Y}_{\cdot j\cdot} = \frac{1}{an} Y_{\cdot j\cdot} = \frac{1}{an}\sum_{i=1}^{a}\sum_{k=1}^{n} Y_{ijk} \tag{11}$$

where abn is the total number of experiments and "dot" subscript notation represents the summation over the subscript that it replaced. Thus, the total sum of squares S_T can be expressed and then decomposed as

$$\begin{aligned} S_T &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk} - \bar{Y}\right)^2 \\ &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left[\left(\bar{Y}_{i\cdot\cdot} - \bar{Y}\right) + \left(\bar{Y}_{\cdot j\cdot} - \bar{Y}\right) + \left(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}\right) + \left(Y_{ijk} - \bar{Y}_{ij\cdot}\right)\right]^2 \\ &= bn\sum_{i=1}^{a}\left(\bar{Y}_{i\cdot\cdot} - \bar{Y}\right)^2 + an\sum_{j=1}^{b}\left(\bar{Y}_{\cdot j\cdot} - \bar{Y}\right)^2 + n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}\right)^2 \\ &+ \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk} - \bar{Y}_{ij\cdot}\right)^2 \end{aligned} \tag{12}$$

The ANOVA table for this two-way fixed effects model is given in Table 4. F-test can be used to determine whether there exists a significant difference among treatment means of one factor or interactions between two factors, at a significance level of α. The higher the F value, the more significant the main effect or interaction effect is to the factor.

Table 4 ANOVA table for the two-way fixed effects model

Source of variation	Sum of squares	Degree of freedom	Mean square	F statistic[a]
Factor A	$S_A = bn\sum_{i=1}^{a}(\bar{Y}_{i\cdot} - \bar{Y})^2$	$a-1$	$MS_A = \frac{S_A}{a-1}$	$F_A = \frac{MS_A}{MS_E}$
Factor B	$S_B = an\sum_{j=1}^{b}(\bar{Y}_{\cdot j} - \bar{Y})^2$	$b-1$	$MS_B = \frac{S_B}{b-1}$	$F_B = \frac{MS_B}{MS_E}$
Interaction	$S_{A\times B} = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y})^2$	$(a-1)$ $(b-1)$	$MS_{A\times B} = \frac{S_{A\times B}}{(a-1)(b-1)}$	$F_{A\times B} = \frac{MS_{A\times B}}{MS_E}$
Error	$S_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{ijk} - \bar{Y}_{ij\cdot})^2$	$ab(n-1)$	$MS_E = \frac{S_E}{ab(n-1)}$	
Total	$S_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{ijk} - \bar{Y})^2$	$abn-1$		

[a] $F_A \sim F[a-1,\ ab(n-1)]$, $F_B \sim F[b-1,\ ab(n-1)]$, $F_{A\times B} \sim F[(a-1)(b-1),\ ab(n-1)]$

Frey and Patil (2002) gave a detailed description of the ANOVA method, including its advantages and disadvantages. Mokhtari and Frey (2005) showed that ANOVA is more reliable than correlation and regression methods by applying them for SA of a two-dimensional probabilistic risk assessment model. The reliability of ANOVA method was also illustrated by Tang et al. (2007).

2.2.2 Fourier Amplitude Sensitivity Test

Fourier amplitude sensitivity test (FAST) was presented by Cukier et al. (1973) for SA of multiparameter nonlinear model, in which conditional variances are represented by coefficients from the multiple Fourier series expansion of the response function and the ergodic theorem (Weyl 1938) is applied to transform the multidimensional integral into a one-dimensional integral in the evaluation of the Fourier coefficients. The FAST method is capable of computing the main effect of each parameter to the response variance.

Let $Y = f(\mathbf{X}) = f(X_1, X_2, \ldots, X_n)$, where $X_i \in [0,\ 1]$ and $i = 1, 2, \cdots, n$. Consider a set of transfer functions

$$X_i(s) = G_i[\sin(\omega_i s)] \tag{13}$$

where $\{\omega_i\}$ is a set of frequencies and $s \in (-\infty,\ \infty)$. The key idea of FAST is to apply the ergodic theorem to transform the n-dimensional integral $\int_0^1 \int_0^1 \ldots \int_0^1 f(\mathbf{X})\, \mathrm{d}X_1 \mathrm{d}X_2 \ldots \mathrm{d}X_n$ into a one-dimensional integral $\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} f(s)\ \mathrm{d}s$. Since the numerical computation of this integral is impossible for an incommensurate set of frequencies, an approximate numerical integration can be made by using a set of positive integer frequencies, which makes the search curve s not space-filling but periodic with a 2π period. By considering $f(s)$ within the finite interval $(-\pi,\ \pi)$, the expectation and variance of Y can then be approximated, respectively, by

$$E(Y) \approx \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s)\mathrm{d}s \tag{14}$$

and

$$V(Y) \approx \frac{1}{2\pi} \int_{-\pi}^{\pi} f^2(s)\mathrm{d}s - E^2(Y) \tag{15}$$

Following Parseval's theorem, we have

$$\frac{1}{\pi} \int_{-\pi}^{\pi} f^2(s)\mathrm{d}s = \frac{1}{2}A_0^2 + \sum_{p=1}^{\infty} \left(A_p^2 + B_p^2\right) \tag{16}$$

where $A_p = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \cos(ps)\mathrm{d}s$ and $B_p = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \sin(ps)\mathrm{d}s$ are the Fourier coefficients. By applying the above equality to the formulas of expectation and variance, we can get

$$V(Y) \approx 2 \sum_{p=1}^{\infty} \left(A_p^2 + B_p^2\right) \tag{17}$$

The first-order sensitivity index can be defined as

$$S_i = \frac{V_i}{V(Y)} = \frac{2 \sum_{q=1}^{\infty} \left(A_{q\cdot\omega_i}^2 + B_{q\cdot\omega_i}^2\right)}{2 \sum_{p=1}^{\infty} \left(A_p^2 + B_p^2\right)} \approx \frac{\sum_{q=1}^{M} \left(A_{q\cdot\omega_i}^2 + B_{q\cdot\omega_i}^2\right)}{\sum_{i=1}^{n} \sum_{q=1}^{M} \left(A_{q\cdot\omega_i}^2 + B_{q\cdot\omega_i}^2\right)} \tag{18}$$

where V_i is the estimated conditional variance of the ith parameter and M is the maximum harmonic usually taken to be 4 or higher. A large index means a significant first-order effect.

A lot of transfer functions have been proposed to provide uniformly distributed samples in the n-dimensional unit hypercube. Saltelli et al. (1999) suggested a popular periodic transfer function

$$X_i(s) = \frac{1}{2} + \frac{1}{\pi} \arcsin(\sin \omega_i s + \varphi_i) \tag{19}$$

where φ_i is a random phase shift chosen uniformly in $[0,\ 2\pi)$. The advantage of this function is that the starting point of the curve can be anywhere within the unit hypercube. By selecting N_r sets $\{\varphi_1, \varphi_2, \ldots, \varphi_n\}$, N_r search curves can then be generated, and this procedure was named "resampling" by Saltelli et al. (1999). The sample size of FAST is therefore

$$N = N_r(2M\omega_{\max} + 1) \tag{20}$$

where $\omega_{\max}$ is the maximum frequency. The minimum sample size is $2M\omega_{\max} + 1$ when there is only a single search curve.

Schaibly and Shuler (1973) applied FAST to two chemical reaction systems involving sets of coupled nonlinear rate equations and verified its effectiveness in determining the parameter sensitivities of nonlinear complex systems. FAST was adopted by Collins and Avissar (1994) and Rodríguez-Camino and Avissar (1998) to estimate the relative importance of land surface model (LSM) parameters to the variability of surface heat fluxes.

2.2.3 Extended Fourier Amplitude Sensitivity Test

Saltelli et al. (1999) proposed an extension of the FAST to calculate parameter total effect, which is known as extended FAST (EFAST). Assign a frequency ω_i for the ith ($i = 1, 2, \cdots, n$) parameter and a different frequency ω_{i-} for all the remaining parameters, where $i-$ means all parameters but the ith one. By evaluating the spectrum at the frequency ω_{i-} and higher harmonics $q \cdot \omega_{i-}$, the total sensitivity index of the ith parameter can be estimated by

$$S_{Ti} = 1 - \frac{V_{i-}}{V(Y)} = 1 - \frac{2\sum_{q=1}^{\infty}\left(A^2_{q\cdot\omega_{i-}} + B^2_{q\cdot\omega_{i-}}\right)}{2\sum_{p=1}^{\infty}\left(A_p^2 + B_p^2\right)} \approx 1 - \frac{\sum_{q=1}^{M}\left(A^2_{q\cdot\omega_{i-}} + B^2_{q\cdot\omega_{i-}}\right)}{\sum_{i=1}^{n}\sum_{q=1}^{M}\left(A^2_{q\cdot\omega_i} + B^2_{q\cdot\omega_i}\right)} \tag{21}$$

where V_{i-} is the estimated conditional variance except for the ith parameter. A large index means a significant total effect. EFAST needs to choose two frequencies ω_i and ω_{i-} for each parameter, and usually a higher value is assigned to ω_i. Unlike FAST method that all indices can be calculated from a single curve, EFAST requires n curves for calculating all n S_{Ti}. Therefore, the sample size needed by EFAST is

$$N = nN_r(2M\omega_{\max} + 1) \tag{22}$$

where $\omega_{\max} = \max\{\omega_i, \omega_{i-}\} \equiv \omega_i$ and N_r is the number of resampling times as in FAST. The minimum sample size for EFAST is $n(2M\omega_{\max} + 1)$.

Wang et al. (2013) adopted the EFAST method to analyze the parameter sensitivity of the World Food Studies (WOFOST) crop growth model. Other applications of the EFAST method can be found in Confalonieri et al. (2010) and Reusser et al. (2011).

2.2.4 McKay Correlation Ratios

McKay (1995) makes ANOVA-like decomposition of response variances for calculating correlation ratio, which is a ratio of the variance of expectation conditioned on one parameter and the total variances and is the representation of parameter main effect. Tong (2005) extended the idea for main effect analysis to two-way interaction effect analysis for uncorrelated parameters.

Let $E(Y)$ and $V(Y)$ be the expectation and variance of the response Y, respectively, thus $V(Y)$ can be decomposed as

$$V(Y) = V[E(Y|X_i)] + E[V(Y|X_i)] = V\left[E\left(Y|X_i, X_j\right)\right] + E\left[V\left(Y|X_i, X_j\right)\right] \quad (23)$$

where X_i and X_j are the ith and jth parameter, respectively, $V[E(Y|X_i)]$ is the variance of the conditional expectation of Y conditioned on X_i, $E[V(Y|X_i)]$ is the residual term measuring the estimated variance of Y by fixing X_i, $V[E(Y|X_i, X_j)]$ is the variance of the conditional expectation of Y conditioned on X_i and X_j, and $E[V(Y|X_i, X_j)]$ is the residual term measuring the estimated variance of Y by fixing X_i and X_j. The correlation ratios of McKay main effect and two-way interaction effect are defined, respectively, as

$$S_i = \frac{V[E(Y|X_i)]}{V(Y)} = \frac{V[E(Y|X_i)]}{V[E(Y|X_i)] + E[V(Y|X_i)]} \quad (24)$$

and

$$S_{ij} = \frac{V\left[E\left(Y|X_i,\ X_j\right)\right]}{V(Y)} = \frac{V\left[E\left(Y|X_i,\ X_j\right)\right]}{V\left[E\left(Y|X_i,\ X_j\right)\right] + E\left[V\left(Y|X_i,\ X_j\right)\right]} \quad (25)$$

The former measures the relative contribution of parameter X_i to the response variance, while the latter measures the relative contributions of parameters X_i and X_j together to the response variance. The higher the parameter correlation ratio is, the more significant the parameter effect is.

2.2.5 Sobol' Sensitivity Indices

The global method proposed by Sobol' (1993, 2001) is a milestone for global SA of nonlinear models, which makes ANOVA-like decomposition of response variances for calculating specific order sensitivity indices. This method has received much attention because it can provide accurate and robust sensitivity measures of any orders (Nossent et al. 2011; Wagener et al. 2009).

Let the function $Y = f(\mathbf{X}) = f(X_1, X_2, \ldots, X_n)$, where $X_i \in [0,\ 1]$ and $i = 1, 2, \cdots, n$. Assume that the model response can be decomposed into 2^n summands of increasing dimensions as

$$\begin{aligned} Y &= f(\mathbf{X}) \\ &= f_0 + \sum_{i=1}^{n} f_i(X_i) + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} f_{i,j}\left(X_i, X_j\right) + \cdots \\ &\quad + f_{1,2,\cdots,n}(X_1, X_2, \cdots, X_n) \end{aligned} \quad (26)$$

where f_0 is a constant, $f_i(X_i)$ are the functions of one parameter, and $f_{i,j}(X_i, X_j)$ are the functions of two parameters, etc. The above formula is called ANOVA representation of $f(\mathbf{X})$ if the integral of every summand is zero

$$\int_0^1 f_{i_1,\ldots,i_s}(X_{i_1},\ldots,X_{i_s})dX_k = 0, \text{ for } k = i_1,\ldots,i_s \tag{27}$$

where $1 \le i_1 < \cdots < i_s \le n$.

Assume that $f(\mathbf{X})$ is square integrable. The total response variance can be written as

$$V(Y) = \int_0^1 \cdots \int_0^1 f^2(\mathbf{X})\mathrm{d}\mathbf{X} - f_0^2 \tag{28}$$

While the contribution of a generic term $f_{i_1,\cdots,i_s}(1 \le i_1 < \cdots < i_s \le n)$ to the total variance can be written as

$$V_{i_1,\ldots,i_s} = \int_0^1 \cdots \int_0^1 f^2_{i_1,\ldots,i_s}(X_{i_1},\ldots,X_{i_s})\mathrm{d}X_{i_1}\ldots\mathrm{d}X_{i_s} \tag{29}$$

Thus the ANOVA-like decomposition of total variance can be expressed as

$$V(Y) = \sum_{s=1}^{n}\sum_{i_1<\cdots<i_s}^{n} V_{i_1,\cdots,i_s} = \sum_{i=1}^{n} V_i + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} V_{i,j} + \cdots + V_{1,\cdots,n} \tag{30}$$

The Sobol' sensitivity indices are defined as

$$S_{i_1,\cdots,i_s} = \frac{V_{i_1,\cdots,i_s}}{V(Y)}, \quad 1 \le i_1 < \cdots < i_s \le n \tag{31}$$

and the sum of all indices is $\sum_{s=1}^{n}\sum_{i_1<\cdots<i_s}^{n} S_{i_1,\cdots,i_s} = 1$.

Theoretically, this global method can compute sensitivity index of any order. However, the computation for higher-order terms is impractical when the number of parameters is large. Homma and Saltelli (1996) provided a simple way for computing the total effect of each parameter as

$$S_{Ti} = S_i + S_{i,ci} = 1 - S_{ci} \tag{32}$$

where S_i and $S_{i,ci}$ are representations of first-order effect and higher-order effect, respectively, and S_{ci} is the sum of all the $S_{i_1,\cdots,i_s}$ terms that excludes the index i.

Example 2.2

Situation. Suppose that we are planning to join a bungee jumping club and would like to enjoy real excitement but stay alive by approaching the ground as close as possible. The minimum distance to the ground during the oscillation can be

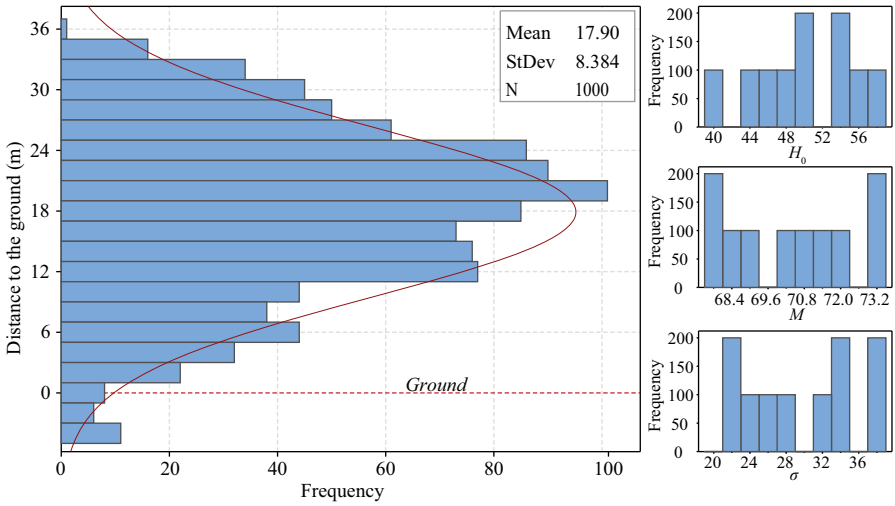

Fig. 4 Probability distributions of the model response and parameters for the bungee experiment

expressed as $h_{\min} = H_0 - 2Mg/(k\sigma)$, where $g = 9.8\ \mathrm{m/s^2}$ is the acceleration of gravity and $k = 1.5\ \mathrm{N/m}$ is the elastic constant of one strand. The uncertainties are from the height of the platform H_0 (40–60 m), the mass of our body M (67–74 kg), and the number of strands in the cord σ (20–40). In view of uncertainties, please evaluate the risk of safe jumps, and identify the main impact factors.

Solution. Assume that the uncertain factors H_0, M, and σ are uniformly distributed in their ranges. One thousand LH samples are generated from the three-dimensional parameter space with each of the three factors having ten levels. Samples are then designed as separate experiments to run the model. The probability distributions of the model response and parameters are shown in Fig. 4. In 980 cases out of 1000, the jump is successful, that is, the risk of this bungee experiment is about 2%.

It is observable from the function that the model is linear on factors H_0 and M/σ, but not on M and σ separately. Therefore, we use the model-independent method to analyze parameter sensitivities. Fig. 5 shows the main effect and two-way interaction effect of different experimental factors using McKay's method. The analysis indicates that the number of strands (σ) is the most important factor, the height of the platform (H_0) is of secondary importance, and the influence of the mass of our body (M) is ignorable. The relative contribution of the first two factors (σ and H_0) account for more than 90% of the total variance. Hence, we should not waste much time on the accuracy of our weight but focus on the accuracy of the number of strands and the height of the platform.

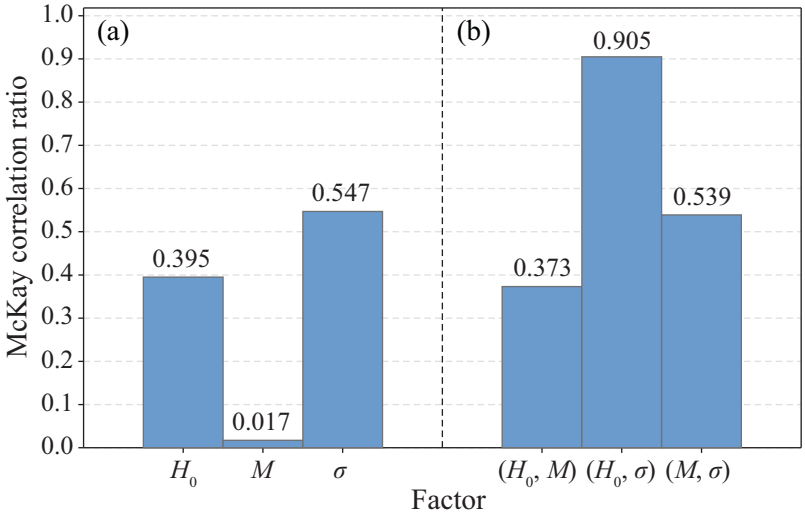

Fig. 5 McKay main effect and two-way interaction effect analysis for the bungee experiment

2.3 Regression-Based Methods

Since today's models are becoming more detailed and realistic and hence have many parameters, high computational costs are prohibitive to quantitative SA using variance-based methods. Therefore, regression-based methods, such as linear regression (LR) (Galton 1886), multivariate adaptive regression splines (MARS) (Friedman 1991), sum-of-trees (SOT) (Chipman et al. 2010), delta test (DT) (Pi and Peterson 1994), and Gaussian process (GP) (MacKay 1998) models, are often employed to screen out insensitive parameters by qualitatively evaluating parameter overall effects. On the other hand, many researchers have also investigated the possibility of replacing the original simulation models with computationally cheaper surrogate models (also called response surface models, metamodels, or emulators) that perform a similar function (Borgonovo et al. 2012; Shahsavani and Grimvall 2011). Regression-based methods are often used to construct surrogate models to improve overall computational efficiency (Wang and Shan 2007). Quantitative SA can then be applied to the surrogate model if it has been proved to be effective for approximating the simulation model. Shahsavani and Grimvall (2011) demonstrated the performance of variance-based SA using surrogate models. Detailed review of available surrogate models can be found in Storlie et al. (2009) and Razavi et al. (2012). A brief introduction of the LR, MARS, SOT, DT, and GP models is given as follows:

2.3.1 Linear Regression

The generalized form of a LR model relating model parameters and response is

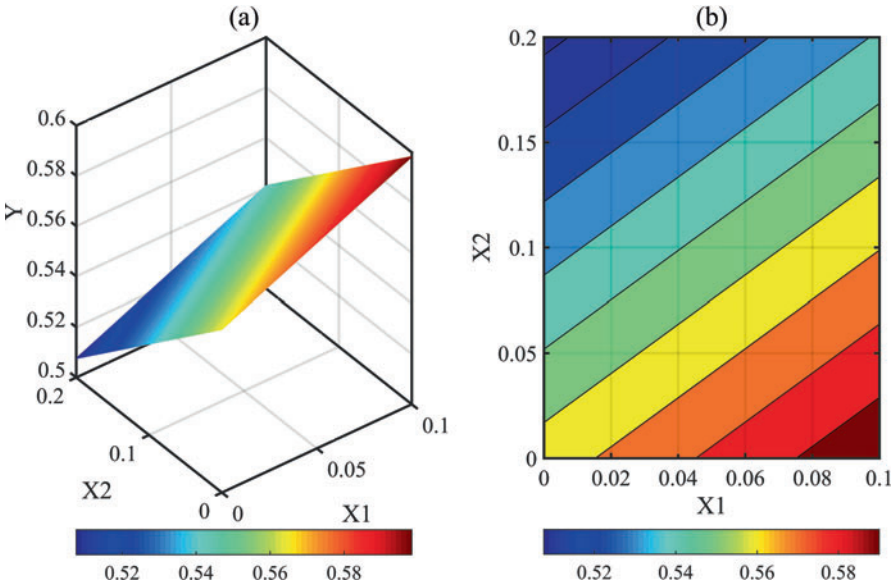

Fig. 6 (a) response surface and (b) contour plot for a two-parameter LR model

$$\hat{Y} = b_0 + \sum_{i=1}^{n} b_i X_i \tag{33}$$

While the actual simulation model response can be expressed as

$$Y = b_0 + \sum_{i=1}^{n} b_i X_i + \varepsilon \tag{34}$$

where b_0 is the intercept, b_i is the regression coefficient of the ith parameter X_i, and $\varepsilon \sim N(0, \sigma^2)$ is the error term between the simulation model response and the regression model response. Under the assumption of Gaussian errors, the regression coefficients can be obtained using the least squares approach. An example showing the parameter-response relationship of the LR model is given in Fig. 6.

Utilizing the means and standard deviations of the parameter and response, the LR model is usually normalized to

$$\frac{\hat{Y}^k - \bar{Y}}{\hat{s}} = \sum_{i=1}^{n} \frac{b_i \hat{s}_i}{\hat{s}} \frac{X_i^k - \bar{X}_i}{\hat{s}_i} \tag{35}$$

where k represents the kth sample and $b_i \hat{s}_i / \hat{s}$ is defined as the standardized regression coefficient (SRC), with

$$\hat{s} = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}\left(Y^k - \bar{Y}\right)^2} \tag{36}$$

and

$$\hat{s}_i = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}\left(X_i^k - \bar{X}_i\right)^2} \tag{37}$$

as the standard deviations of Y and X_i, respectively. A positive value of SRC indicates that X_i and Y tend to move in the same direction, otherwise in the opposite direction. The larger the absolute value of SRC, the more sensitive is the parameter X_i.

SRC is a sensitivity measure based on the linear parameter-response relationship, and it cannot provide reliable indication of parameter sensitivity when the underlying relationship is nonlinear. However, the transformation of raw data into ranks has been proven to work quite well when the parameter-response relationship is monotonic (Iman and Conover 1979). Therefore, standardized rank regression coefficient (SRRC) can be used as parameter sensitivity measure for nonlinear but monotonic problems.

2.3.2 Multivariate Adaptive Regression Splines

Multivariate adaptive regression splines (MARS) is an extension of LR models, which makes use of the LR, the mathematical construction of splines, the binary recursive partitioning, and brute search intelligent algorithms (Friedman 1991). The general form of MARS can be represented as

$$Y = f(\mathbf{X}) = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m}\left[s_{km}\left(X_{v(k,\,m)} - t_{km}\right)\right]_+^q \tag{38}$$

where a_0 is a constant, a_m are fitting coefficients, M is the number of basis functions, K_m is the number of factors in the mth basis function, s_{km} takes on values of either 1 or -1 and indicates the right or left sense of the associated step function, $v(k, m)$ is the label of the independent parameter and $1 \le v(k, m) \le n$, t_{km} indicates the knot location, and the exponent q is the order of the spline approximation. The subscript "+" means the function is a truncated power function

$$\left[s_{km}\left(X_{v(k,\,m)} - t_{km}\right)\right]_+^q = \begin{cases} \left[s_{km}\left(X_{v(k,\,m)} - t_{km}\right)\right]^q & s_{km}\left(X_{v(k,\,m)} - t_{km}\right) > 0 \\ 0 & s_{km}\left(X_{v(k,\,m)} - t_{km}\right) \le 0 \end{cases} \tag{39}$$

MARS builds a model in two phases: the forward pass and the backward pass, which is the same as that used by recursive partitioning trees. The forward pass

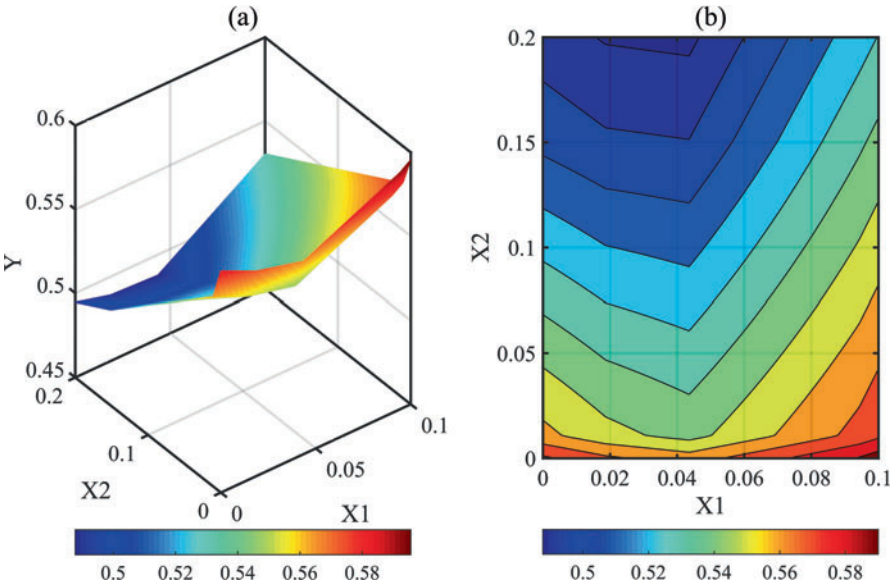

Fig. 7 (a) response surface and (b) contour plot for a two-parameter MARS model

builds an overfit model using all parameters, while the backward pass prunes the overfit model by removing one parameter from the model at a time. The lack-of-fit criterion called generalized cross-validation (GCV) criterion is then computed for both the overfit model and the pruned model

$$GCV(M) = \frac{1}{N} \frac{\sum_{l=1}^{N} \left[Y_l - \hat{f}_M(\mathbf{X}_l) \right]^2}{\left[1 - C(M)/N \right]^2} \tag{40}$$

with

$$C(M) = 1 + c(M)d \tag{41}$$

where N is the number of observations in the data set, M is the number of non-constant basis functions in the model $\hat{f}_M(\mathbf{X})$, d is the effective degrees of freedom, and $c(M)$ is a penalty for adding a basis function. An example showing the parameter-response relationship of the MARS model is given in Fig. 7.

The increase in GCV values between the pruned model and the overfitted model can be considered as the importance measure of the removed parameter (Steinberg et al. 1999). The most important parameter is the one that, when omitted, degrades the model fit the most. The score of the ith ($i = 1, 2, \ldots, n$) parameter is given by

$$S_i = \frac{\Delta g(i)}{\max\{\Delta g(1), \Delta g(2), \ldots, \Delta g(n)} \times 100 \tag{42}$$

where $\Delta g(i)$ is the increase in *GCV* when *i*th parameter is removed. The larger the *GCV* increase, the more important is the removed parameter.

2.3.3 Sum-of-Trees

Sum-of-trees (SOT) model is fundamentally a classification (or Bayesian) additive regression tree model with multivariate components (Chipman et al. 2010). Let T denotes a binary tree consisting of a set of interior node decision rules and a set of terminal nodes, and let $M = \{\mu_1, \mu_2, \ldots, \mu_b\}$ denotes a set of values associated with each of the b terminal nodes of T. Thus the SOT model can be represented as

$$Y = \sum_{j=1}^{m} g(\mathbf{X}; T_j, M_j) + \varepsilon \tag{43}$$

where for each binary regression tree T_j and its associated terminal node values M_j, $g(\mathbf{X}; T_j, M_j)$ is the function which assigns $\mu_{ij} \in M_j$ to parameter set $\mathbf{X}$; m is the total number of trees, and $\varepsilon \sim (0, \sigma^2)$. An example showing the parameter-response relationship of the SOT model is given in Fig. 8.

The residual sum of squares is used as the criteria for node splitting. A parameter that has the maximum decrease of residual sum of squares will be chosen to split the

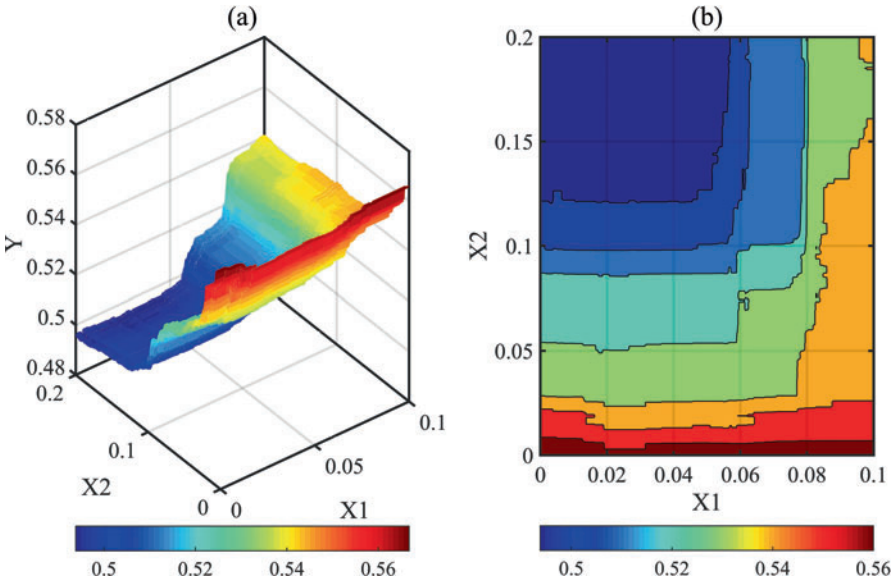

Fig. 8 (**a**) response surface and (**b**) contour plot for a two-parameter SOT model

node. The splitting process will not be stopped until per terminal node has minimum number of data points. The total number of splits for each parameter is then taken as the scoring criterion of sensitivity. The score for ith parameter is expressed as

$$S_i = \frac{p(i)}{\max\{p(1),\ p(2),\ \cdots,\ p(n)\}} \times 100 \tag{44}$$

where $p(i)$ is the number of splits for ith parameter. The more splits the parameter has, the more sensitive is the parameter.

An illustration of the SOT model is given in Fig. 9. As can be seen from this figure, the two-dimensional space is split into seven subspaces (i.e., subtrees) by six splitting nodes, and the number of splits for X_1 and X_2 is four and two, respectively. Therefore, the sensitivity scores for them are 100 and 50, respectively.

2.3.4 Delta Test

Delta test (DT) is based on the nearest neighbor method for estimating the variance of the residuals (Pi and Peterson 1994). It is founded on the hypothesis of the continuity of the regression function, i.e., if two sample points are close in the

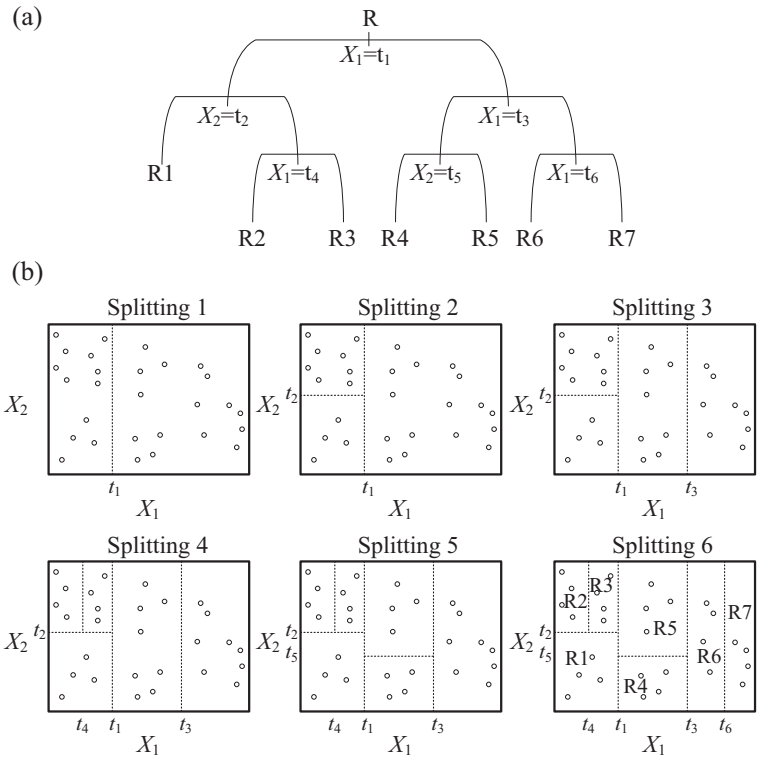

Fig. 9 Illustration of the SOT method for a two-parameter problem

parameter space, the responses of these two points will be close enough in the response space. Or else, it can be explained by the influence of noise. Assume that we have n parameters and sample points $\mathbf{X}_k \in [0,\ 1]^n$ for $1 \le k \le N$. Let $Y_k = f(\mathbf{X}_k) + \varepsilon_k$, where f is a continuous function with bounded first and second partial derivatives and the residuals $\varepsilon_k \sim (0,\ \sigma^2)$. Then the points $(\mathbf{X}_k,\ Y_k)_{k=1}^N$ comprise imitation data set. Let the DT metric that is restricted to the parameter subset space S be

$$\delta_s = \frac{1}{N}\sum_{k=1}^{N}\left(Y_k - Y_{N_s(k)}\right)^2 \approx Var(\varepsilon) \tag{45}$$

where the nearest neighbor of kth sample is

$$N_s(k) = \arg\min_{l \neq k} \|\mathbf{X}_k - \mathbf{X}_l\|_S^2 \tag{46}$$

and the semi-norm

$$\|\mathbf{X}_k - \mathbf{X}_l\|_S^2 = \sum_{p \in S}\left(\mathbf{X}_k^{(p)} - \mathbf{X}_l^{(p)}\right)^2 \tag{47}$$

Thus the DT metrics for all $2^n - 1$ non-empty parameter subsets can be calculated. Fig. 10 presents an illustration of the nearest neighbors of a point in different subset spaces. An example showing the parameter-response relationship of the nearest neighbor model is given in Fig. 11.

DT was proposed for parameter selection by Eirola et al. (2008). It takes the subset of parameters that minimize the noise variance from all the parameter combinations as sensitive ones. However, this procedure needs an efficient search algorithm to find this subset of parameter combinations. This search process can be too time-consuming, and usually it is impossible to do an exhaustive search of all

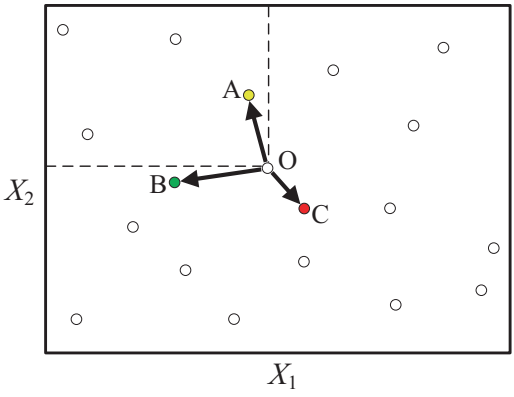

Fig. 10 Illustration of the nearest neighbor of DT method for a two-parameter problem, where points A, B, and C are nearest neighbors of point O in subset spaces $\{X_1\}$, $\{X_2\}$, and $\{X_1, X_2\}$, respectively

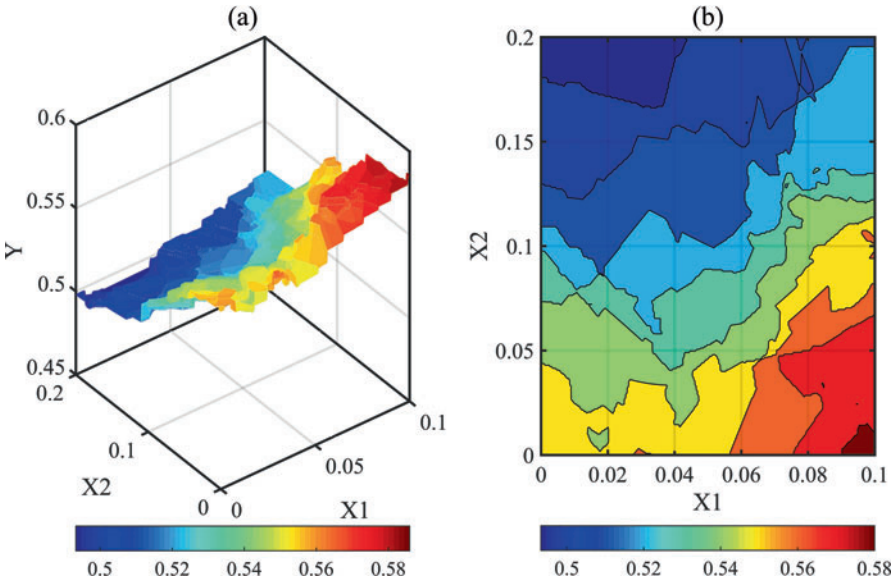

Fig. 11 (a) response surface and (b) contour plot for a two-parameter nearest neighbor model

combinations. DT assesses the final choice using forward sweep and uses genetic algorithm to speed up the search. The first 50 subsets which have the lowest value of DT metrics are taken for sensitivity scoring. The score of the ith ($i = 1, 2, \cdots, n$) parameter is given by

$$S_i = \frac{\sum_{m=1}^{50} \delta_S^{(m)} \times I_i^{(m)}}{\sum_{m=1}^{50} \delta_S^{(m)}} \times 100 \tag{48}$$

where $\delta_S^{(m)}$ is the DT metric of the mth subset and $I_i^{(m)} = 1$ if the ith parameter is included in the mth subset, or else $I_i^{(m)} = 0$. A higher score means a more sensitive parameter.

2.3.5 Gaussian Process

Gaussian process (GP) method characterizes simulation responses over the parameter space as a multivariate Gaussian distribution (MacKay 1998). Let the training data set consist of parameter vectors $\{\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_N\}$ and the corresponding set of response values $\{Y_1, Y_2, \cdots, Y_N\}$, where N is the sample size. A GP is a collection of variables $\mathbf{Y} = \{Y_1, Y_2, \cdots, Y_N\}$ which have a joint probability distribution

$$P(\mathbf{Y}|\mu(\mathbf{X}),\ \mathbf{C}) = \frac{1}{Z}\exp\left\{-\frac{1}{2}[\mathbf{Y}-\mu(\mathbf{X})]^T\mathbf{C}^{-1}[\mathbf{Y}-\mu(\mathbf{X})]\right\} \tag{49}$$

where $\mathbf{C} = \{C(\mathbf{X}_k, \mathbf{X}_l;\ \Theta)\}_{k,\ l=1}^{N}$ is a parameterized covariance function with hyperparameters Θ, $\mu(\mathbf{X})$ is the mean function of the distribution, and Z is the normalization factor. That is, random function $\mathbf{Y}$ can be specified by its mean function $\mu(\mathbf{X})$ and covariance function $C(\mathbf{X},\ \mathbf{X}')$. An example showing the parameter-response relationship of the GP model is given in Fig. 12.

Different kinds of mean and covariance functions lead to different GPs. Gibbs and MacKay (1997) presented a software package called "Tpros" for regression problem using GP. The form of covariance function given by them is

$$C(\mathbf{X}_k, \mathbf{X}_l;\ \Theta) = \theta_1 \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\frac{\left(\mathbf{X}_k^{(i)} - \mathbf{X}_l^{(i)}\right)^2}{r_i^2}\right] + \theta_2 + \varepsilon_{kl}(\mathbf{X}_k, \mathbf{X}_l) \tag{50}$$

where θ_1 is the hyperparameter that gives the overall vertical scale, θ_2 is the hyperparameter that gives the vertical uncertainty, $\varepsilon_{kl}(\mathbf{X}_k, \mathbf{X}_l)$ is the noise model, $\mathbf{X}_k^{(i)}$ and $\mathbf{X}_l^{(i)}$ are the ith components of sample points $\mathbf{X}_k$ and $\mathbf{X}_l$, respectively, and r_i is the length scale that characterizes the distance in the direction of ith parameter over which $\mathbf{Y}$ is expected to vary significantly.

As can be seen from the covariance function that when two sample points are close (with respect to their length scales) in parameter space, the exponent is small,

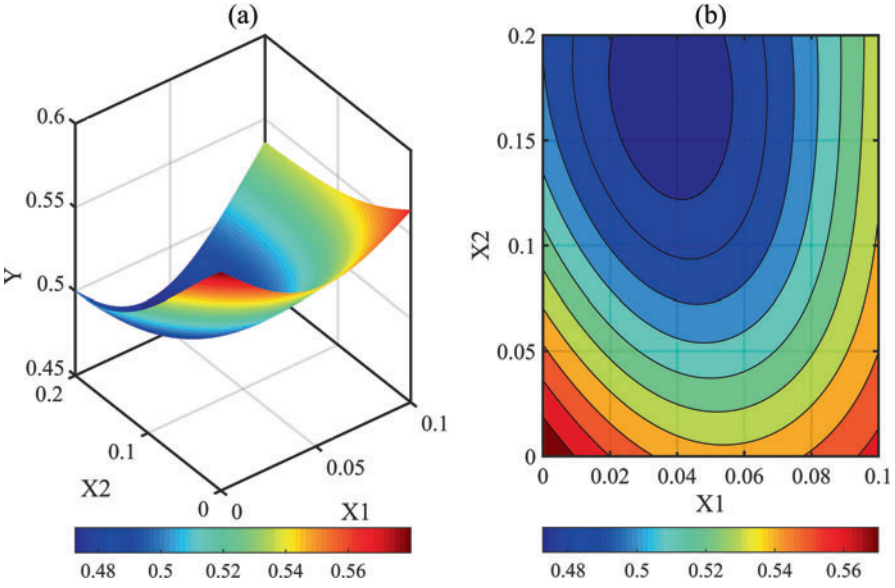

Fig. 12 (a) response surface and (b) contour plot for a two-parameter GP model

and thus the covariance is large, which means their corresponding response values are highly correlated. That is to say, points that are close in parameter space give rise to similar response values. On the contrary, a smaller length scale of a parameter leads to larger difference of response values for two close sample points, which means a more significant influence of this parameter on model response. Therefore, the length scales can be taken as the scoring criteria for parameter screening. The score for ith parameter is expressed as

$$S_i = \frac{1/r_i}{\max\{1/r_1, 1/r_2, \cdots, 1/r_n\}} \times 100 \tag{51}$$

Example 2.3

Situation. Consider the artificial computational model proposed by Morris (1991) which contains 20 parameters and has the form as $Y = \beta_0 + \sum_{i=1}^{20} \beta_i w_i + \sum_{i<j}^{20} \beta_{i,j} w_i w_j + \sum_{i<j<l}^{20} \beta_{i,j,l} w_i w_j w_l + \sum_{i<j<l<s}^{20} \beta_{i,j,l,s} w_i w_j w_l w_s$, where $w_i = 2(X_i - 1/2)$ except for $i = 3$, 5, and 7, where $w_i = 2[1.1X_i/(X_i + 0.1) - 1/2]$. Each parameter X_i is supposed to be uniformly distributed in [0, 1]. Coefficients with relatively large values are as follows: $\beta_i = 20$, with $i = 1, \ldots, 10$; $\beta_{i,j} = -15$, with $i, j = 1, \ldots, 6$; $\beta_{i,j,l} = -10$, with, $i, j, l = 1, \ldots, 5$; and $\beta_{i,j,l,s} = 5$, with $i, j, l, s = 1, \ldots, 4$. The remainders of the first- and second-order coefficients are independently generated from a standard normal distribution. The remainders of the third- and fourth-order coefficients are set to zero. Assess parameter sensitivity, and reduce parameter dimensionality to a reasonable number.

Solution. We design 1000 LH experiments with each parameter having 1000 levels to run the model. Surrogate-based methods as MARS, SOT, DT, and GP are then adopted to qualitatively evaluate parameter sensitivity scores and quantitatively evaluate Sobol' total sensitivity indices (Sobol'-t). Analysis results for the 20-parameter test function are shown in Fig. 13. As it can be seen from the figure, parameter sensitivity rankings vary across SA methods, but parameter categories of different methods are consistent, that is, parameters 11 to 20 have lower sensitivities than the other 10 parameters. They can therefore be regarded as insensitive parameters and set to any fixed values over their ranges.

3 Which SA Methods to Use?

Many researchers have reviewed some popular SA methods in specific scientific fields, such as chemical reactions (Saltelli et al. 2005; Turányi 1990), ecological modeling (Cariboni et al. 2007), environmental modeling (Hamby 1994; Helton 1993), risk assessment (Frey et al. 2003; Frey and Patil 2002), linear programming (Ward and Wendell 1990), surface hydrology, and water quality modeling (Reusser

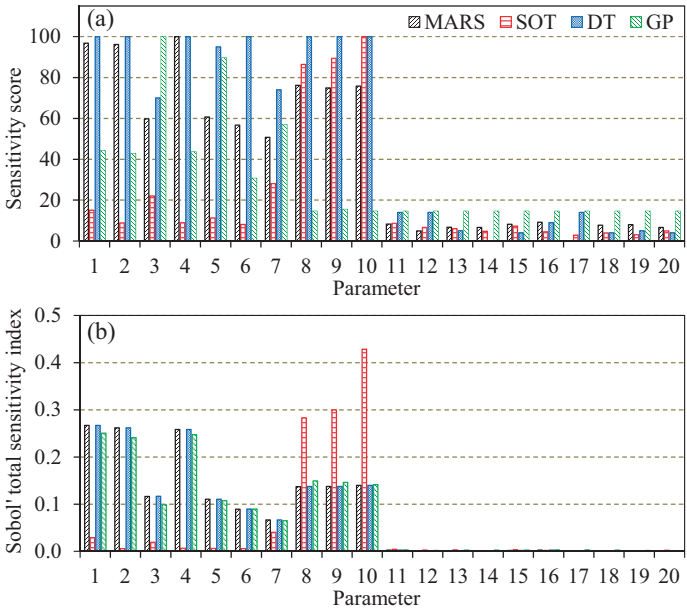

Fig. 13 Surrogate-based (**a**) parameter sensitivity scores and (**b**) Sobol' total sensitivity indices for the Morris 20-parameter test function

et al. 2011), among many other disciplines. Here we provide an extensive discussion on suitability of some commonly used SA methods for different applications. Given the myriad of SA methods, there is a need for further investigations regarding which methods should be used for specific problems.

Various attempts have been made over the years to answer the above question. Campolongo and Saltelli (1997) compared the performances of MOAT screening, SRC, and Sobol' sensitivity indices for SA of an environmental model. Their comparison showed that MOAT screening is the most efficient method, while Sobol' sensitivity indices are the most robust method. Saltelli and Bolado (1998) investigated the relationship between FAST method and Sobol' sensitivity indices and showed that FAST method is equivalent to the first-order Sobol' sensitivity indices but is computationally more efficient than Sobol's method. Saltelli et al. (1999) showed that EFAST method is more efficient than Sobol's method in computing the total-effect indices. By contrasting two variance-based global methods – EFAST and Sobol' sensitivity indices – with the most widely used local method OAT, Saltelli et al. (2000) showed that the variance-based global SA methods are robust, model independent, and computationally convenient. Furthermore, EFAST is numerically more efficient than Sobol's method. By defining different ranges of variation, Lenhart et al. (2002) compared two forms of a partial derivative-based local method using the hydrologic model SWAT (Arnold et al. 1998). Results indicate that both approaches provide similar results and hence can be considered as equivalent. Marino et al. (2008) reviewed and compared partial rank

correlation coefficient (PRCC) with the EFAST method. Their results show that PRCC relies on the monotonicity assumption between parameter and response, whereas EFAST is computationally more expensive. Reusser et al. (2011) compared FAST, EFAST, and Sobol' sensitivity indices by applying them to the hydrologic model TOPMODEL (Beven 1997) in a small mountainous catchment. Their comparison shows that the three methods give comparable results, while FAST is computationally more efficient. Tang et al. (2007) compared four SA methods, including the local analysis using parameter estimation software (PEST), regional SA (RSA), ANOVA, and Sobol's method, for the application to the lumped Sacramento soil moisture accounting model (SAC-SMA) (Burnash et al. 1973) coupled with SNOW-17 (Anderson 1973). Their conclusion is that ANOVA and Sobol's method are overall superior to RSA and PEST, and ANOVA is more efficient but less robust than Sobol's method. Confalonieri et al. (2010) performed SA on a crop model using the MOAT screening; regression-based methods with LH, MC, and QMC sampling; and two variance-based methods: EFAST and Sobol' sensitivity indices. Their experiments demonstrate that the simplest method MOAT screening produced results comparable to those obtained by methods more computationally expensive. Sun et al. (2012) employed the OAT, MOAT, and RSA methods to assess parameter sensitivities of a water quality model. They concluded that the three methods are complementary, but the use of OAT method for interpreting parameter behaviors should be avoided unless the model uncertainty is small. A comparison of the interpretation and computational cost of the local SA method, MOAT screening, and Sobol's method was made by Wainwright et al. (2014) in the application to a pressure propagation problem. The three SA methods were shown to give similar interpretations and importance rankings of model parameters. Although Sobol's method is illustrated to be computationally less efficient than MOAT screening, it is nonsubstitutable because of its capability of interpreting the contribution of each parameter to the response uncertainty.

Generally, global, quantitative, and model-independent SA methods are advocated for all problem settings where finite parameter variations are involved (Saltelli 1999). Considering the large computational cost of global SA, Foglia et al. (2009) even argued that local SA is sufficient to identify insensitive parameters in preliminary model evaluation. Wainwright et al. (2014) think the reason for this argument is because the value of global methods has not been fully appreciated. Global methods are often limited to parameter importance ranking, even though they can provide additional information for systematic understanding of model behavior. The strengths and limitations of several qualitative and quantitative global SA methods were discussed by Gan et al. (2014). Overall, qualitative SA methods are more efficient but less accurate and robust than quantitative ones. The stepwise SA framework proposed by Gan et al. (2015), using qualitative SA method for preliminary parameter screening and then quantitative SA method for assessing each parameter's contribution to the variance of model response, is an effective and efficient solution for understanding and simplifying complex system models.

4 Summary

The importance of SA for computer-based models is universally recognized. We reviewed a number of commonly used SA methods as gradient-based, variance-based, and regression-based methods. Features and applicability of those methods were described and illustrated with a few examples. Merits and limitations were also given by reviewing the literature on those different SA methods.

The choice of an appropriate SA method depends on (1) the number of considered parameters, (2) the computational costs of the model and the SA method, (3) the ability of the SA method to account for nonlinear and non-monotonic parameter-response relationship, and (4) the ability of the SA method to account for parameter interactions. We therefore emphasize several recommendations for SA based on the selection criteria and characteristics of different methods: (1) local SA methods are effective only for linear and monotonic problem, (2) qualitative global SA methods should be adopted when a single model run takes a significant amount of time and/or the model has a large number of uncertain parameters, (3) quantitative global SA methods can be used to evaluate parameter main effect, interaction effect, and total effect when the parameter dimensionality is low and the model is computationally efficient, and (4) surrogate models are computationally cheaper than time-consuming simulation models and can be used to obtain approximate results for quantitative global SA.

Parameters are often ranked according to the values of specific SA measure, which allows the analyst to focus research efforts on the most sensitive parameters (factor prioritization) and simplify the model by fixing the least sensitive parameters (factor fixing) (Saltelli et al. 2008). Parameter rankings may vary between different SA measures under the same settings, but the sensitivity categories of the parameters should be the same. Therefore, it is important that two types of errors should be avoided in categorizing model parameters, that is, insensitive parameters are classified as sensitive ones (Type I error) and, conversely, sensitive parameters are taken as insensitive ones (Type II error).

Acknowledgment This study was supported by the National Natural Science Foundation of China (41505092) and National Key Research and Development Program of China (2017YFC1404000).

References

E.A. Anderson, *National Weather Service River Forecast System – Snow Accumulation and Ablation Model* (NOAA, Silver Spring, 1973)

J.G. Arnold, R. Srinivasan, R.S. Muttiah, J.R. Williams, Large area hydrologic modeling and assessment. Part I: model development. J. Am. Water Resour. Assoc. **34**, 73–89 (1998). https://doi.org/10.1111/j.1752-1688.1998.tb05961.x

D.L. Beres, D.M. Hawkins, Plackett – Burman technique for sensitivity analysis of many-parametered models. Ecol. Model. **141**, 171–183 (2001). https://doi.org/10.1016/S0304-3800(01)00271-X

K. Beven, TOPMODEL: a critique. Hydrol. Process. **11**, 1069–1085 (1997). https://doi.org/10.1002/(SICI)1099-1085(199707)11:9<1069::AID-HYP545>3.0.CO;2-O

E. Borgonovo, W. Castaings, S. Tarantola, Model emulation and moment-independent sensitivity analysis: an application to environmental modelling. Environ. Model. Softw. **34**, 105–115 (2012). https://doi.org/10.1016/j.envsoft.2011.06.006

G.E.P. Box, J.S. Hunter, The 2^{k-p} fractional factorial designs: part I. Technometrics **3**, 311–351 (1961a). https://doi.org/10.1080/00401706.1961.10489951

G.E.P. Box, J.S. Hunter, The 2^{k-p} fractional factorial designs: part II. Technometrics **3**, 449–458 (1961b). https://doi.org/10.1080/00401706.1961.10489967

R.J.C. Burnash, R.L. Ferral, R.A. McGuire, *A Generalized Streamflow Simulation System: Conceptual Modeling for Digital Computers* (US Department of Commerce, National Weather Service, Sacramento, 1973)

F. Campolongo, A. Saltelli, Sensitivity analysis of an environmental model: an application of different analysis methods. Reliab. Eng. Syst. Saf. **57**, 49–69 (1997). https://doi.org/10.1016/S0951-8320(97)00021-5

F. Campolongo, J. Cariboni, A. Saltelli, An effective screening design for sensitivity analysis of large models. Environ. Model. Softw. **22**, 1509–1518 (2007). https://doi.org/10.1016/j.envsoft.2006.10.004

F. Campolongo, A. Saltelli, J. Cariboni, From screening to quantitative sensitivity analysis. A unified approach. Comput. Phys. Commun. **182**, 978–988 (2011). https://doi.org/10.1016/j.cpc.2010.12.039

J. Cariboni, D. Gatelli, R. Liska, A. Saltelli, The role of sensitivity analysis in ecological modelling. Ecol. Model. **203**, 167–182 (2007). https://doi.org/10.1016/j.ecolmodel.2005.10.045

H.A. Chipman, E.I. George, R.E. McCulloch, BART: Bayesian additive regression trees. Ann. Appl. Stat. **4**, 266–298 (2010). https://doi.org/10.1214/09-AOAS285

D.C. Collins, R. Avissar, An evaluation with the Fourier amplitude sensitivity test (FAST) of which land-surface parameters are of greatest importance in atmospheric modeling. J. Clim. **7**, 681–703 (1994). https://doi.org/10.1175/1520-0442(1994)007<0681:AEWTFA>2.0.CO;2

R. Confalonieri, G. Bellocchi, S. Bregaglio, M. Donatelli, M. Acutis, Comparison of sensitivity analysis techniques: a case study with the rice model WARM. Ecol. Model. **221**, 1897–1906 (2010). https://doi.org/10.1016/j.ecolmodel.2010.04.021

S.A. Cryer, P.L. Havens, Regional sensitivity analysis using a fractional factorial method for the USDA model GLEAMS. Environ. Model. Softw. **14**, 613–624 (1999). https://doi.org/10.1016/S1364-8152(99)00003-1

R.I. Cukier, C.M. Fortuin, K.E. Shuler, A.G. Petschek, J.H. Schaibly, Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I theory. J. Chem. Phys. **59**, 3873–3878 (1973). https://doi.org/10.1063/1.1680571

C. Daniel, On varying one factor at a time. Biometrics **14**, 430–431 (1958)

R.E. Dickinson, A. Henderson-Sellers, P.J. Kennedy, M.F. Wilson, *Biosphere-Atmosphere Transfer Scheme (BATS) for the NCAR Community Climate Model* (NCAR, Boulder, 1986). https://doi.org/10.5065/D6668B58

E. Dion, L. VanSchalkwyk, E.F. Lambin, The landscape epidemiology of foot-and-mouth disease in South Africa: a spatially explicit multi-agent simulation. Ecol. Model. **222**, 2059–2072 (2011). https://doi.org/10.1016/j.ecolmodel.2011.03.026

Q. Duan, J. Schaake, V. Andreassian, S. Franks, G. Goteti, H.V. Gupta, Y.M. Gusev, F. Habets, A. Hall, L. Hay, T. Hogue, M. Huang, G. Leavesley, X. Liang, O.N. Nasonova, J. Noilhan, L. Oudin, S. Sorooshian, T. Wagener, E.F. Wood, Model parameter estimation experiment (MOPEX): an overview of science strategy and major results from the second and third workshops. J. Hydrol. **320**, 3–17 (2006). https://doi.org/10.1016/j.jhydrol.2005.07.031

E. Eirola, E. Liitiäinen, A. Lendasse, F. Corona, M. Verleysen, Using the Delta test for variable selection, in ESANN 2008 Proceedings, European Symposium on Artificial Neural Networks, Bruges, Belgium (2008)

L. Foglia, M.C. Hill, S.W. Mehl, P. Burlando, Sensitivity analysis, calibration, and testing of a distributed hydrological model using error-based weighting and one objective function. Water Resour. Res. **45**, W6427 (2009). https://doi.org/10.1029/2008WR007255

A. Francos, F.J. Elorza, F. Bouraoui, G. Bidoglio, L. Galbiati, Sensitivity analysis of distributed environmental simulation models: understanding the model behaviour in hydrological studies at the catchment scale. Reliab. Eng. Syst. Saf. **79**, 205–218 (2003). https://doi.org/10.1016/S0951-8320(02)00231-4

H.C. Frey, S.R. Patil, Identification and review of sensitivity analysis methods. Risk Anal. **22**, 553–578 (2002). https://doi.org/10.1111/0272-4332.00039

H.C. Frey, A. Mokhtari, T. Danish, Evaluation of selected sensitivity analysis methods based upon applications to two food safety process risk models. Department of Civil, Construction, and Environmental Engineering, North Carolina State University, Raleigh, NC (2003)

J.H. Friedman, Multivariate adaptive regression splines. Ann. Stat. **19**, 1–67 (1991). https://doi.org/10.1214/aos/1176347963

F. Galton, Regression towards mediocrity in hereditary stature. J. Anthropol. Inst. G. B. Irel. **15**, 246–263 (1886)

Y. Gan, Q. Duan, W. Gong, C. Tong, Y. Sun, W. Chu, A. Ye, C. Miao, Z. Di, A comprehensive evaluation of various sensitivity analysis methods: a case study with a hydrological model. Environ. Model. Softw. **51**, 269–285 (2014). https://doi.org/10.1016/j.envsoft.2013.09.031

Y. Gan, X.-Z. Liang, Q. Duan, H.I. Choi, Y. Dai, H. Wu, Stepwise sensitivity analysis from qualitative to quantitative: application to the terrestrial hydrological modeling of a conjunctive surface-subsurface process (CSSP) land surface model. J. Adv. Model. Earth Syst. **7**, 648–669 (2015). https://doi.org/10.1002/2014MS000406

M. Gibbs, D.J.C. MacKay, *Efficient implementation of Gaussian processes*. Unpublished manuscript (1997)

J. Grant, K.J. Curran, T.L. Guyondet, G. Tita, C. Bacher, V. Koutitonsky, M. Dowd, A box model of carrying capacity for suspended mussel aquaculture in Lagune de la Grande-Entrée, Iles-de-la-Madeleine, Québec. Ecol. Model. **200**, 193–206 (2007). https://doi.org/10.1016/j.ecolmodel.2006.07.026

A. van Griensven, T. Meixner, S. Grunwald, T. Bishop, M. Diluzio, R. Srinivasan, A global sensitivity analysis tool for the parameters of multi-variable catchment models. J. Hydrol. **324**, 10–23 (2006). https://doi.org/10.1016/j.jhydrol.2005.09.008

D.M. Hamby, A review of techniques for parameter sensitivity analysis of environmental models. Environ. Monit. Assess. **32**, 135–154 (1994). https://doi.org/10.1007/BF00547132

J.C. Helton, Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. Reliab. Eng. Syst. Saf. **42**, 327–367 (1993). https://doi.org/10.1016/0951-8320(93)90097-I

A. Henderson-Sellers, A factorial assessment of the sensitivity of the BATS land-surface parameterization scheme. J. Clim. **6**, 227–247 (1993). https://doi.org/10.1175/1520-0442(1993)006<0227:AFAOTS>2.0.CO;2

B. Henderson-Sellers, A. Henderson-Sellers, Sensitivity evaluation of environmental models using fractional factorial experimentation. Ecol. Model. **86**, 291–295 (1996). https://doi.org/10.1016/0304-3800(95)00066-6

T. Homma, A. Saltelli, Importance measures in global sensitivity analysis of nonlinear models. Reliab. Eng. Syst. Saf. **52**, 1–17 (1996). https://doi.org/10.1016/0951-8320(96)00002-6

R.L. Iman, W.J. Conover, The use of the rank transform in regression. Technometrics **21**, 499–509 (1979). https://doi.org/10.1080/00401706.1979.10489820

A.J. Jakeman, R.A. Letcher, J.P. Norton, Ten iterative steps in development and evaluation of environmental models. Environ. Model. Softw. **21**, 602–614 (2006). https://doi.org/10.1016/j.envsoft.2006.01.004

J.P.C. Kleijnen, Sensitivity analysis and related analyses: a review of some statistical techniques. J. Stat. Comput. Simul. **57**, 111–142 (1997). https://doi.org/10.1080/00949659708811805

T. Lenhart, K. Eckhardt, N. Fohrer, H.G. Frede, Comparison of two different approaches of sensitivity analysis. Phys. Chem. Earth **27**, 645–654 (2002). https://doi.org/10.1016/S1474-7065(02)00049-9

D.J.C. MacKay, Introduction to Gaussian processes, in *Neural Networks and Machine Learning*, ed. by C.M. Bishop, (Springer, Berlin, 1998). pp 133–165

S. Marino, I.B. Hogue, C.J. Ray, D.E. Kirschner, A methodology for performing global uncertainty and sensitivity analysis in systems biology. J. Theor. Biol. **254**, 178–196 (2008). https://doi.org/10.1016/j.jtbi.2008.04.011

L.S. Matott, J.E. Babendreier, S.T. Purucker, Evaluating uncertainty in integrated environmental models: a review of concepts and tools. Water Resour. Res. **45**, W6421 (2009). https://doi.org/10.1029/2008WR007301

M.D. McKay, *Evaluating Prediction Uncertainty* (Los Alamos National Laboratory (LANL), Los Alamos, 1995)

M.D. McKay, R.J. Beckman, W.J. Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics **21**, 239–245 (1979). https://doi.org/10.1080/00401706.2000.10485979

A. Mokhtari, H.C. Frey, Sensitivity analysis of a two-dimensional probabilistic risk assessment model using analysis of variance. Risk Anal. **25**, 1511–1529 (2005). https://doi.org/10.1111/j.1539-6924.2005.00679.x

M.D. Morris, Factorial sampling plans for preliminary computational experiments. Technometrics **33**, 161–174 (1991). https://doi.org/10.1080/00401706.1991.10484804

J. Nossent, P. Elsen, W. Bauwens, Sobol' sensitivity analysis of a complex environmental model. Environ. Model. Softw. **26**, 1515–1525 (2011). https://doi.org/10.1016/j.envsoft.2011.08.010

A.B.. Owen, Orthogonal arrays for computer experiments, integration and visualization. Stat. Sin. **2**, 439–452 (1992)

H. Pi, C. Peterson, Finding the embedding dimension and variable dependencies in time series. Neural Comput. **6**, 509–520 (1994). https://doi.org/10.1162/neco.1994.6.3.509

R.L. Plackett, J.P. Burman, The design of optimum multifactorial experiments. Biometrika **33**, 305–325 (1946). https://doi.org/10.1093/biomet/33.4.305

O. Rakovec, M.C. Hill, M.P. Clark, A.H. Weerts, A.J. Teuling, R. Uijlenhoet, Distributed evaluation of local sensitivity analysis (DELSA), with application to hydrologic models. Water Resour. Res. **50**, 409–426 (2014). https://doi.org/10.1002/2013WR014063

S. Razavi, B.A. Tolson, D.H. Burn, Review of surrogate modeling in water resources. Water Resour. Res. **48**, W7401 (2012). https://doi.org/10.1029/2011WR011527

B. Renard, D. Kavetski, G. Kuczera, M. Thyer, S.W. Franks, Understanding predictive uncertainty in hydrologic modeling: the challenge of identifying input and structural errors. Water Resour. Res. **46**, W5521 (2010). https://doi.org/10.1029/2009WR008328

D.E. Reusser, W. Buytaert, E. Zehe, Temporal dynamics of model parameter sensitivity for computationally expensive models with the Fourier amplitude sensitivity test. Water Resour. Res. **47**, W7551 (2011). https://doi.org/10.1029/2010WR009947

E. Rodríguez-Camino, R. Avissar, Comparison of three land-surface schemes with the Fourier amplitude sensitivity test (FAST). Tellus A **50**, 313–332 (1998). https://doi.org/10.3402/tellusa.v50i3.14529

A. Saltelli, Sensitivity analysis: could better methods be used? J. Geophys. Res. **104**, 3789–3793 (1999). https://doi.org/10.1029/1998JD100042

A. Saltelli, R. Bolado, An alternative way to compute Fourier amplitude sensitivity test (FAST). Comput. Stat. Data Anal. **26**, 445–460 (1998). https://doi.org/10.1016/S0167-9473(97)00043-1

A. Saltelli, S. Tarantola, K.P.S. Chan, A quantitative model-independent method for global sensitivity analysis of model output. Technometrics **41**, 39–56 (1999). https://doi.org/10.2307/1270993

A. Saltelli, S. Tarantola, F. Campolongo, Sensitivity analysis as an ingredient of modeling. Stat. Sci. **15**, 377–395 (2000). https://doi.org/10.1214/ss/1009213004

A. Saltelli, M. Ratto, S. Tarantola, F. Campolongo, Sensitivity analysis for chemical models. Chem. Rev. **105**, 2811–2827 (2005). https://doi.org/10.1021/cr040659d

A. Saltelli, M. Ratto, S. Tarantola, F. Campolongo, Sensitivity analysis practices: strategies for model-based inference. Reliab. Eng. Syst. Saf. **91**, 1109–1125 (2006). https://doi.org/10.1016/j.ress.2005.11.014

A. Saltelli, M. Ratto, T. Andres, F. Campolongo, *Global Sensitivity Analysis: The Primer* (Wiley, Chichester, 2008)

J.H. Schaibly, K.E. Shuler, Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. II applications. J. Chem. Phys. **59**, 3879–3888 (1973). https://doi.org/10.1063/1.1680572

D. Shahsavani, A. Grimvall, Variance-based sensitivity analysis of model outputs using surrogate models. Environ. Model. Softw. **26**, 723–730 (2011). https://doi.org/10.1016/j.envsoft.2011.01.002

A. Sieber, S. Uhlenbrook, Sensitivity analyses of a distributed catchment model to verify the model structure. J. Hydrol. **310**, 216–235 (2005). https://doi.org/10.1016/j.jhydrol.2005.01.004

I.M. Sobol', Quasi-Monte Carlo methods. Prog. Nucl. Energy **24**, 55–61 (1990). https://doi.org/10.1016/0149-1970(90)90022-W

I.M. Sobol', Sensitivity analysis for nonlinear mathematical models. Math. Model. Comput. Exp. **1**, 407–414 (1993)

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

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

C.B. Storlie, L.P. Swiler, J.C. Helton, C.J. Sallaberry, Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. Reliab. Eng. Syst. Saf. **94**, 1735–1763 (2009). https://doi.org/10.1016/j.ress.2009.05.007

X.Y. Sun, L. Newham, B. Croke, J.P. Norton, Three complementary methods for sensitivity analysis of a water quality model. Environ. Model. Softw. **37**, 19–29 (2012). https://doi.org/10.1016/j.envsoft.2012.04.010

B. Tang, Orthogonal array-based Latin hypercubes. J. Am. Stat. Assoc. **88**, 1392–1397 (1993). https://doi.org/10.2307/2291282

Y. Tang, P. Reed, T. Wagener, K. van Werkhoven, Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. Hydrol. Earth Syst. Sci. **11**, 793–817 (2007). https://doi.org/10.5194/hess-11-793-2007

C. Tong, *PSUADE User's Manual* (Lawrence Livermore National Laboratory (LLNL), Livermore, 2005)

T. Turányi, Sensitivity analysis of complex kinetic systems. Tools and applications. J. Math. Chem. **5**, 203–248 (1990). https://doi.org/10.1007/BF01166355

L. Uusitalo, A. Lehikoinen, I. Helle, K. Myrberg, An overview of methods to evaluate uncertainty of deterministic models in decision support. Environ. Model. Softw. **63**, 24–31 (2015). https://doi.org/10.1016/j.envsoft.2014.09.017

T. Wagener, K. van Werkhoven, P. Reed, Y. Tang, Multiobjective sensitivity analysis to understand the information content in streamflow observations for distributed watershed modeling. Water Resour. Res. **45**, W2501 (2009). https://doi.org/10.1029/2008WR007347

H.M. Wainwright, S. Finsterle, Y. Jung, Q. Zhou, J.T. Birkholzer, Making sense of global sensitivity analyses. Comput. Geosci. UK **65**, 84–94 (2014). https://doi.org/10.1016/j.cageo.2013.06.006

W.E. Walker, P. Harremoës, J. Rotmans, J.P. van der Sluijs, M.B. van Asselt, P. Janssen, M.P. Krayer Von Krauss, Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. Integr. Assess. **4**, 5–17 (2003). https://doi.org/10.1076/iaij.4.1.5.16466

G.G. Wang, S. Shan, Review of metamodeling techniques in support of engineering design optimization. J. Mech. Des. **129**, 370–380 (2007). https://doi.org/10.1115/1.2429697

J. Wang, X. Li, L. Lu, F. Fang, Parameter sensitivity analysis of crop growth models based on the extended Fourier amplitude sensitivity test method. Environ. Model. Softw. **48**, 171–182 (2013). https://doi.org/10.1016/j.envsoft.2013.06.007

C. Wang, Q. Duan, C. Tong, W. Gong, A GUI platform for uncertainty quantification of complex dynamical models. Environ. Model. Softw. **76**, 1–12 (2016). https://doi.org/10.1016/j.envsoft.2015.11.004

J.E. Ward, R.E. Wendell, Approaches to sensitivity analysis in linear programming. Ann. Oper. Res. **27**, 3–38 (1990). https://doi.org/10.1007/BF02055188

H. Weyl, Mean motion. Am. J. Math. **60**, 889–896 (1938)