Environmental Modelling & Software 41 (2013) 39–52

Contents lists available at SciVerse ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

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

A R T I C L E I N F O

Article history:
Received 14 December 2011
Received in revised form
7 October 2012
Accepted 8 October 2012
Available online 17 December 2012

Keywords:
Global sensitivity analysis
Statistical emulator
DTVGM
Response surface model
RSMSobol' method

A B S T R A C T

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.

1364-8152/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.envsoft.2012.10.009

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, $i = 1, 2, \ldots, n$, 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

$$d_i(X) = \frac{f(X_1, \ldots, X_{i-1}, X_i + \Delta, X_{i+1}, \ldots, X_n) - f(X_1, \ldots, X_{i-1}, X_i, \ldots, X_n)}{\Delta} \tag{1}$$

where Δ is a value in $\{1/(p-1), 2/(p-1), \ldots, 1 - 1/(p-1)\}$, p is the number of levels, and $X = (x_1, \ldots, x_{i-1}, x_i, \ldots, x_n)$ is a random sample in the parameter space so that the transformed point $(x_1, \ldots, x_{i-1}, x_i + \Delta, \ldots, x_n)$ 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 $X^{(1)}, X^{(2)}, \ldots, X^{(R)}$ 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:

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

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

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 non-parametric methods such as Friedman’s multivariate adaptive regression splines