Available online at www.sciencedirect.com
ScienceDirect
Reliability Engineering and System Safety 93 (2008) 1563–1573
RELIABILITY ENGINEERING & SYSTEM SAFETY
www.elsevier.com/locate/ress

Uncertainty and sensitivity analysis for models with correlated parameters

Chonggang Xu, George Zdzislaw Gertner*

Department of Natural Resources and Environmental Sciences, University of Illinois, 1102 South Goodwin Avenue, Urbana, IL 61801, USA

Received 15 February 2006; received in revised form 25 April 2007; accepted 7 June 2007
Available online 26 June 2007

Abstract

When conducting sensitivity and uncertainty analysis, most of the global sensitivity techniques assume parameter independence. However, it is common that the parameters are correlated with each other. For models with correlated inputs, we propose that the contribution of uncertainty to model output by an individual parameter be divided into two parts: the correlated contribution (by the correlated variations, i.e. variations of a parameter which are correlated with other parameters) and the uncorrelated contribution (by the uncorrelated variations, i.e. the unique variations of a parameter which cannot be explained by any other parameters). So far, only a few studies have been conducted to obtain the sensitivity index for a model with correlated input. But these studies do not distinguish between the correlated and uncorrelated contribution of a parameter. In this study, we propose a regression-based method to quantitatively decompose the total uncertainty in model output into partial variances contributed by the correlated variations and partial variances contributed by the uncorrelated variations. The proposed regression-based method is then applied in three test cases. Results show that the regression-based method can successfully measure the uncertainty contribution in the case where the relationship between response and parameters is approximately linear.

Keywords: Correlated parameters; Latin hypercube sampling; Linear regression; Sensitivity analysis; Uncertainty analysis

1. Introduction

Nowadays many complex models have been developed in physics, chemistry, environmental sciences and risk analysis. A consequence of model complexity is that the uncertainty in both model structure and parameter estimation has increased. Thus, the identification and representation of uncertainty is recognized as an essential component in model application [1–3]. In the study of uncertainty, we need to know how much uncertainty there is in the model output (uncertainty analysis) and where the uncertainty comes from (sensitivity analysis).

Many uncertainty and sensitivity analysis techniques are now available [4–16]. In the case of simple mathematical models, Taylor series expansions can be used to approximate the model and the analytical differential sensitivity index can be derived [5,8,17]. However, for complex models, it would be difficult to use the Taylor series approximation, and more advanced techniques are needed. McRae et al. [18] and Saltelli et al. [8] classified the sensitivity techniques into two groups: local sensitivity analysis methods and global sensitivity analysis methods. The local sensitivity analysis techniques examine the local response of the output(s) by varying input parameters one at a time by holding other parameters at central values. The global sensitivity techniques examine the global response (averaged over the variation of all the parameters) of the model output(s) by exploring a finite (or even an infinite) region. The local sensitivity analysis is easy to implement. However, local sensitivity analysis can only inspect one point at a time. Thus, more and more studies nowadays are using global sensitivity analysis methods instead of local sensitivity analysis. Many global sensitivity analysis techniques are available, such as Fourier amplitude sensitivity test (FAST) [18–24]; fractional factorial

*Corresponding author. Tel.: +1 217 333 9346; fax: +1 217 244 3219.
E-mail address: gertner@uiuc.edu (G.Z. Gertner).

0951-8320/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ress.2007.06.003

design method [25–27]; Plackett–Burman technique [28]; Morris method [29]; sampling-based methods [3,10,11,16,17]; Sobol's method [30]; and McKay's method based on an one-way ANOVA [31].

When conducting sensitivity and uncertainty analysis, most of the global sensitivity techniques such as FAST, Sobol's method and sampling-based methods rely on the assumption of parameter independence. However, in many cases, the parameters are correlated with one another. For example, in meteorology, the central pressure of the storm is correlated with the radius of the maximum wind [1]. For models with correlated inputs, we propose that the contribution of uncertainty to model output by an individual parameter be divided into two parts: the correlated contribution (by the correlated variations, i.e. the variations of a parameter which are correlated with other parameters) and the uncorrelated contribution (by the uncorrelated variations, i.e. the unique variations of a parameter which cannot be explained by any other parameters). This distinction between correlated and uncorrelated contribution of uncertainty for an individual parameter is very important, since it can help us decide if we need to focus on the correlated variations among specific parameters (if the correlated contribution dominates) or the parameter itself (if the uncorrelated contribution dominates).

By now, only a few studies have been conducted to obtain the sensitivity index for models with correlated input [1,32–37]. Iman et al. [38] proposed the partial correlation coefficient (a correlation between model output and parameter that remains after adjusting for other parameters) as a measure of parameter sensitivity for models with correlated input based on Latin hypercube sampling [1,38]. Bedford [34] proposed a Gram–Schmidt orthogonalization to obtain the first-order Sobol's indices for correlated input. Saltelli et al. [33,39] proposed a correlation ratio method based on McKay's method. Fang et al. [32] proposed sequential sampling to approximate the differential sensitivity index. However, all methods except for the partial correlation coefficient method only provide an overall sensitivity index of one parameter, which does not distinguish the correlated or uncorrelated contribution of one parameter. The partial correlation coefficient method only provides a relative sensitivity index for the uncorrelated contribution but not for the correlated contribution. If two variables are highly correlated, then neither one will show up as being important in a sensitivity analysis based on the partial correlation coefficient. In addition, if the model is close to linear for each variable, then each variable will have a partial correlation coefficient close to 1 in absolute value even though their effect on the uncertainty in model results may be very different.

In this study, we use a regression-based method to quantitatively decompose the variances in the model output into partial variances contributed by the correlated and uncorrelated variations of parameters.

2. Methods

2.1. Variance decomposition by regression with independent input

For a model $y = f(x)$ $[x = (x_1, x_2 \ldots, x_i, \ldots, x_K)]$, if the effect of each parameter x_i on model output is linear and we only care about the main effects of each parameter, the model can be simplified as follows:

$$y = \beta_0 + \sum_{i=1}^{K} \beta_i x_i + e, \tag{1}$$

where $\beta_0 \ldots \beta_k$ are regression coefficients and e is the error. We can use the regression based on Eq. (1) to obtain the sensitivity index of each parameter given that the model at hand is approximately linear. If the parameters are mutually independent and are independent of e, based on Eq. (1), the variance in model output can be decomposed as follows:

$$V = \sum_{i=1}^{K} V_i + V_e, \tag{2}$$

where V_e is the variance contributed by the error in Eq. (1) and V_i is the variance contributed by parameter x_i. V_e is the variance in model output unaccounted for after removing the linear main effects of all the parameters. It can include non-linear effects or higher order interaction effects which can be present in the model.

Latin hypercube sampling can be used to explore the sensitivity of parameters based on the regression using Eq. (1). The Latin hypercube sampling method is an efficient sampling method compared to the simple random sampling method [16,40]. Latin hypercube sampling first stratifies the range of each variable (i.e. x_i) into N disjoint intervals of equal probability and then one random value is drawn from each interval. Thus, each parameter will have N random values. The sampled random values of all parameters are randomly permuted to form a sample of size N,

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1K} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{NK} \end{bmatrix}, \tag{3}$$

where the ith column of the matrix represents the N permuted random sample points for x_i. The model is then run on the sample and N response values of y are generated. Namely,

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{bmatrix} f(x_{11}, \ldots, x_{1i}, .., x_{1K}) \\ \vdots \\ f(x_{N1}, \ldots, x_{Ni}, .., x_{NK}) \end{bmatrix}. \tag{4}$$