Reliability Engineering and System Safety 94 (2009) 596–603

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Parameter uncertainty effects on variance-based sensitivity analysis

W. Yu, T.J. Harris *

Department of Chemical Engineering, Queen's University, Kingston, Ont., Canada K7L 3N6

ARTICLE INFO

Article history:
Received 31 December 2006
Received in revised form
17 June 2008
Accepted 25 June 2008
Available online 3 July 2008

Keywords:
Variance-based sensitivity analysis
Parameter uncertainty
Variance decomposition
Linear-in-parameter model

ABSTRACT

In the past several years there has been considerable commercial and academic interest in methods for variance-based sensitivity analysis. The industrial focus is motivated by the importance of attributing variance contributions to input factors. A more complete understanding of these relationships enables companies to achieve goals related to quality, safety and asset utilization. In a number of applications, it is possible to distinguish between two types of input variables—regressive variables and model parameters. Regressive variables are those that can be influenced by process design or by a control strategy. With model parameters, there are typically no opportunities to directly influence their variability. In this paper, we propose a new method to perform sensitivity analysis through a partitioning of the input variables into these two groupings: regressive variables and model parameters. A sequential analysis is proposed, where first an sensitivity analysis is performed with respect to the regressive variables. In the second step, the uncertainty effects arising from the model parameters are included. This strategy can be quite useful in understanding process variability and in developing strategies to reduce overall variability. When this method is used for nonlinear models which are linear in the parameters, analytical solutions can be utilized. In the more general case of models that are nonlinear in both the regressive variables and the parameters, either first order approximations can be used, or numerically intensive methods must be used.

1. Introduction

Sensitivity analysis (SA) methods are well known in the engineering literature. They have been used across various disciplines such as weather forecasting, chemical engineering and economics. Readers are referred to [1–4] and the references contained therein for recent reviews of these topics and a number of extensions and modifications.

The objective of SA is to ascertain how the model output depends on its input factors. Three general classes of SA techniques can be defined [4]:

- factor screening methods (e.g. one-at-a-time (OAT) experiments, factorial designs);
- local methods (e.g. differential or nominal value analysis);
- global methods (e.g. Monte Carlo analysis, FAST and Sobol's method).

Screening methods are typically qualitative tools providing only a ranking of the importance of the regressors/parameters. Screening methods can usually be further characterized as retaining properties of either local or global methods. One advantage of the screening methods is that they are computationally efficient. Their disadvantage is that they tend to address only a specific point, or local region in the regressor/parameter space. Mostly based on partial derivatives, local methods usually embody a univariate assessment framework among the regressors/parameters. As such they are valid only in a small range taken about the nominal values of the regressors/parameters. Global methods apportion the total variance of the output to each input factor and the interactions among the input factors. All input factors are allowed to vary simultaneously over their ranges taking into account the shapes of their probability density functions. Global methods are far more computationally demanding and involve various methods of sampling the input factor space (e.g. random sampling, quasi-random sampling, Latin hypercube sampling (LHS), etc.) [5]. A review of global SA methods, including Monte Carlo based regression–correlation measures and variance-based method can be found in [3,6,7].

Consider a system described by a uniresponse model of the form

$$Y = f(X, \Theta), \tag{1}$$

where Y is the value of the output or response variable, and $[X, \Theta]$ are the input factors. f is the model that maps the input factors to

* Corresponding author. Tel.: +1 613 533 2765; fax: +1 613 533 6637.
E-mail addresses: weiyu@chee.queensu.ca (W. Yu), harrist@post.queensu.ca (T.J. Harris).

0951-8320/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ress.2008.06.016

the output. The input factors have been partitioned into two sets. X is an $m \times 1$ vector of regressive variables and Θ is a $p \times 1$ vector of model parameters. The regressive variables would typically correspond to the input settings of the model. In chemical engineering, these might correspond to material flow rates, temperatures, pressures, etc. The parameters are also inputs, but they are fundamentally different from the regressive inputs. In chemical engineering, these inputs include the physicochemical parameters of the model, such as those related to reaction kinetics and thermodynamic equilibria. Most often the physicochemical parameters are estimated from experiment data or they are obtained from standard correlations. In either case, there is always uncertainty with these values.

A variance-based SA for the model in Eq. (1) that accounts for uncertainties in both regressors, X, and parameters, Θ, can be conducted in a straightforward fashion using correlation ratio-based techniques [8,9], FAST [10–12], or Sobol' techniques [13] and variants on these methods. (In this paper we will utilize the latter two approaches.) Unfortunately, the full SA may not always provide entirely satisfactory results. In some cases, understanding and quantifying the regressor uncertainty may have a higher priority than the parameter uncertainty. For example, in controller performance assessment [14], a variance decomposition of important process variables is undertaken with respect to measured or unmeasured variables. The results can be directly used to assist in output variance reduction. After the preliminary SA is completed, the impact of parameter uncertainty effects on these results might be considered.

In this paper, the SA is approached in a sequential fashion. First, the sensitivity with respect to the regressor variables ($X's$) is determined using the nominal or expected values for the parameters. We call this regressive sensitivity analysis (RSA). The analysis with respect to the parameters is then undertaken using the results from the first step to determine the effects of the parameter uncertainty. This approach relies on a fundamental variance decomposition identity [15, p. 55] (see Eq. (14)). Considerable simplification arises when a model is linear in the parameters, or can be adequately approximated as such. Polynomial models are quite simple, typically involving a bilinear product of the form $f(X,\Theta) = H^{\mathrm{T}}(X)\Theta$, where $H(X)$ is a polynomial in the regressive variables. If the model is linear in its parameters, the effects of the parameter uncertainties on the RSA can be deduced using the RSA and the first two moments of the distribution of the parameter uncertainties. When the model cannot be partitioned in this fashion, computational intensive methods, such as Monte Carlo techniques, must be applied.

The layout of the paper is as follows: In Section 2, the global variance-based SA is introduced. In Section 3, an approach for incorporating the parameter uncertainty effects on RSA for general nonlinear models is outlined. In most cases, numerically intensive methods are required to account for these uncertainties. However, for models that are linear in the parameters, analytical solutions are readily obtained. Some simple methods to cope with the more general case are proposed. In Section 4, two examples are used to illustrate the methodology. This is followed by the conclusions and a discussion of areas for future work.

2. Variance-based SA

Variance-based SA aims to ascertain how much a model output depends on each or some of its input factors. A review of variance-based SA can be found in [16]. For more detailed information, readers are referred to [17]. The importance of the given independent inputs $X_1, X_2, \ldots, X_m$ can be measured via the sensitivity index, which is defined as the fractional contribution to the output variance due to the uncertainties in inputs. To calculate the sensitivity index, it is first necessary to undertake a decomposition of the total output variance, $\mathrm{Var}(Y)$, using the decomposition formula [13,18,19,24]

$$\mathrm{Var}(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \cdots + V_{12\ldots m}, \tag{2}$$

where

$$V_i = \mathrm{Var}[E[Y|X_i = x_i^*]]$$

$$V_{ij} = \mathrm{Var}[E[Y|X_i = x_i^*, X_j = x_j^*]] - \mathrm{Var}[E[Y|X_i = x_i^*]] - \mathrm{Var}[E[[Y|X_j = x_j^*]] \tag{3}$$

and so on. $E[Y|X_i = x_i^*]$ denotes the expectation of Y conditional on X_i having a fixed value x_i^*, and Var stands for variance over all the possible values of x_i^*. The decomposition described in Eq. (2) is commonly referred to as analysis of variance (ANOVA) in the statistics literature [20,21].

The decomposition via Eq. (2) only holds if the factors, X_i, are independent of each other. Solutions for variance-based SA with correlated input factors are addressed in [22,23]. If a model has m factors, the total number of terms in Eq. (2) is as high as $2^m - 1$.

Sensitivity indices $S_{i_1,i_2,\ldots,i_s}$ [10,12,13,24] can be computed as

$$S_{i_1,i_2,\ldots,i_s} = \frac{V_{i_1,i_2,\ldots,i_s}}{\mathrm{Var}(Y)}. \tag{4}$$

All the $S_{i_1,i_2,\ldots,i_s}$ are nonnegative and their sum is

$$S = \sum_{i=1}^{m} S_i + \sum_{1 \leqslant i < j \leqslant m} S_{ij} + \cdots + S_{12\ldots m} = 1. \tag{5}$$

The total sensitivity index (TSI), S_{T_i}, of one factor, X_i, is defined as the sum of all the sensitivity indices involving that factor. For example, in a three-factor case, the TSI can be written as

$$\begin{aligned} S_{\mathrm{T}_1} &= S_1 + S_{12} + S_{13} + S_{123} \\ S_{\mathrm{T}_2} &= S_2 + S_{12} + S_{23} + S_{123} \\ S_{\mathrm{T}_3} &= S_3 + S_{23} + S_{13} + S_{123}. \end{aligned} \tag{6}$$

Evaluation of the sensitivity indices is most often accomplished through analytical means or applications of Monte Carlo methods [24]. Efficient numerical methods are required for large-scale problems [19]. The FAST [10,12] and Sobol's [13,24] methods have been developed to cope with this dimensionality problem. Many applications of these methods are described by Saltelli et al. [3] and Saltelli and Homma [24]. The latter reference contains a number of insightful applications of Monte Carlo methods to evaluate the sensitivity indices discussed in this paper. There are, however, a number of important special cases for which analytical results can be obtained.

2.1. Special case: bilinear models

Bilinear models are often used in empirical statistical models under the framework of factorial models. Frequently more complex models are approximated by bilinear models to facilitate analysis. A bilinear model is often represented as a polynomial of the form

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \sum_{i=1}^{n} \beta_{ii} X_i^2 + \sum_{i=1}^{n} \sum_{j>i}^{n} \beta_{ij} X_i X_j. \tag{7}$$

We will assume that the regressive variables $X_i, i = 1 \ldots n$ are independently distributed and follow some distributions with finite moments up to order four and that each regressive variable has zero mean and constant variance σ_i^2. This is not restrictive as a bilinear model can always be cast in this form. Using the ANOVA-like decomposition method, the variance decomposition of the