

# Multiple predictor smoothing methods for sensitivity analysis: Description of techniques

Curtis B. Storlie<sup>a</sup>, Jon C. Helton<sup>b,\*</sup>

<sup>a</sup>Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA

<sup>b</sup>Department of Mathematics and Statistics, Arizona State University, Tempe, AZ 85287-1804, USA

Received 26 September 2006; received in revised form 11 October 2006; accepted 11 October 2006

Available online 26 January 2007

## Abstract

The use of multiple predictor smoothing methods in sampling-based sensitivity analyses of complex models is investigated. Specifically, sensitivity analysis procedures based on smoothing methods employing the stepwise application of the following nonparametric regression techniques are described: (i) locally weighted regression (LOESS), (ii) additive models, (iii) projection pursuit regression, and (iv) recursive partitioning regression. Then, in the second and concluding part of this presentation, the indicated procedures are illustrated with both simple test problems and results from a performance assessment for a radioactive waste disposal facility (i.e., the Waste Isolation Pilot Plant). As shown by the example illustrations, the use of smoothing procedures based on nonparametric regression techniques can yield more informative sensitivity analysis results than can be obtained with more traditional sensitivity analysis procedures based on linear regression, rank regression or quadratic regression when nonlinear relationships between model inputs and model predictions are present.

© 2006 Elsevier Ltd. All rights reserved.

**Keywords:** Additive models; Epistemic uncertainty; Locally weighted regression; Nonparametric regression; Projection pursuit regression; Recursive partitioning regression; Scatterplot smoothing; Sensitivity analysis; Stepwise selection; Uncertainty analysis

## 1. Introduction

The importance of uncertainty analysis and sensitivity analysis as components of analyses for complex systems is almost universally recognized, where uncertainty analysis designates the determination of the uncertainty in analysis results that derives from the uncertainty in analysis inputs and sensitivity analysis designates the determination of the contributions of individual uncertain analysis inputs to the uncertainty in analysis results [1–11]. A number of approaches to uncertainty and sensitivity analysis have been developed, including differential analysis [12–17], response surface methodology [18–26], Monte Carlo analysis [27–38], and variance decomposition procedures [39–43]. Overviews of these approaches are available in several reviews [44–52].

The focus of this presentation is on Monte Carlo (i.e., sampling-based) approaches to uncertainty and sensitivity analysis. Such analyses involve the consideration of models of the form

$$\mathbf{y} = \mathbf{f}(\mathbf{x}), \quad (1.1)$$

where

$$\mathbf{y} = [y_1, y_2, \dots, y_n]^T \quad (1.2)$$

is a vector of analysis results and

$$\mathbf{x} = [x_1, x_2, \dots, x_n]^T \quad (1.3)$$

is a vector of imprecisely known analysis inputs. In general, the model  $\mathbf{f}$  can be quite large and involved (e.g., a system of nonlinear partial differential equations requiring numerical solution (see Ref. [53]) or possibly a sequence of complex, linked models as is the case in a probabilistic risk assessment for a nuclear power plant (see Refs. [54,55]) or a performance assessment for a radioactive waste disposal facility (see Refs. [56,57])); the vector  $\mathbf{y}$  of analysis results can be of high dimension and complex structure (e.g., the

\*Corresponding author. Department 1544, MS 0779, Sandia National Laboratories, Albuquerque, NM 87185-0779, USA.

Tel.: +1 505 284 4808; fax: +1 505 844 2348.

E-mail address: [jchelto@sandia.gov](mailto:jchelto@sandia.gov) (J.C. Helton).

elements of  $\mathbf{y}$  might be several hundred temporally or spatially dependent functions); and the vector  $\mathbf{x}$  of analysis inputs can also be of high dimension and complex structure (e.g., several hundred variables, with some variables corresponding to physical properties of the system under study and other variables corresponding to parameters in probability distributions or perhaps to designators for alternative models).

The uncertainty in the elements of  $\mathbf{x}$  is characterized by a sequence of probability distributions

$$D_1, D_2, \dots, D_{n_X}, \quad (1.4)$$

where  $D_j$  is a probability distribution characterizing the uncertainty in  $x_j$ . Correlations and other restrictions involving the relations between the  $x_j$  are also possible. Such distributions and any associated restrictions are intended to numerically capture the existing knowledge about the elements of  $\mathbf{x}$  and are often developed through an expert review process [58–73].

The uncertainty characterized by the distributions  $D_1, D_2, \dots, D_{n_X}$  in Eq. (1.4) is often referred to as epistemic uncertainty. Alternate designations for epistemic uncertainty include state of knowledge, subjective, reducible, and type B [74–82]. In particular, epistemic uncertainty derives from a lack of knowledge about the appropriate value to use for a quantity that is assumed to have a fixed value in the context of a particular analysis. In the conceptual and computational organization of an analysis, epistemic uncertainty is generally considered to be distinct from aleatory uncertainty, which arises from an inherent randomness in the behavior of the system under study [74–83].

Sampling-based uncertainty and sensitivity analyses are based on a sample

$$\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in_X}], \quad i = 1, 2, \dots, n_S, \quad (1.5)$$

from the possible values for  $\mathbf{x}$  generated in consistency with the distributions in Eq. (1.4) and any associated restrictions. Random sampling is one possibility for the generation of this sample. However, owing to its efficient stratification properties, Latin hypercube sampling is widely used in analyses of this type, especially when computationally intensive models are involved [27,37,38].

The analysis evaluations

$$\mathbf{y}_i = \mathbf{y}(\mathbf{x}_i) = \mathbf{f}(\mathbf{x}_i), \quad i = 1, 2, \dots, n_S, \quad (1.6)$$

provide a mapping between analysis inputs (i.e.,  $\mathbf{x}_i$ ) and analysis results (i.e.,  $\mathbf{y}_i$ ) that forms the basis for both uncertainty analysis and sensitivity analysis. Once the preceding mapping is available, the determination of uncertainty analysis results is generally straightforward and involves the generation of summary results such as histograms, density functions, cumulative distribution functions (CDFs), complementary cumulative distribution functions (CCDFs), and box plots for individual elements of  $\mathbf{y}$  [29, Section 6.5]). The determination of sensitivity analysis results involves the exploration of the preceding

mapping with techniques such as examination of scatterplots, regression analysis, correlation and partial correlation analysis, and searches for nonrandom patterns [29, Section 6.6].

The determination of sensitivity analysis results is generally more demanding than the determination of uncertainty analysis results. In particular, the popular regression and correlation-based techniques can fail to appropriately identify the effects of the individual elements of  $\mathbf{x}$  on the elements of  $\mathbf{y}$  when nonlinear and nonmonotonic relations are present [29, Section 6.6]. Possible approaches to sensitivity analysis to use in such situations include grid-based statistical analyses of scatterplots [30,84], distance-based statistical analyses of scatterplots [85–98], multidimensional Kolmogorov–Smirnov tests [99–102], rank-concordance tests [103,104], and classification trees [105,106]. However, the preceding approaches lack the intuitive appeal of regression-based approaches to sensitivity analysis. In particular, regression-based sensitivity analysis can be carried out in a sequential manner with variable importance being indicated by the order in which variables enter the regression model and by the fraction of total variance that can be accounted for as successive variables enter the regression model.

The purpose of this presentation is to describe regression-based techniques for sensitivity analysis that are based on multiple predictor smoothing methods. Such methods are conceptually consistent with regression-based methods that have been widely used in the past in sensitivity analysis [29, Section 6.6], but have the important advantage that they are capable of incorporating local changes in the relationship between a dependent variable (i.e., an element of  $\mathbf{y}$ ) and multiple independent variables (i.e., elements of  $\mathbf{x}$ ). As a result, these methods can be successfully applied in situations involving nonlinear relationships between analysis inputs and analysis results where more traditional regression-based approaches would fail to appropriately capture these relationships.

This presentation is divided into two parts. In this, the first part, traditional approaches to regression-based sensitivity analysis are briefly described (Section 2), nonparametric approaches to regression analysis based on local data smoothing are introduced (Section 3), algorithms for the stepwise implementation of the procedures described in Section 3 as part of a sensitivity analysis are described (Section 4), and a brief summary discussion is given (Section 5). Then, in the second part of the presentation [107], the described procedures are illustrated in sensitivity analyses involving both simple test problems and results from a performance assessment for a radioactive waste disposal facility (i.e., the Waste Isolation Pilot Plant) [56,57].

Although analyses for real systems almost always involve multiple output variables as indicated in conjunction with Eqs. (1.1)–(1.3), the following discussions assume that a single real-valued result of the form

$$y = f(\mathbf{x}) \quad (1.7)$$

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات