

Multiple Predictor Smoothing Methods for Sensitivity Analysis: Description of Techniques

Curtis B. Storlie^a and Jon C. Helton^b

^aDepartment of Statistics, North Carolina State University, Raleigh, NC 27695-8203 USA

^bDepartment of Mathematics and Statistics, Arizona State University, Tempe, AZ 85287-1804 USA

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.

Key Words: Additive models, Epistemic uncertainty, Locally weighted regression, Nonparametric regression, Projection pursuit regression, Recursive partitioning regression, Scatterplot smoothing, Sensitivity analysis, Stepwise selection, Uncertainty analysis.

Please send correspondence to:

Jon C. Helton
Department 6874, MS 0779
Sandia National Laboratories
Albuquerque, NM 87185-0779 USA
Phone: 505-284-4808
Fax: 505-844-2348
email: jchelto@sandia.gov

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

$$\mathbf{y} = [y_1, y_2, \dots, y_{nY}] \quad (1.2)$$

is a vector of analysis results and

$$\mathbf{x} = [x_1, x_2, \dots, x_{nX}] \quad (1.3)$$

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 (e.g., Ref. 53) or possibly a sequence of complex, linked models as is the case in a probabilistic risk assessment for a nuclear power plant (e.g., Refs. 54, 55) or a performance assessment for a radioactive waste disposal facility (e.g., Refs. 56, 57); the vector \mathbf{y} of analysis results can be of high dimension and complex structure (e.g., the 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

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

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_{nX} 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

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

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

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

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} (Sect. 6.5, Ref. 29). 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 (Sect. 6.6, Ref. 29).

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 (Sect. 6.6, Ref. 29). 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 (Sect. 6.6, Ref. 29), 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 (Sect. 2), nonparametric approaches to regression analysis based on local data smoothing are introduced (Sect. 3), algorithms for the stepwise implementation of the procedures described in Sect. 3 as part of a sensitivity analysis are described (Sect. 4), and a brief summary discussion is given (Sect. 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

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

is under consideration. Similarly,

$$y_i = f(\mathbf{x}_i), i = 1, 2, \dots, nS, \tag{1.8}$$

is used to represent the result of evaluating y with the sample in Eq. (1.5). This simplifies the notation and results in no loss in generality as the results under discussion are valid for individual elements of \mathbf{y} . All statistical analyses in this presentation are carried out within the R statistical computing environment,¹⁰⁸ which is an open source equivalent to the S -Plus statistical package.¹⁰⁹

2. Traditional Parametric Regression Models

Several parametric regression models used in sensitivity analysis are briefly reviewed. More information on such models can be obtained in a number of excellent texts (e.g., Ref. 110-114).

2.1 Linear Regression

Linear regression has long been the method of choice for researchers wishing to approximate a surface. This regression model is predicated on a relation of the form

$$y = \beta_0 + \sum_{j=1}^{nX} \beta_j x_j + \varepsilon, \quad (2.1)$$

where ε is a random error term with an expected value of zero (i.e., $E(\varepsilon) = 0$).

The approximate form of the relation in Eq. (2.1) is

$$\hat{y} = b_0 + \sum_{j=1}^{nX} b_j x_j, \quad (2.2)$$

where the b_j are typically estimated with least squares procedures from observations of the form $[\mathbf{x}_i, y_i]$, $i = 1, 2, \dots, nS$. In turn, the preceding approximation is often algebraically reformulated as

$$(\hat{y} - \bar{y})/\hat{s} = \sum_{j=1}^{nX} (b_j \hat{s}_j / \hat{s})(x_j - \bar{x}_j) / \hat{s}_j, \quad (2.3)$$

where

$$\bar{y} = \sum_{i=1}^{nS} y_i / nS, \quad \hat{s} = \left[\sum_{i=1}^{nS} (y_i - \bar{y})^2 / (nS - 1) \right]^{1/2}$$

$$\bar{x}_j = \sum_{i=1}^{nS} x_{ij} / nS, \quad \hat{s}_j = \left[\sum_{i=1}^{nS} (x_{ij} - \bar{x}_j)^2 / (nS - 1) \right]^{1/2}.$$

The coefficients $b_j \hat{s}_j / \hat{s}$ in Eq. (2.3) are called standardized regression coefficients. When the x_j are independent, $|b_j \hat{s}_j / \hat{s}|$ can be used as a measure of variable importance. Specifically, $|b_j \hat{s}_j / \hat{s}|$ indicates the effect of moving a variable away from its expected value by a fixed fraction of its standard deviation while holding all other variables fixed at their expected values. Statistical tests can be used to indicate if the coefficients in Eqs. (2.2) and (2.3) appear to be different from zero. However, in the context of sensitivity analysis, it is important to recognize that such tests are simply one form of guidance with respect to variable importance as the underlying distributional assump-

tions with respect to the error term ε are not satisfied when deterministic models are under consideration (Sect. 6.6.3, Ref. 29).

The following identity holds when the relation in Eq. (2.2) is estimated with least squares procedures:

$$\sum_{i=1}^{nS} (y_i - \bar{y})^2 = \sum_{i=1}^{nS} (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{nS} (\hat{y}_i - y_i)^2, \quad (2.4)$$

where \hat{y}_i denotes the estimate of y_i obtained from the regression model (Sect. 3.4, Ref. 110). In order from left to right, the three summations in the preceding equation are referred to as the total sum of squares (SS_{tot}), the regression sum of squares (SS_{reg}), and the residual sum of squares (SS_{res}). Since SS_{res} provides a measure of variability about the regression model,

$$R^2 = SS_{reg} / SS_{tot} = \sum_{i=1}^{nS} (\hat{y}_i - \bar{y})^2 / \sum_{i=1}^{nS} (y_i - \bar{y})^2 \quad (2.5)$$

provides a measure of the extent to which the regression model can match the observed results. Specifically, R^2 is close to 1 when the variation about the regression model is small (i.e., when SS_{res} is small relative to SS_{tot}), which indicates that the regression model is successful in matching the observed results. Similarly, R^2 is close to 0 when the variation about the regression model is large (i.e., when SS_{reg} is small relative to SS_{tot}), which indicates that the regression model is not successful in matching the observed results.

When linear regression is used as a sensitivity analysis technique, the regression is usually performed in a step-wise manner (Sect. 6.6.4, Ref. 29). With this approach, the most influential variable is added to the model first (producing a model of the form in Eq. (2.2) with one independent variable); then the next most influential variable is added to the model (producing a model of the form in Eq. (2.2) with two independent variables); and the process is continued in this manner until no more influential variables can be identified. Variable importance is then indicated by the order in which variables entered the regression model, the changes in R^2 values as successive variables entered the regression model, and the standardized regression coefficients for the variables in the final regression model. However, it is important to recognize that standardized regression coefficients can produce very misleading indications of variable importance when highly correlated variables are included in the regression model (Sect. 6.6.7, Ref. 29).

An important special situation exists when the values for the x_j used in construction of the regression model in Eq. (2.2) (i.e., in the sample in Eq. (1.5)) are independent. Technically, this is equivalent to $\mathbf{X}^T \mathbf{X}$ being a diagonal matrix, where

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1,nX} \\ 1 & x_{21} & x_{22} & \cdots & x_{2,nX} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{nS,1} & x_{nS,2} & & x_{nS,nX} \end{bmatrix}. \quad (2.6)$$

In this situation,

$$R^2 = R_1^2 + R_2^2 + \cdots + R_{nX}^2, \quad (2.7)$$

where R_j^2 is the R^2 value that results from regressing y on only x_j (p. 99, Ref. 111). Thus, R_j^2 is the contribution of x_j to R^2 when the sampled inputs are independent (i.e., when the design matrix \mathbf{X} is orthogonal). As a result, the incremental R^2 values in a stepwise regression are equal to the contributions of the individual independent variables to the total R^2 value for the regression.

There are many favorable properties of linear regression such as computational speed and interpretability. Hypothesis testing for input variable importance can be performed with ease. When the surface to be approximated is nearly linear in the inputs (i.e., the x_j), there is no better technique. However, in situations where the underlying relationship (i.e., the model in Eq. (1.7)) is far from linear, linear regression will produce a very poor approximation (Fig. 1). As a result, a number of alternatives to linear regression have been developed, including rank regression (Sect. 2.2), quadratic regression (Sect. 2.3), and nonlinear regression (Sect. 2.4).

The results in Fig. 1 come from an uncertainty and sensitivity analysis carried out for a two phase fluid flow model. This analysis will be described in greater detail in Sect. 5.2 where it is used to illustrate multiple predictor smoothing methods. This analysis involved 31 uncertain variables (i.e., $nX = 31$ in Eq. (1.3)). The regression line in Fig. 1 involves only one uncertain variable. Owing to the extreme nonlinearity of the relationships involved, the inclusion of additional uncertain variables in the regression model fails to produce a satisfactory representation. For example, use of an α -value cutoff of 0.02 for entry of a variable into the regression model produces a model with five variables and an R^2 value of only 0.27. Additional discussion is given in Sect. 5.2.6.

2.2 Rank Regression

The results obtained with linear regression can often be improved with suitable transformations of the independent (i.e., y) and dependent (i.e., x_1, x_2, \dots, x_{nX}) variables. For example, logarithmic or square root transformations may make the underlying relationships more linear and hence more amenable to analysis by linear regression. The identification of effective transformations is often subjective and thus difficult to automate. As a result, the effective use of transformations in a large sensitivity study can be difficult due to the large number of independent and dependent variables under consideration.

One broadly applicable transformation is the rank transformation, which is effective when the relationships between independent and dependent variables are monotonic (Ref. 115; Sect. 6.6.6, Ref. 29). The use of the rank transformation in conjunction with linear regression is straightforward. The smallest value of a variable is given a rank of 1; the next largest value is given a rank of 2; and so on up to the largest value which is given a rank of nS , where nS is the sample size. Equal variable values are assigned the average of what their ranks would have been. Then, the usual regression procedures are carried out with the original variable values replaced by their ranks (Ref. 115; Sect. 6.6.6, Ref. 29).

The rank transformation converts monotonic relationships into linear relationships (Fig. 2). As a result, a linear regression in this situation with rank transformed data (i.e., a rank regression) provides a better approximation to the underlying relationships than would be obtained with a linear regression on the original (i.e., raw) data. Rank regressions have been successfully used in a large number of sensitivity analyses (e.g., Refs. 116-118). However, rank regressions cannot significantly improve the quality of a regression analysis when the underlying relations are nonlinear and nonmonotonic (Fig. 3).

2.3 Quadratic Regression

Quadratic regression is used as a designator for linear regression that includes individual variables (i.e., the x_j), variable squares (i.e., x_j^2), and multiplicative interaction terms (i.e., $x_j x_k$). Formally, quadratic regression is predicated on a model of the form

$$y = \alpha + \sum_{j=1}^{nX} (\beta_j x_j + \beta_{jj} x_j^2) + \sum_{j=1}^{nX} \sum_{k=j+1}^{nX} \beta_{jk} x_j x_k + \varepsilon. \quad (2.8)$$

More generally, polynomial regression models that involve additional powers of the x_j and more complex multiplicative interaction terms are also possible.

Quadratic regression removes the assumption that the effects of the individual x_j are completely additive but still cannot model completely general interactions. Further, quadratic regression has difficulty representing functions with asymptotes and other complex behavior. Still, quadratic regression has been used with considerable success in industrial applications for many years.^{20, 119}

A quadratic regression model (Fig. 4) shows significant improvement over the results previously shown for the application of linear and rank regression to a nonlinear and nonmonotonic relationship (Figs. 1, 3).

2.4 Nonlinear Regression

Nonlinear regression involves estimating the coefficients in a nonlinear relationship between y and the independent variables under consideration.¹²⁰ In particular, the regression models introduced in Sects. 2.1 – 2.3 are re-

ferred to as linear models because y is expressed as a linear combination of the variables in the regression model. In contrast, nonlinear regression involves estimating the coefficients $\beta_j, j = 0, 1, 2, \dots$, in a hypothesized relationship such as

$$y = \beta_0 + \beta_1 x + \beta_2 \exp(\beta_3 x) + \varepsilon, \quad (2.9)$$

where the relationships between y and at least some of the independent variables are nonlinear in the sense that y is not represented as a linear combination of these variables. Once the candidate form for the nonlinear regression model is decided on (e.g., the relationship in Eq. (2.9)), the β_j 's can be estimated with techniques based on least squares, which is the maximum likelihood estimate when the ε 's are normally distributed.

A major drawback to nonlinear regression is the requirement to decide on the form of the nonlinear regression model before the regression process can be initiated. This can be a particularly daunting challenge in a sensitivity analysis where several hundred different dependent variables (i.e., y 's) may be under consideration with each dependent variable potentially requiring the formulation of a different nonlinear regression model. Further, model fitting, hypothesis testing, and interpreting of results is more difficult than is the case for linear regression. For the preceding reasons, nonlinear regression models are not considered in this study. The nonparametric regression approaches introduced in the next section (Sect. 3) have advantages over nonlinear regression in that they can incorporate nonlinear relationships without the need to provide *a priori* specifications of model form.

3. Nonparametric Regression

Linear regression analysis has many desirable properties. When the underlying relationships are close to linear, no better technique is available. However, when nonlinear relationships are present, linear regression analysis can give misleading results and possibly no results at all. This potential failing provides the motivation for nonparametric regression.

Nonparametric regression, which is often called smoothing, is a form of surface approximation that is based on an assumed relationship of the form

$$y = f(\mathbf{x}) + \varepsilon, \quad \mathbf{x} = [x_1, x_2, \dots, x_{nX}], \quad (3.1)$$

where $E(\varepsilon) = 0$ and, as a result, $E(y|\mathbf{x}) = f(\mathbf{x})$. Usually, very few restrictions or assumptions are made about the properties of f . In particular, f is not assumed to take a particular parametric form such as a multivariate polynomial involving the elements of \mathbf{x} . Sometimes f is assumed to be “smooth” in the sense that certain continuity restrictions are imposed on f and possibly its derivatives.

To facilitate the introduction of the concept of smoothing, \mathbf{x} is initially assumed to be univariate and smoothing is discussed in this context (Sect. 3.1); that is, the relation in Eq. (3.1) is assumed to be of the form $y = f(x) + \varepsilon$. Such univariate smoothing is often referred to as scatterplot smoothing. Next, the concept of degrees of freedom in association with smoothing is discussed (Sect. 3.2). Then, multivariate smoothing is described for relationships of the form in Eq. (3.1); that is, for the case where \mathbf{x} is a vector rather than a scalar (Sect. 3.3). Finally, hypothesis testing for variable importance in nonparametric regression is discussed (Sect. 3.4).

3.1 Univariate Scatterplot Smoothers

The following provides a brief overview of scatterplot smoothing. More information is available in several references.¹²¹⁻¹²⁴ As previously indicated, scatterplot smoothers are used when there is one independent variable (i.e., x) and one dependent variable (i.e., y). Specifically, a data set of the form (x_i, y_i) , $i = 1, 2, \dots, nS$, is under consideration throughout this section. As suggested by the name, scatterplot smoothing involves fitting a curve to the data represented in a scatterplot. There are many ways to construct (i.e., fit) such a curve. The most familiar approach to such construction is simple linear regression (Sect. 2.1), although this approach is hardly nonparametric. In contrast to the parametric character of linear regression, the following nonparametric approaches to scatterplot smoothing are introduced: running means (Sect. 3.1.1), locally weighted means (Sect. 3.1.2), locally weighted regression (Sect. 3.1.3), and smoothing splines (Sect. 3.1.4).

3.1.1 Running Means

With running means, or possibly running medians, the predicted (i.e., estimated) value $\hat{y}(x_0)$ of $y(x_0)$ at a value x_0 of x is given by the mean (or median) of the y_i 's associated with x_i 's close to x_0 . Typically, a fixed number r of values for x_i is selected for use. Then, $\hat{y}(x_0)$ is defined on the basis of the r values for x_i that are closest to x_0 . For running means, this leads to the following approximation for an arbitrary value of x :

$$\hat{y}(x) = \hat{f}(x) = \frac{1}{r} \sum_{i=1}^{nS} I_{[0, d_r(x)]}(|x_i - x|) y_i, \quad (3.2)$$

where $d_r(x)$ denotes the distance along the x -axis to the r^{th} nearest neighbor of x (i.e., r values of x_i satisfy $|x_i - x| \leq d_r(x)$) and

$$I_{[0, d_r(x)]}(|x_i - x|) = \begin{cases} 1 & \text{if } 0 \leq |x_i - x| \leq d_r(x) \\ 0 & \text{otherwise.} \end{cases} \quad (3.3)$$

A minor modification is required if multiple observations satisfy $|x_i - x| = d_r(x)$ (e.g., increase the value for r to incorporate these values or leave r fixed and average the corresponding y_i values). An analogous relationship holds for running medians except that medians over the y_i 's associated with the r x_i 's closest to x are calculated rather than means.

Although running means or medians are appealing because of their simplicity, they tend to produce a very wiggly function $\hat{f}(x)$. Specifically, as the values for x move along the x -axis, the sets of x_i 's in use change, with these changes resulting in discontinuities in $\hat{f}(x)$. This behavior is illustrated in Fig. 5 for running means with the previously introduced two-phase flow data and $r = 20$.

3.1.2 Locally Weighted Means: Kernel Smoothers

Smoothing based on locally weighted means is employed to keep the intuitively appealing idea of a moving average while, concurrently, producing less small-scale erratic behavior in $\hat{f}(x)$. Specifically, locally weighted averaging with a kernel function $k(z; h)$ produces the approximation

$$\hat{f}(x) = \frac{\sum_{i=1}^{nS} k(x_i - x; h) y_i}{\sum_{i=1}^{nS} k(x_i - x; h)}. \quad (3.4)$$

The role of $k(z; h)$ is to place more weight on the y_i 's associated with x_i 's close to x and less weight on y_i 's associated with x_i 's farther away from x . The kernel function $k(z; h)$ is usually chosen to have a maximum at $z = 0$ and to decrease monotonically to zero as $|z|$ increases. If $k(z; h)$ is a continuous function of z , then $\hat{f}(x)$ will be a continuous function of x . The bandwidth h , also known as the smoothing parameter, determines the amount of smoothing to be done to the data. Larger values of h result in more smoothing and smaller values of h result in greater fidelity to the data. A commonly used kernel function is

$$k(z; h) = \left(1/h\sqrt{2\pi}\right) \exp\left(-z^2/2h^2\right), \quad (3.5)$$

which corresponds to the normal density function with $\mu = 0$ and $\sigma = h$. Other viable choices for $k(z; h)$ also exist (e.g., see Sect. 2.6, Ref. 121). As discussed in more detail in Sect. 3.2, there is no universally accepted approach to determining the best value for h for a given kernel function and data set. However, it is widely accepted that the choice of the bandwidth h has more effect on the smoothing process than the choice of the kernel function (p. 19, Ref. 121).

The use of a kernel smoother with the kernel function in Eq. (3.5) and a bandwidth of $h = 0.6$ is illustrated in Fig. 6. Comparison of Figs. 5 and 6 illustrates the smoother form for $\hat{f}(x)$ produced by the use of locally weighted means than is the case for running means.

Kernel smoothers are “linear smoothers” in the sense that

$$\hat{\mathbf{y}} = [\hat{f}(x_1), \hat{f}(x_2), \dots, \hat{f}(x_{nS})]^T \quad (3.6)$$

can be represented by

$$\hat{\mathbf{y}} = \mathbf{S}\mathbf{y}, \quad (3.7)$$

where $\mathbf{y} = [y_1, y_2, \dots, y_{nS}]^T$ and the i^{th} row of the matrix \mathbf{S} contains the kernel weights in the linear combination in Eq. (3.4). Specifically, the value in row i and column j of \mathbf{S} is

$$s_{ij} = k(x_j - x_i; h) / \sum_{k=1}^{nS} k(x_k - x_i; h), \quad (3.8)$$

which is the weight on the j^{th} observation for prediction at x_i .

Edge effects are a potential drawback with locally weighted means. Such effects can be manifested near the largest and smallest observed values for x and result because of the unequal numbers of observations to the left and right of such values. Specifically, there are few observations to the left of small values for x_i and few observations to the right of large values for x_i . This imbalance in the number of observations can result in an overemphasis in the averaging process of observations on one side of such values and thus distort $\hat{f}(x)$ for values of x near the upper or lower ends of the range of values for the x_i . This effect can be seen for the smaller values of $x = BHPRM$ in Fig. 6, where the value for $\hat{f}(x)$ determined with locally weighted means appears to fall below the overall trend of the data.

3.1.3 Locally Weighted Regression

An approach similar to the kernel smoother (Sect. 3.1.2) that reduces the problem of edge effects involves the use of a locally weighted regression line.¹²⁵ With locally weighted regression,

$$\hat{f}(x) = \hat{\alpha}(x) + \hat{\beta}(x)x, \quad (3.9)$$

where $\hat{\alpha}(x)$ and $\hat{\beta}(x)$ are estimated for individual values of x . In particular and for a specific value of x , the quantities $\hat{\alpha}(x)$ and $\hat{\beta}(x)$ are defined to be the values for α and β that minimize the sum

$$\sum_{i=1}^{nS} (\alpha + \beta x_i - y_i)^2 k(x - x_i; h), \quad (3.10)$$

where $k(x - x_i; h)$ is an appropriately defined kernel function. The indicated minimization of α and β is straightforward with an appropriate matrix formulation of the problem (p. 84, Ref. 124).

Locally weighted regression is actually equivalent to the determination of a locally weighted mean (Sect. 3.1.2) with a complicated kernel function that derives from the estimation of $\hat{\alpha}(x)$ and $\hat{\beta}(x)$. This kernel function will not be given here but can be found elsewhere (p. 241, Ref. 126). Thus, locally weighted regression is also a linear smoother as it can be put in the form $\hat{\mathbf{y}} = \mathbf{S}\mathbf{y}$ indicated in Eq. (3.7).

Both kernel smoothing of the mean and locally weighted regression have a problem when data are sparse in a particular region. In this situation, there are few points close to some x values to use in the averaging process and, depending on the kernel in use, $\hat{f}(x)$ may not even be defined for such x values. Cleveland recognized this problem and mitigated its effects by incorporating a nearest neighbors approach with the locally weighted regression line.¹²⁵ This procedure is often referred to by the designator LOESS, which is short for *local regression* and was chosen in allusion to the fact that LOESS is a deposit of fine clay or silt along a river valley and is thus a surface of sorts (p. 314, Ref. 127).

With LOESS, the kernel function is modified to take into account the distance $d_r(x)$ to the r^{th} nearest neighbor of a point x . Specifically, Cleveland¹²⁵ proposed that $\hat{\alpha}(x)$ and $\hat{\beta}(x)$ should be estimated by minimizing the expression in Eq. (3.10) with the kernel function $k(z; h)$ defined by

$$k(z; h) = \left[1 - (|z|/h)^3 \right]^3 I_{[0, h)}(|z|), \quad (3.11)$$

where $I_{[0, h)}(|z|)$ is defined analogously to the expression in Eq. (3.3) (i.e., $I_{[0, h)}(|z|) = 1$ if $0 \leq |z| < h$ and 0 otherwise) and h corresponds to $d_r(x)$. With this formulation, $\hat{\alpha}(x)$ and $\hat{\beta}(x)$ are defined to be the values for α and β that minimize the expression

$$\sum_{i=1}^{nS} (\alpha + \beta x_i - y_i)^2 \left\{ 1 - \left[|x - x_i| / d_r(x) \right]^3 \right\}^3 I_{[0, d_r(x))}(|x - x_i|). \quad (3.12)$$

The use of $h = d_r(x)$ in the definition of $k(z; h)$ allows the bandwidth to vary along the x -axis. This assures that $r - 1$ of the nS observations will have nonzero weights when computing the local regression line $\hat{f}(x)$ for each x regardless of how sparse the data is. If several points are tied for being the r^{th} nearest neighbor to x , then there will actually be less than $r - 1$ points with nonzero weight for this special case. An analysis employing LOESS is often described by its span, which is the ratio r/nS . Intuitively, the span is the ratio of the number of observations with nonzero weight used in the estimation of $\hat{\alpha}(x)$ and $\hat{\beta}(x)$ to the total number of observations although this is not quite correct as only $r - 1$ observations typically have nonzero weight.

The improvement in the estimate of $\hat{f}(x)$ with LOESS over the estimate obtained with locally weighted means can be seen by comparing the results in Figs. 6 and 7. In particular, the estimate for $\hat{f}(x)$ in Fig. 7 is obtained from LOESS with $r = 60$ and a corresponding span of 0.20. This estimate tracks the data near the ends of the range for $x = BHPRM$ more faithfully than is the case for $\hat{f}(x)$ in Fig. 6 obtained with locally weighted means. This is particularly evident for the smaller values of x . Due to its good performance, LOESS has become one of the most popular scatterplot smoothers.

3.1.4 Smoothing Splines

Another popular scatterplot smoother is the cubic smoothing spline. A cubic smoothing spline is a function \hat{f} that minimizes the penalized residual sum of squares

$$\sum_{i=1}^{nS} [y_i - \hat{f}(x_i)]^2 + \lambda \int_a^b [d^2 \hat{f}(x)/dx^2]^2 dx \quad (3.13)$$

over all continuously differentiable functions f , where $a \leq x_{(1)} = \min\{x_i: 1 \leq i \leq nS\}$, $\max\{x_i: 1 \leq i \leq nS\} = x_{(nS)} \leq b$, and λ is a constant (Sect. 2.10, Ref. 121). The first term in the preceding expression is the residual sum of squares and measures fidelity to the data; the second term constitutes a penalty for \hat{f} having too much curvature.

There is a unique, explicit solution to the minimization problem associated with Eq. (3.13). This solution is a natural cubic polynomial spline with knots (i.e., locations of change in the structure of the spline) at the observed values for x (Sect. 2.10, Ref. 121). A cubic polynomial spline is a function that is a cubic polynomial on any interval defined by adjacent knots, has two continuous derivatives, and has a third derivative that is a step function with jumps at the knots. A natural cubic spline is a cubic spline that is restricted to be linear on $(-\infty, x_{(1)})$ and $(x_{(nS)}, \infty)$.

The quantity λ in Eq. (3.13) plays the role of a smoothing parameter. As with the smoothing parameters associated with the previously introduced methods, the appropriate value to use for λ is not intuitively apparent. Typically, the equivalent degrees of freedom (df) described in the next section (Sect. 3.2) is used to determine the value for λ for smoothing splines. Fig. 8 shows a cubic smoothing spline involving *BHPRM* with $df = 8$. As comparison of Figs. 7 and 8 shows, the behavior of this cubic smoothing spline is similar to that of LOESS.

3.2 Equivalent Degrees of Freedom and Smoothing Parameters

Automated methods of selecting smoothing parameters for the techniques presented in Section 3.1 are now discussed. To do this, the related topic of degrees of freedom is introduced. In linear model theory, the degrees of freedom df of a model is defined to be the number of linearly independent columns in the design matrix \mathbf{X} defined in Eq. (2.6). This is the same as the number of parameters included in the associated linear model. An equivalent definition is

$$df = tr(\mathbf{H}), \quad (3.14)$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$, often called the hat matrix, is the perpendicular projection matrix that projects a vector onto the \mathbf{X} space (i.e., the space spanned by the vectors corresponding to the columns in \mathbf{X}) and $tr(\mathbf{H})$ denotes the trace of \mathbf{H} (i.e., the sum of the diagonal elements of \mathbf{H}). Predicted values for the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$ are obtained from the relationship $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$. More on the projection matrix \mathbf{H} can be found elsewhere (p. 68, Ref. 128; p. 393, Ref. 129).

The nonparametric techniques discussed in Sects. 3.1.1 – 3.1.3 can be put in the form $\hat{\mathbf{y}} = \mathbf{S}\mathbf{y}$ indicated in Eq. (3.7). Such techniques are said to be linear. For convenience, \mathbf{S} is referred to as the smoother matrix. The symbol \mathbf{S} is used to denote the smoother matrix to distinguish it from \mathbf{H} since \mathbf{S} is not, in general, a perpendicular projection onto the \mathbf{X} space. A natural generalization of the concept of degrees of freedom is to define the degrees of freedom associated with a smoother matrix \mathbf{S} to be

$$df = tr(\mathbf{S}), \quad (3.15)$$

where $tr(\mathbf{S})$ denotes the trace of \mathbf{S} . In turn, the degrees of freedom for error $dferr$ can then be defined by

$$dferr = nS - tr(\mathbf{S}) \quad (3.16)$$

in analogy to the corresponding definition

$$dferr = nS - tr(\mathbf{H}) \quad (3.17)$$

for linear models.

The preceding definitions make some intuitive sense if the two extreme prediction cases, simple averaging and interpolation, are considered. For a simple average, the diagonal elements of \mathbf{S} are given by $s_{ii} = 1/nS$. As a result, $tr(\mathbf{S}) = 1$ or, equivalently, one degree of freedom (i.e., $df = 1$) is being used to estimate the overall mean value. In the interpolation case, $s_{ii} = 1$ and the other weights in a row must be 0 so that the predicted value is given by $\hat{f}(x_i) = y_i$. In this case, each observation has its own value and $tr(\mathbf{S}) = nS$, which implies a model with nS degrees of freedom (i.e., $df = nS$). Most models fall somewhere in between these two extremes. Additional discussion of degrees of freedom in the context of nonparametric regression is available elsewhere (pp. 52 – 55, Ref. 121).

Degrees of freedom will be used for inference later in this presentation. However, degrees of freedom can also be used to obtain some insight with respect to appropriate values to use for smoothing parameters. For a particular kernel, a desired value of df for the smoother matrix can be specified, and then the value of the smoothing parameter that produces this value can be determined. This still leaves open the question of what is an appropriate value for df . The approaches below offer a better guide to smoothing parameter selection.

A widely used automatic selection procedure for smoothing parameters is the cross validation (CV) approach. With this approach, the jackknifed (or leave one out) residuals are obtained by fitting the model without the i^{th} observation and then predicting y_i . The deleted residual is then

$$r_{(i)} = y_i - \hat{y}_{(i)}, \quad (3.18)$$

where $\hat{y}_{(i)}$ is the jackknifed (i.e., predicted) value for y_i obtained with y_i omitted from the prediction process, and the predicted residual sum of squares (PRESS)¹³⁰ is given by

$$PRS = \sum_{i=1}^{nS} r_{(i)}^2 = \sum_{i=1}^{nS} [y_i - \hat{y}_{(i)}]^2. \quad (3.19)$$

The PRESS value PRS is then used in the selection of the smoothing parameter. In particular, different values for the smoothing parameter result in different values for PRS . The preferred smoothing parameter value is the value that minimizes PRS .

For representations of the form $\hat{\mathbf{y}} = \mathbf{S}\mathbf{y}$, it is not necessary to fit the model multiple times to obtain the deleted residuals. Instead, all the deleted residuals can be obtained from the usual residuals

$$r_i = y_i - \hat{y}_i \quad (3.20)$$

and the leverage values s_{ii} , which are the diagonal elements of \mathbf{S} . In particular, the deleted residuals are given by

$$r_{(i)} = r_i / (1 - s_{ii}) = (y_i - \hat{y}_i) / (1 - s_{ii}) \quad (3.21)$$

(see p. 47, Ref. 121). This makes cross validation easy to apply for linear smoothing provided \mathbf{S} is relatively easy to calculate.

In practice, the preceding cross validation criterion tends to result in the selection of smoothing parameters that undersmooth. To correct for this, a generalized cross validation criterion has been suggested (p. 49, Ref. 121). This generalized criterion employs an adjusted PRESS value given by

$$PRS_A = \sum_{i=1}^{nS} \left[\frac{r_i}{1 - \text{tr}(\mathbf{S})/nS} \right]^2 = \sum_{i=1}^{nS} \left[\frac{y_i - \hat{y}_i}{1 - \text{tr}(\mathbf{S})/nS} \right]^2 \quad (3.22)$$

in the determination of the smoothing parameter. Given that

$$\text{tr}(\mathbf{S}) = \sum_{i=1}^{nS} s_{ii}, \quad (3.23)$$

each deleted residual $r_{(i)}$ is in essence being calculated with an average leverage value given by

$$\bar{s} = \text{tr}(\mathbf{S})/nS = \sum_{i=1}^{nS} s_{ii}/nS. \quad (3.24)$$

This approach puts less emphasis on observations with high leverage values. Another way to write PRS_A is

$$PRS_A = \left[\frac{1}{1 - df/nS} \right]^2 \sum_{i=1}^{nS} r_i^2, \quad (3.25)$$

which shows that PRS_A can be viewed as the error sum of squares penalized by the degrees of freedom associated with the model used in smoothing the data. With this criterion, the preferred smoothing parameter value is the value that minimizes PRS_A .

3.3 Multivariate Smoothers

More general relationships of the form $y = f(\mathbf{x})$ indicated in Eq. (1.7) are now considered. Further, a mapping $y_i = f(\mathbf{x}_i)$, $i = 1, 2, \dots, nS$, from analysis inputs to analysis results as shown in Eq. (1.8) is assumed to be available for analysis. In this framework, approximations $\hat{f}(\mathbf{x})$ to a relationship of the form

$$E(y|\mathbf{x}) = f(\mathbf{x}) = f(x_1, x_2, \dots, x_{nX}) \quad (3.26)$$

are sought. The kernel methods described for the univariate case in Sect. 3.1 have immediate and straightforward generalizations to this multivariate context. These generalizations are often referred to as multiple predictor techniques. In particular, the following multiple predictor techniques are considered in this section: locally weighted regression (Sect. 3.3.1), additive models (Sect. 3.3.2), projection pursuit regression (Sect. 3.3.3), and recursive partitioning regression (Sect. 3.3.4).

3.3.1 Locally Weighted Regression: LOESS

The LOESS technique in multiple dimensions is analogous to the same technique in one dimension (Sect. 3.1.3). In particular, the relationship between y and \mathbf{x} is assumed to be of the form

$$y = f(\mathbf{x}) = \alpha(\mathbf{x}) + \boldsymbol{\beta}(\mathbf{x})\mathbf{x} + \varepsilon, \quad (3.27)$$

where $\boldsymbol{\beta}(\mathbf{x}) = [\beta_1(\mathbf{x}), \beta_2(\mathbf{x}), \dots, \beta_{nX}(\mathbf{x})]$, $\mathbf{x} = [x_1, x_2, \dots, x_{nX}]^T$, and $E(\varepsilon) = 0$. In turn, an approximate relationship of the form

$$\hat{y} = \hat{f}(\mathbf{x}) = \hat{\alpha}(\mathbf{x}) + \hat{\boldsymbol{\beta}}(\mathbf{x})\mathbf{x} \quad (3.28)$$

is sought with LOESS, with the corresponding one dimensional special case appearing in Eq. (3.9).

The quantities $\hat{\alpha}(\mathbf{x})$ and $\hat{\boldsymbol{\beta}}(\mathbf{x})$ for a given value of \mathbf{x} are defined to be the values for α and $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_{nX}]$ that minimize the sum

$$\sum_{i=1}^{nS} (\alpha + \boldsymbol{\beta} \mathbf{x}_i - y_i)^2 \left[1 - \left(\frac{\|\mathbf{x} - \mathbf{x}_i\|}{d_r(\mathbf{x})} \right)^3 \right]^3 I_{[0, d_r(\mathbf{x})]}(\|\mathbf{x} - \mathbf{x}_i\|), \quad (3.29)$$

where (i) $d_r(\mathbf{x})$ is the distance to the r^{th} nearest neighbor of \mathbf{x} in nX -dimensional Euclidean space, (ii) $I_{[0, d_r(\mathbf{x})]}(\|\mathbf{x} - \mathbf{x}_i\|)$ is defined analogously to $I_{[0, d_r(x)]}(|x - x_i|)$ in Eq. (3.12), and (iii) the individual independent variables (i.e., x_1, x_2, \dots, x_{nX}) are normalized to mean zero and standard deviation one so that the value for the norm $\|\cdot\|$ is not dominated by the units used for these variables. The determination of α and $\boldsymbol{\beta}$ is straightforward with the use of appropriate matrix techniques (p. 139, Ref. 122). Except for use of the norm $\|\cdot\|$ instead of absolute value $|\cdot|$, the expression in Eq. (3.29) with LOESS for multidimensional \mathbf{x} is the same as the expression in Eq. (3.12) for the one-dimensional case.

The determination of $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}(\mathbf{x})$ provides an estimate of \hat{y} for one value of \mathbf{x} as indicated in Eq. (3.28). Estimates of y for additional values of \mathbf{x} require the solution of an additional minimization problem for each \mathbf{x} . This may seem computationally demanding but LOESS is actually quite fast computationally even with multiple independent variables.

The obvious benefit to using LOESS in multiple dimensions is that it can capture nonlinear behavior that a typical parametric model cannot. A more subtle advantage is the capability to capture very general interactions between input variables. The indicated capabilities derive from the property that LOESS is inherently local in its approximations to the relationship $y = f(\mathbf{x})$. For example, a LOESS surface fitted to two variables is shown in Fig. 9. The actual functional relationship is

$$y = f(x_1, x_2) = (1/2\pi) \exp\left\{-\left[\frac{(x_1 - 5)^2 + (x_2 - 5)^2}{2}\right]\right\}, \quad (3.30)$$

which corresponds to the density function for a bivariate normal distribution. The surface in Fig. 9 was constructed with LOESS and a random sample of size $nS = 100$ from $\mathbf{x} = [x_1, x_2]$ with x_1 and x_2 uniform on $[0, 10]$. In this example, LOESS captures the nonlinear interaction between x_1 and x_2 in the determination of y .

The LOESS technique in multiple dimensions is also a linear smoother in the sense that it can be expressed in the form shown in Eq. (3.7). The actual form of the kernel function is a generalization of the univariate case given in Schimek (p.241, Ref. 126).

A drawback with LOESS and other local averaging techniques in higher dimensions is that the closest observed values \mathbf{x}_i to the value \mathbf{x} under consideration are not necessarily local (i.e., nearby) along the axes for the individual variables $x_j, j = 1, 2, \dots, nX$, contained in \mathbf{x} . This is sometimes referred to as the curse of dimensionality. To illustrate this, first consider one independent variable. To include 30% of the data in a local average, it is necessary to span approximately 30% of the corresponding axis if the variable values are approximately uniformly distributed.

With the same distributional assumption and two independent variables, including 30% of the data now requires spanning 55% of the range of each of the variables. This requirement results because the joint range of the two variables is now a rectangle and covering 30% of this rectangle requires covering 55% of the range of each of the two variables (i.e., $(0.55)^2 \cong 0.30$). As the number of independent variables increases, the problem becomes worse. With five independent variables, use of 30% of the data requires spanning 79% of the range of each of the individual variables. This hardly constitutes a local average anymore. The span (i.e., percent coverage) can be made smaller but then there is a danger of undersmoothing unless the number of observations is substantially increased.

The LOESS procedure will work in higher dimensions and actually works quite well for $nX \leq 3$. For $nX > 3$, however, LOESS starts to be affected by the curse of dimensionality. As will be illustrated later, this can cause LOESS to miss the effects of important variables in the estimation of f (Sect. 5).

Several procedures have been developed in an attempt to overcome the dimensionality problem. These procedures implement one or more of the following strategies as discussed in subsequent sections: additive modeling (Sect. 3.3.2), dimension reduction (Sect. 3.3.3), and recursive partitioning (Sect. 3.3.4).

3.3.2 Additive Models

For additive modeling, the function $f(\mathbf{x})$ in Eq. (3.27) is assumed to have the form

$$f(\mathbf{x}) = \sum_{j=1}^{nX} f_j(x_j), \quad (3.31)$$

where the f_j are arbitrary functions that will be determined as part of the analysis process. This is analogous to multiple linear regression where the effects of the independent variables are additive. The difference is that $y = f(\mathbf{x})$ is not assumed to be a linear function of the x_j . This representation is not completely general as it does not allow for interactions between the independent variables. However, nothing prevents the inclusion of multiplicative interactions $x_r x_s$ as in linear regression.

Additive models are usually constructed with a method known as backfitting suggested by Friedman and Stutzel.¹³¹ The algorithm that is used in the software packages R and S-Plus to implement this method is described in Chambers and Hastie (p. 300, Ref. 127). The indicated algorithm is more efficient than the approach that is described below. However, the described approach provides a more intuitive introduction to the ideas involved in additive model construction.

The observed values for y are assumed to be of the form

$$y_i = f(\mathbf{x}_i) + \varepsilon_i = \sum_{j=1}^{nX} f_j(x_{ij}) + \varepsilon_i. \quad (3.32)$$

Given initial estimates $\hat{f}_2, \hat{f}_3, \dots, \hat{f}_{nX}$ for f_2, f_3, \dots, f_{nX} , (e.g., $\hat{f}_j(x_j) = b_j x_j$ for $j = 2, 3, \dots, nX$, where the b_j are coefficients from a regression model of the form indicated in Eq. (2.2)), an estimate \hat{f}_1 for f_1 can be obtained through use of the relationship

$$y_i - \sum_{j=2}^{nX} \hat{f}_j(x_{ij}) \cong f_1(x_{i1}) + \varepsilon_i \quad (3.33)$$

for $i = 1, 2, \dots, nS$. In particular, one of the scatterplot smoothers introduced in Sect. 3.1 can be used to smooth the partial residuals on the left hand side of Eq. (3.33) across x_1 . This produces an estimate \hat{f}_1 for f_1 defined across the range of values for x_1 . Given this estimate for f_1 , the estimate \hat{f}_2 for f_2 can be refined in the same manner across the range of values for x_2 with $\hat{f}_1, \hat{f}_3, \hat{f}_4, \dots, \hat{f}_{nX}$. This procedure then continues and repetitively cycles through the variables. The cycling continues until convergence is achieved.

The result is \hat{f}_j defined over the range of x_j for $j = 1, 2, \dots, nX$. In turn, $y = f(\mathbf{x})$ can be estimated for arbitrary values of $\mathbf{x} = [x_1, x_2, \dots, x_{nX}]$ by

$$y \cong \sum_{j=1}^{nX} \hat{f}_j(x_j) \quad (3.34)$$

Additional detail is available elsewhere (pp. 90 – 91, Ref. 121; pp. 300 – 302, Ref. 127).

Additive models can be used to develop representations for complex nonlinear behavior as indicated in Fig. 10 by the approximation to

$$y = f(x_1, x_2) = \sin(x_1) + (x_2 - 5)^2 \quad (3.35)$$

obtained from a random sample of size $nS = 100$ from $\mathbf{x} = [x_1, x_2]$ with x_1 and x_2 uniform on $[0, 10]$. Additive models also work well in higher dimensions with a large number of independent variables as will be illustrated in Sect. 5. However, successful construction of an additive model is dependent on the actual relationship between y and \mathbf{x} involving limited interactions between the elements of \mathbf{x} .

The procedure indicated in this section to construct an approximation to the function $f(\mathbf{x})$ in Eq. (3.31) is a linear smoother provided a linear scatterplot smoother is used in the backfitting algorithm in the sense that this procedure can be formally represented in the form shown in Eq. (3.7). The smoother matrix \mathbf{S} in Eq. (3.7) is difficult to compute in a closed form as the overall analysis involves an iterative process. An approximation to $tr(\mathbf{S})$, which corresponds to the number of degrees of freedom associated with the procedure, is given by

$$tr(\mathbf{S}) \cong \sum_{j=1}^{nX} df_j, \quad (3.36)$$

where df_j is the degrees of freedom used in the scatterplot smoother for x_j in the backfitting algorithm (p. 129, Ref. 121).

3.3.3 Projection Pursuit Regression

Projection pursuit regression involves both dimension reduction and additive modeling and is based on the assumption that the function $f(\mathbf{x})$ in Eq. (3.27) has the form

$$f(\mathbf{x}) = \sum_{s=1}^{nD} g_s(\boldsymbol{\alpha}_s \mathbf{x}), \quad (3.37)$$

where $\boldsymbol{\alpha}_s = [\alpha_{1s}, \alpha_{2s}, \dots, \alpha_{nXs}]$, $\boldsymbol{\alpha}_s$ and $\boldsymbol{\alpha}_t$ are orthogonal for $s \neq t$, $\mathbf{x} = [x_1, x_2, \dots, x_{nX}]^T$, $\boldsymbol{\alpha}_s \mathbf{x}$ corresponds to a linear combination of the elements of \mathbf{x} , and g_s is an arbitrary function. Values for g_s , $\boldsymbol{\alpha}_s$ and nD are determined as part of the analysis procedure. The expression in Eq. (3.37) is an additive model with the quantities $\boldsymbol{\alpha}_s \mathbf{x}$ replacing the elements x_j of \mathbf{x} as the independent variables. Further, this expression involves a reduction in dimension as nD is usually smaller than nX .

The representation for f in Eq. (3.37) allows for interactions between variables, which is not the case for the additive representation in Eq. (3.31). To see this, consider the example in which $\mathbf{x} = [x_1, x_2]^T$, $\boldsymbol{\alpha}_1 = [1, 1]$ and $g_1(u) = u^2$. The result is

$$g_1(\boldsymbol{\alpha}_1 \mathbf{x}) = (x_1 + x_2)^2 = x_1^2 + 2x_1x_2 + x_2^2, \quad (3.38)$$

which involves the interaction term x_1x_2 .

The entities $\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\alpha}}_2, \dots, \hat{\boldsymbol{\alpha}}_{nD}$ and $\hat{g}_1, \hat{g}_2, \dots, \hat{g}_{nD}$ are estimated as part of the construction process. This is accomplished by first estimating $\boldsymbol{\alpha}_1$ and g_1 . Specifically, $\hat{\boldsymbol{\alpha}}_1$ and \hat{g}_1 are defined to be the values for $\boldsymbol{\alpha}$ and $g_{\boldsymbol{\alpha}}$ that minimize the sum

$$\sum_{i=1}^{nS} [y_i - g_{\boldsymbol{\alpha}}(\boldsymbol{\alpha} \mathbf{x}_i)]^2, \quad (3.39)$$

where $\boldsymbol{\alpha} \in R^{nX}$, $\|\boldsymbol{\alpha}\| = 1$, and $g_{\boldsymbol{\alpha}}$ is the outcome of using a scatterplot smoother (e.g., LOESS; see Sect. 3.1.3) on the points $[y_i, \boldsymbol{\alpha} \mathbf{x}_i]$, $i = 1, 2, \dots, nS$. Once $\hat{\boldsymbol{\alpha}}_1$ and \hat{g}_1 are estimated, the partial residuals $y_i - \hat{g}_1(\hat{\boldsymbol{\alpha}}_1 \mathbf{x}_i)$, $i = 1, 2, \dots, nS$, are used to obtain $\hat{\boldsymbol{\alpha}}_2$ and \hat{g}_2 . Specifically, $\hat{\boldsymbol{\alpha}}_2$ and \hat{g}_2 are defined to be the values for $\boldsymbol{\alpha}$ and $g_{\boldsymbol{\alpha}}$ that minimize the sum

$$\sum_{i=1}^{nS} \{ [y_i - \hat{g}_1(\hat{\boldsymbol{\alpha}}_1 \mathbf{x}_i)] - g_{\boldsymbol{\alpha}}(\boldsymbol{\alpha} \mathbf{x}_i) \}^2, \quad (3.40)$$

where $\boldsymbol{\alpha} \in R^{nX}$, $\|\boldsymbol{\alpha}\| = 1$, $\boldsymbol{\alpha}$ and $\hat{\boldsymbol{\alpha}}_1$ are orthogonal, and $g_{\boldsymbol{\alpha}}$ is the outcome of using a scatterplot smoother on the points $[y_i - \hat{g}_1(\hat{\boldsymbol{\alpha}}_1 \mathbf{x}), \boldsymbol{\alpha} \mathbf{x}_i]$, $i = 1, 2, \dots, nS$. This process continues until no appreciable improvement based on a relative error criterion is observed. Unlike additive models, backfitting is typically not used in projection pursuit regression.

The scatterplot smoother typically used at each step in projection pursuit regression is a variable span version of LOESS, called the supersmoother (*supsmu*) in R and *S-Plus* (p. 318, Ref. 109). This presentation will actually use smoothing splines instead. Further discussion on smoothing in the context of projection pursuit regression, smoothing parameter selection and determination of the number of projection terms nD is given in Sect. 4. Additional information on projection pursuit regression is available elsewhere.¹³¹

As indicated in Eq. (3.37), the outcome of a projection pursuit regression consists of the vectors $\boldsymbol{\alpha}_s$, defined for $s = 1, 2, \dots, nD$ and corresponding functions g_s defined for $\boldsymbol{\alpha}_s \mathbf{x}$. Predictions of $y = f(\mathbf{x})$ are then given by

$$y \cong \sum_{s=1}^{nD} g_s(\boldsymbol{\alpha}_s \mathbf{x}), \quad (3.41)$$

for $\mathbf{x} = [x_1, x_2, \dots, x_{nX}]$.

Projection pursuit regression can represent very general situations involving nonlinearity and variable interactions. Further, it avoids the dimensionality problem by using projection terms and additive modeling. However, this generality can come at a price. Results in Sect. 5 suggest that projection pursuit regression has a tendency to overfit the data by including spurious variables in the model.

3.3.4 Recursive Partitioning Regression

Recursive partitioning regression is most commonly known in the form of regression trees.¹³² A regression tree splits the data into subgroups where the observations within each subgroup are more homogeneous than they are over the set of all observations. Then, $f(\mathbf{x})$ in Eq. (3.27) is estimated by the sample mean over each subgroup. The resultant estimate for f is a piecewise constant function, which is also known as a simple function. More precisely, the estimate $\hat{f}(\mathbf{x})$ is given by

$$\hat{f}(\mathbf{x}) = \sum_{s=1}^{nP} c_s I_s(\mathbf{x}), \quad (3.42)$$

where (i) \mathcal{A}_s , $s = 1, 2, \dots, nP$, are the disjoint sets into which the observed values \mathbf{x}_i , $i = 1, 2, \dots, nS$, are partitioned (usually on the basis of the values for y_i), (ii) the mean c_s over each set \mathcal{A}_s is defined by

$$c_s = \sum_{\mathbf{x}_i \in \mathcal{A}_s} y_i / C(\mathcal{A}_s) \quad (3.43)$$

with $C(\mathcal{A}_s)$ denoting the cardinality of \mathcal{A}_s , and (iii) $I_s(\mathbf{x})$ is the indicator function such that $I_s(\mathbf{x}) = 1$ if $\mathbf{x} \in \mathcal{A}_s$ and 0 otherwise. The use of regression trees in sensitivity analysis is illustrated in Mishra et al.¹⁰⁶

Regression trees can be generalized by replacing the mean c_s in Eq. (3.42) with a linear function. In particular, $\hat{f}(\mathbf{x})$ can be defined by

$$\hat{f}(\mathbf{x}) = \sum_{s=1}^{nP} (\hat{\alpha}_s + \hat{\beta}_s \mathbf{x}) I_s(\mathbf{x}), \quad (3.44)$$

where $\hat{\alpha}_s + \hat{\beta}_s \mathbf{x}$ is the least squares linear fit to the data associated with \mathcal{A}_s and I_s is defined the same as in Eq. (3.42). An example of $\hat{f}(\mathbf{x})$ for a single independent variable is given in Fig. 11. The individual regressions can also be constrained so that the regression lines (in one dimension) and regression surfaces (in two or more dimensions) meet continuously. Examples for one and two dimensions are given in Figs. 12 and 13.

The individual regression lines in Figs. 11b and 12b are constructed with a robust regression procedure in which the sum of squares is minimized over the middle two quartiles of the deviations from the regression line (see Ref. 133 for additional information on robust regression). In contrast, the individual regression lines in Figs. 11a and 12a are constructed with the traditional least squares procedure in which the sum of squares is minimized over all deviations from the regression line. The robust regression procedure reduces the effects of large deviations from the overall trend in the data. The effect of this reduction in the examples presented in Figs. 11 and 12 is to produce regression lines that more closely match a visual impression of the trends in the data. The visually appealing nature of the results in Figs. 11b and 12b suggests that robust regression procedures could have a useful role to play in sensitivity analysis due to their effectiveness in reducing the influence of outliers. Although all of the least squares procedures in this presentation are carried out in the traditional manner, the use of robust regression procedures in sensitivity analysis is an area that merits additional investigation.

The linear fit associated with $\hat{f}(\mathbf{x})$ in Eq. (3.44) reduces the need to split the data as many times as is typically the case when a regression tree is used. This approach will certainly outperform a regression tree when the relationship between y and the \mathbf{x}_i 's is close to linear for each partition set \mathcal{A}_s . The interpretation of the representation for $\hat{f}(\mathbf{x})$ in Eq. (3.44) with the linear fit is perhaps less obvious than the interpretation for $\hat{f}(\mathbf{x})$ in Eq. (3.42) with means. However, the primary concern in this presentation is constructing close approximations to the function $f(\mathbf{x})$ that defines y . Which independent variables are important in this approximation can be easily determined by observing the fidelity of $\hat{f}(\mathbf{x}_i)$ to the corresponding values y_i when $\hat{f}(\mathbf{x})$ is constructed with and without the inclusion of individual independent variables. In the examples of Section 5.1, the recursive partitioning approach given here outperformed the regression tree approach indicated in Eqs. (3.42) and (3.43), particularly in terms of estimation of η^2 defined in Eq. (5.13).

The determination of the partition sets \mathcal{A}_s , $s = 1, 2, \dots, nP$, and the associated function $\hat{f}(\mathbf{x})$ is now considered. Let $x_{(r)j}$, $r = 1, 2, \dots, nS$, represent the sampled values for x_j ordered by size (i.e., $x_{(r)j} \leq x_{(r+1)j}$ for $r = 1, 2, \dots, nS - 1$), and let \mathcal{A}_{rj1} and \mathcal{A}_{rj2} denote the sets defined by

$$\mathcal{A}_{rj1} = \left\{ \mathbf{x}_i : x_{ij} \leq x_{(r)j} \right\} \quad (3.45)$$

and

$$\mathcal{A}_{rj2} = \left\{ \mathbf{x}_i : x_{ij} > x_{(r)j} \right\} \quad (3.46)$$

for $r = 1, 2, \dots, nS$ and $j = 1, 2, \dots, nX$. A separate linear regression is performed for the set of (y_i, \mathbf{x}_i) pairs associated with each of the sets \mathcal{A}_{rj1} and \mathcal{A}_{rj2} . Some of the sets will have too few data pairs (i.e., less than $nX + 1$) to fit a linear regression model and are excluded from consideration. This results in a total of $nX(nS - 2nX - 1)$ pairs $[\mathcal{A}_{rj1}, \mathcal{A}_{rj2}]$ that are candidates to define initial values for \mathcal{A}_1 and \mathcal{A}_2 .

The pair $[\mathcal{A}_{rj1}, \mathcal{A}_{rj2}]$ with regressions that together provide the best representation for y are selected as the initial values for \mathcal{A}_1 and \mathcal{A}_2 . This determination is made on the basis of the R^2 value given by

$$R_{rj}^2 = \frac{SST - SSE_{rj}}{SST} \quad (3.47)$$

for each pair $[\mathcal{A}_{rj1}, \mathcal{A}_{rj2}]$, where

$$SST = \sum_{i=1}^{nS} (y_i - \bar{y})^2 = \sum_{i=1}^{nS} \left(y_i - \frac{\sum_{i=1}^{nS} y_i}{nS} \right)^2,$$

$$SSE_{rj} = SSE(\mathcal{A}_{rj1}) + SSE(\mathcal{A}_{rj2}),$$

and $SSE(\mathcal{A}_{rj1})$ and $SSE(\mathcal{A}_{rj2})$ denote the error sum of squares for the linear regressions associated with \mathcal{A}_{rj1} and \mathcal{A}_{rj2} , respectively. The selection

$$[\mathcal{A}_1, \mathcal{A}_2] = [\mathcal{A}_{rj1}, \mathcal{A}_{rj2}] \quad (3.48)$$

is then made for the pair $[\mathcal{A}_{rj1}, \mathcal{A}_{rj2}]$ that has the largest value for R_{rj}^2 .

With initial values for \mathcal{A}_1 and \mathcal{A}_2 determined, consideration is given to splitting \mathcal{A}_1 and \mathcal{A}_2 into two subsets to produce three sets of \mathbf{x}_i values. This involves consideration of triples of sets of the form $[\mathcal{U}_1, \mathcal{V}_1, \mathcal{A}_2]$ and $[\mathcal{A}_1, \mathcal{U}_2, \mathcal{V}_2]$, where (i) \mathcal{U}_1 and \mathcal{V}_1 correspond to subsets of \mathcal{A}_1 obtained in a manner analogous to that used in the definition of \mathcal{A}_{rj1} and \mathcal{A}_{rj2} and (ii) \mathcal{U}_2 and \mathcal{V}_2 correspond to subsets of \mathcal{A}_2 also obtained in a manner analogous to that used in the definition of \mathcal{A}_{rj1} and \mathcal{A}_{rj2} . A triple of sets $[\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3]$ is then defined that is equal to the $[\mathcal{U}_1, \mathcal{V}_1, \mathcal{A}_2]$ or $[\mathcal{A}_1, \mathcal{U}_2, \mathcal{V}_2]$.

$\mathcal{U}_2, \mathcal{V}_2]$ triple that has the highest R^2 value obtained in a manner analogous to that described in conjunction with Eq. (3.47) except that results obtained from regressions involving three sets are involved. This process of constructing additional sets is then continued in an analogous manner until further splitting would not be beneficial as determined by some stopping criterion.

Prediction of $y = f(\mathbf{x})$ for arbitrary values of \mathbf{x} is straightforward once the construction process to obtain $\mathcal{A}_s, \hat{\alpha}_s$ and $\hat{\beta}_s$ is complete. Specifically, the desired prediction follows directly from Eq. (3.44).

Since the determination of the partition regions is data driven (i.e., based on the observed y values), the smoother matrix \mathbf{S} for recursive partitioning regression depends on the y values and is hence not a linear smoother. Because of this, an equivalent degrees of freedom is hard to define. However, a possible definition is to use the degrees of freedom from the model obtained as if the partitions had been specified *a priori*, and then add a certain number of degrees of freedom for each partition.

If the partitions were known *a priori*, then the smoother matrix derives from the regression analyses carried out for each set $\mathcal{A}_s, s = 1, 2, \dots, nP$, and can be constructed from the design matrices \mathbf{X}_s associated with these regressions (see Eq. (2.6)). In particular, \mathbf{S} is constructed from the matrices

$$\mathbf{H}_s = \mathbf{X}_s (\mathbf{X}_s^T \mathbf{X}_s)^{-1} \mathbf{X}_s^T \quad (3.49)$$

and has the form

$$\mathbf{S} = \begin{bmatrix} \mathbf{H}_1 & 0 & \dots & 0 \\ 0 & \mathbf{H}_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \mathbf{H}_{nP} \end{bmatrix} \quad (3.50)$$

when the rows of \mathbf{X} are rearranged so rows that correspond to elements of \mathcal{A}_s are next to each other. Further, the degrees of freedom for the model with the sets \mathcal{A}_s specified *a priori* is

$$df_{ap} = tr(\mathbf{S}) = \sum_{s=1}^{nP} tr(\mathbf{H}_s), \quad (3.51)$$

which corresponds to the number of degrees of freedom associated with \mathbf{S} . Given that the regression for each set \mathcal{A}_s involves the determination of $nX + 1$ parameters (i.e., the coefficients in the regression model), $tr(\mathbf{H}_s) = nX + 1$; as a result,

$$df_{ap} = nP(nX + 1) \quad (3.52)$$

is the degrees of freedom for \mathbf{S} .

However, as the sets $\mathcal{A}_s, s = 1, 2, \dots, nP$, are not specified *a priori*, additional degrees of freedom are involved in the estimation of these partitions. Since the complexity of the partitioned regions increases with the number of independent variables, each additional partition can be viewed as involving another nX degrees of freedom. As a result,

$$df = nP(nX + 1) + nX(nP - 1) \quad (3.53)$$

is an estimate of the degrees of freedom for the entire recursive partitioning model.

In our experience this rule works quite well for determining equivalent degrees of freedom for the recursive partitioning procedure described above. Additional discussion about equivalent degrees of freedom for adaptive or “data driven” approaches such as recursive partitioning and Multivariate Adaptive Regression Splines (MARS) is given in Hastie et al.¹³⁴

Determining the number of sets $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_{nP}$ to use in recursive partitioning regression is analogous to choosing the smoothing parameter in previously described methods. Therefore, a reasonable approach is to determine a stopping point in the partitioning process with a criterion similar to that used for the selection of a smoothing parameter such as cross validation or generalized cross validation (Sect. 3.2). With cross validation, the PRESS value PRS is calculated as indicated in Eq. (3.19); similarly, with generalized cross validation, the adjusted PRESS value PRS_A is calculated as indicated in Eqs. (3.22) and (3.25). Then, if $PRS_s, s = 1, 2, \dots$, and $PRS_{As}, s = 1, 2, \dots$, represent values for PRS and PRS_A , respectively, calculated at successive steps in the partitioning process, an appropriate stopping point would be the last step before these values begin to increase as such an increase is indicative of an overfitting of the data.

As is the case for LOESS (Sect. 3.3.1), additive models (Sect. 3.3.2), and projection pursuit regression (Sect. 3.3.3), recursive partitioning regression can model very general nonlinear relationships. It also models very general interactions and performs well in higher dimensions. Unlike projection pursuit regression, the results in Sect. 5 do not indicate a tendency to overfit the data. However, these desirable properties come at a cost as recursive partitioning regression can require an order of magnitude more computational effort than the other indicated methods.

3.4 Hypothesis Testing for Variable Importance

A number of possibilities exist for hypothesis testing for smoothing methods, including the use of approximate distributions and bootstrapping.¹²³ For reasons of computational efficiency, the following approach to hypothesis testing is practicable for use in sensitivity analyses employing stepwise nonparametric regression and is used in this presentation for that reason.

It is desired to compare results (i.e., estimates of y) obtained with a model constructed from n independent variables (i.e., variables corresponding to n different elements of $\mathbf{x} = [x_1, x_2, \dots, x_n]$) with results obtained from a model constructed without one of these n variables, say x_j . The goal is to test the following hypotheses:

H_0 : Results obtained with and without the inclusion of x_j are the same.

H_a : Results obtained with the inclusion of x_j are different from the results obtained with the exclusion of x_j .

If H_0 can be rejected in favor of H_a , then x_j is an important variable and should be included in the model being constructed by the smoothing process.

The usual test statistic for choosing between H_0 and H_a for linear models is

$$F^* = \frac{(SSE_R - SSE_F)/(df_F - df_R)}{SSE_F/df_F}, \quad (3.54)$$

where SSE_R and df_R are the error sum of squares and degrees of freedom for the reduced model (i.e., the model without x_j) and SSE_F and df_F are defined similarly for the full model (i.e., the model with x_j included) (Ref. 135, p. 169; Ref. 136, Sect. 4.4). In the testing of linear models with normally distributed data, F^* has an F -distribution with $m = df_F - df_R$ and $n = df_F$ degrees of freedom when H_0 is true. As a result, a p -value equal to $prob(F > F^*|H_0)$ can be used to test H_0 against H_a , where $prob(F > F^*|H_0)$ is the probability that a value F for the F -statistic greater than F^* will be obtained by chance if the null hypothesis H_0 is satisfied.

The statistic F^* in Eq. (3.54) can also be defined for models with $n - 1$ and n variables constructed in a smoothing process with df_R and df_F defined by

$$df_R = tr(\mathbf{S}_R) \text{ and } df_F = tr(\mathbf{S}_F), \quad (3.55)$$

where \mathbf{S}_R and \mathbf{S}_F are the smoother matrices associated with the reduced model (i.e., the model without x_j) and the full model (i.e., the model with x_j included), respectively. Unfortunately, the true distribution for F^* is not known for any of the smoothing methods considered in this presentation. However, the distribution for F^* for these smoothing methods can be approximated by an F -distribution with $df_F - df_R$ and df_F degrees of freedom (i.e., F_{rs} with $r = df_F - df_R$ and $s = df_F$; see pp. 66 – 67, Ref. 121).

Determination of whether or not a particular variable should be included in a smoothing process can be made by fitting the associated model with and without the variable and then performing the appropriate F test. This is particularly useful in sensitivity analysis where the objective is to identify the important variables.

Performance of a comprehensive robustness study of this approach would be very beneficial. Our experience with the results contained in this presentation indicates that the approach is quite reasonable. Other approximate

tests for H_0 are also available (pp. 87 – 89, Ref. 123); however, none of these tests are exact. Fortunately, such tests in and of themselves do not have a large bearing on which variables are identified as being important in the stepwise procedures described in this presentation. Rather, it is the contribution of a variable to the model R^2 value that serves as the metric for variable importance (see discussion of R^2 in Sect. 5.1). Hypothesis testing merely serves as a model building tool in the stepwise variable selection discussed in Sect. 4.1.

4. Implementation of Smoothing Methods for Sensitivity Analysis

Explanations are now given on how the different smoothing methods for surface approximation can be used in sensitivity analysis. Details about the forward (i.e., stepwise) model building process and smoothing parameter selection are given. In particular, the following topics are considered in the context of sensitivity analysis: stepwise variable selection (Sect. 4.1), traditional regression methods (Sect. 4.2), locally weighted regression, i.e., LOESS (Sect. 4.3), generalized additive models (Sect. 4.4), projection pursuit regression (Sect. 4.5), and recursive partitioning regression (Sect. 4.6). All of the techniques discussed here and used in the examples of Ref. 107 were implemented using the R language, which is an open source language very similar to S-Plus.

4.1 Stepwise Variable Selection

For purposes of sensitivity analysis, all of the presented regression (i.e., smoothing) methods can be implemented with a forward stepwise variable selection procedure. An approach of this type is essential for sensitivity analyses as there are usually a large number of uncertain analysis inputs under consideration (e.g., $nX \cong 150$ in the NUREG-1150 probabilistic risk assessments,^{55, 137-140} $nX \cong 60$ in the compliance certification application for the Waste Isolation Pilot Plant,¹⁴¹ and $nX \cong 250$ in an analysis for the proposed Yucca Mountain facility for the disposal of high level radioactive waste^{142, 143}). Nonparametric regression techniques are not suitable for constructing models that contain a large number of independent variables unless the sample size is very large. Hence, it is essential to have a method that does not include all the variables under consideration in a model at once. Further, the order in which variables are selected in an appropriately designed stepwise procedure provides important sensitivity information.

A forward stepwise selection procedure operates in the following manner. A single variable model is constructed using each of the independent variables. Thus, if nX independent variables are under consideration, this results in the construction of nX single variable models. The variable, say \tilde{x}_1 , associated with the best of these models is identified and retained. Then, two variable models are constructed using \tilde{x}_1 and each of the remaining $nX - 1$ variables. This results in the construction of $nX - 1$ two variable models. The variable, say \tilde{x}_2 , associated with the best of these models is identified and retained. The process then continues with the construction of three variable models with \tilde{x}_1 , \tilde{x}_2 , and the remaining $nX - 2$ variables, and so on. This process continues until some stopping

criterion is reached that indicates that no additional predictive capability is provided by models with additional variables.

Two important questions are left unanswered in the preceding paragraph: (i) What determines which model, and hence which variable, is best in a set of models?, and (ii) What is an appropriate stopping criterion? The best model is usually determined on the basis of a p -value (Sect. 3.4). The variable associated with the model with the smallest p -value is considered to provide the most predictive capability and is thus the variable retained for use in the next step in the model construction process. The p -value is also used to provide a stopping criterion. In particular, when the minimum p -value over all models is greater than some cutoff value (e.g., 0.02), no variable is selected and the model construction process terminates with the model constructed at the preceding step.

Variable selection can also be based on the PRESS statistic (see Eq. (3.19)). At each step in the selection process, the variable whose inclusion results in the smallest PRESS value is the variable retained. If the minimum PRESS value at a step is larger than the minimum PRESS value of the preceding step, then no variable is selected and the model construction process terminates with the model constructed at the preceding step. Other selection and stopping criteria are also possible, including Akaike's information criterion (p. 158, Ref. 121), adjusted R^2 values (described in Sect. 2 of Ref. 107), and the adjusted PRESS statistic (described in Sect. 3.2). An enhancement of the forward selection procedure is to allow for the possibility of a previously selected variable being dropped from the modeling process if it no longer contributes significant predictive capability as additional variables are selected and included in the model.

Backward stepwise selection involves fitting a model with all nX variables. Then, unimportant variables are sequentially removed until the removal of additional variables reduces the predictive capability of the model. At this point the process is terminated. This selection procedure is not appropriate for sensitivity analysis with nonparametric regression models for two reasons. First, the construction of nonparametric regression models with a large number of variables is not possible with relatively small sample sizes. Second, the backward selection procedure is not useful for identifying the importance of individual variables, which is the primary goal of sensitivity analysis. In contrast, a well designed forward selection procedure identifies the most important variable at the first step, then the next most important variable at the second step, and so on.

The example sensitivity results presented in Ref. 107 use a p -value criterion (Sect. 3.4) for both individual variable selection and termination of the model construction process. Preliminary results indicated that use of a PRESS criterion was too computationally demanding for some of the regression methods and also resulted in models that tended to overfit the data. Our experience is that using either the p -value with a cutoff of $\sigma = 0.02$ or the adjusted PRESS statistic PRS_A for model selection usually results in the same model.

4.2 Traditional Regression: Linear Regression (LIN_REG), Rank Regression (RANK_REG) and Quadratic Regression (QUAD_REG)

Each of the traditional regression approaches (i.e., LIN_REG, RANK_REG and QUAD_REG) can be implemented the same way with a forward stepwise selection procedure and a p -value criterion of $\alpha = 0.02$ (Table 1). The forward selection procedure with QUAD_REG requires some additional explanation as there are many ways to structure this procedure to incorporate variable interactions and squares. The approach taken is to consider a variable, its square, and all two-way interactions at each step of the selection procedure. Thus, if \tilde{x}_1 is the first variable selected, then the corresponding model would be of the form

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \tilde{x}_1 + \hat{\beta}_{11} \tilde{x}_1^2. \quad (4.1)$$

Then, if \tilde{x}_2 is the second variable selected, the corresponding model would be of the form

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \tilde{x}_1 + \hat{\beta}_2 \tilde{x}_2 + \hat{\beta}_{12} \tilde{x}_1 \tilde{x}_2 + \hat{\beta}_{11} \tilde{x}_1^2 + \hat{\beta}_{22} \tilde{x}_2^2, \quad (4.2)$$

and so on.

4.3 Locally Weighted Regression: LOESS

The forward stepwise procedure with LOESS can be implemented with a two stage variable selection process at each step (Table 2). First, multiple spans (i.e., r/nS , where $r - 1$ is the number of points averaged over and nS is the sample size; see Sect. 3.1.3) are considered for each candidate variable, and a LOESS model is constructed for each span. Specifically, models are constructed for the following spans: 10, 0.7, 0.3, 0.1, 0.07 and 0.05, where 10 is simply an indicator for the use of a linear regression model. This results in five models for each candidate variable. Then, the “best” of these five models is selected on the basis of generalized cross validation using the adjusted PRESS value PRS_A defined in Eqs. (3.22) and (3.25). This produces one selected model (i.e., the model with the smallest value for PRS_A) for each candidate variable. Second, the “best” of the selected models for the individual candidate variables is identified with the approximate hypothesis test indicated in conjunction with Eqs. (3.54). Specifically, the model with the smallest p -value is identified, and the associated candidate variable is the variable selected at that step in the stepwise procedure. The procedure terminates with no variable selected if all p -values exceed $\alpha = 0.02$.

The flexibility provided by the different choices for span causes a potential problem with the approximate hypothesis test indicated in conjunction with Eq. (3.54). In particular, adding a variable to an existing model could result in a new model with the same or fewer degrees of freedom. This can happen if the span selected for the new model is much larger than the span for the previous model. This possibility exists because smaller spans produce

more complex models and thus result in models with larger numbers of equivalent degrees of freedom. In turn, this would result in the numerator degrees of freedom (i.e., $df_F - df_R$) for the F statistic in Eq. (3.54) being less than zero.

This problem is handled in the following manner with SSE_R , SSE_F , df_R and df_F defined as in Eq. (3.54):

(i) If $df_F > df_R$, define the p -value with the F test as usual and test against α .

(ii) If $df_F - df_R = 0$, then define the p -value by

$$p\text{-value} = \begin{cases} 0 & \text{if } SSE_F < SSE_R \\ 1 & \text{if } SSE_F \geq SSE_R \end{cases} \quad (4.3)$$

and test against α .

(iii) If $df_F \leq df_R$, a p -value is defined with an F test involving

$$F^* = \frac{(SSE_F - SSE_R)/(df_R - df_F)}{SSE_R/df_R}, \quad (4.4)$$

which is the original F statistic (see Eq. (3.54)) with the roles of the two models (i.e., the reduced or old model and the full or new model) reversed. The new variable should be added to the model only if evidence exists that the resultant model is better than the previous model. Such evidence is provided if the new model results in a significant reduction in the degrees of freedom (i.e., $df_R - df_F > 0$) without significantly increasing the error sum of squares (i.e., if $SSE_F - SSE_R$ is “small”). The preceding implies that the old model should be rejected in favor of the new model for sufficiently small values of F^* as defined in Eq. (4.4). In particular, the associated p -value is given by $prob(F < F^*)$ for an F distribution with $m = df_R - df_F$ and $n = df_R$ degrees of freedom. For the special case $SSE_F - SSE_R \leq 0$, the corresponding p -value is assumed to be zero. The usual test against α is made (i.e., the new model is accepted if the resultant p -value is less than α).

4.4 Generalized Additive Models (GAMs)

Additive models are now considered. Such models are designated as GAMs (generalized additive models) after the *gam* function in the R and S-Plus languages (p. 252, Ref. 121). The descriptor “generalized” is used to indicate fitting of a discontinuous response. The designator GAM is used to be consistent with the *gam* function in R and S-Plus.

Similarly to the forward stepwise procedure with LOESS (Table 2), the forward stepwise procedure with GAMs can be implemented with a two stage variable selection process at each step (Table 3). First, multiple additive models are constructed for each candidate variable, and the “best” of these models is selected for each candidate variable. Second, the “best” of the selected models for the individual variables is identified with the approximate

hypothesis test indicated in conjunction with Eq. (3.54). Specifically, the model with the smallest p -value is identified, and the associated candidate variable is the variable selected at that step. The process terminates with no variable selected if all p -values exceed $\alpha = 0.02$.

An additive model is constructed by repeatedly smoothing residuals across one independent variable at a time (Sect. 3.3.2). In concept, any scatterplot smoother could be used. However, the *gam* function in R and S-Plus is restricted to LOESS (Sect. 3.1.3) and/or smoothing splines (Sect. 3.1.4). Both smoothers usually gave similar results in some preliminary analyses, but an occasional convergence problem was encountered with LOESS. Therefore, smoothing splines are indicated in Table 3 and used in the generation of the GAM results presented in Sect. 5. Specifically, at a given step in the stepwise procedure, multiple degrees of freedom are considered (i.e., 1, 2, 4, 7, 10, 15) for each candidate variable, and a GAM is constructed using smoothing splines for each of these degrees of freedom. This results in six models for each candidate variable. After this construction, it is then necessary to select that “best” of the these six models for each candidate variable.

The indicated selection is made on the basis of generalized cross validation (Sect. 3.2) employing the adjusted PRESS values PRS_A (see Eqs. (3.22) – (3.25)). This criterion for model selection was picked because there is not an option associated with the *gam* function to use cross validation (Sect. 3.2) inside the back fitting algorithm. Further, computing PRESS (see Eq. (3.19)) is difficult because the leverage values s_{ii} are not obtainable for a *gam* fit in R and thus cannot be used in a computationally efficient calculation of PRESS (see Eq. (3.21)). As a result, obtaining the PRESS statistic in R would require fitting a model nS times, where nS is the sample size, and then making nS predictions. Thus, use of cross validation with PRESS in the generation of GAMs with R is computationally very expensive.

In contrast, use of generalized cross validation allows a more computationally efficient determination of the “best” model associated with each candidate variable (Sect. 3.2). In particular, deleted residuals are not needed in generalized cross validation (see Eq. (3.21)). Instead, generalized cross validation is based on the adjusted PRESS value PRS_A , which uses $tr(\mathbf{S})$ in its evaluation (see Eqs. (3.22) – (3.25)). The value for $tr(\mathbf{S})$ can be estimated as indicated in Eq. (3.36). This estimation requires the degrees of freedom df_j (i.e., 1, 2, 4, 7, 10 or 15) used for each variable x_j in the scatterplot smoother employed in the backfitting algorithm. Because the values for df_j are known for each GAM constructed for a given candidate variable (see description of backfitting algorithm in Table 3), the determination of $tr(\mathbf{S})$ (see Eq. (3.36)) and hence PRS_A (see Eq. (3.22)) is straightforward for each of these GAMs. In turn, the selected (i.e., “best”) GAM for a given candidate variable is the model with the smallest value for PRS_A . As already indicated, once the “best” GAM for each variable is identified, the “best” GAM overall is determined on the basis of the p -value associated with the approximate hypothesis test in Eq. (3.54), and the variable selected at the step under consideration is the variable associated with that model.

4.5 Projection Pursuit Regression (PP_REG)

Similarly to the forward stepwise procedures with LOESS (Table 2) and GAMs (Table 3), the forward stepwise procedure with PP_REG can be implemented with a two stage variable selection process at each step (Table 4). First, multiple PP_REG models are constructed for each candidate variable, and the “best” of these models is selected for each candidate variable. Second, the “best” of the selected models for the individual variables is identified with the approximate hypothesis test indicated in conjunction with Eq. (3.54). Specifically, the model with the smallest p -value is identified, and the associated variable is the variable selected at that step. The process terminates with no variable selected if all p -values exceed $\alpha = 0.02$.

The default implementation of PP_REG in the function *ppr* in R and S-Plus uses a scatterplot smoother called *supsmu*, which is a variable span smoother that usually provides a better fit to data than a fixed span smoother.¹³¹ However, the bandwidth at a particular value for \mathbf{x}_i depends on the values for y , which makes this smoother nonlinear. In turn, this makes the equivalent degrees of freedom difficult to define. Without the degrees of freedom, it is difficult to assess the quality of the associated model. Unfortunately, a high R^2 value by itself is not very informative because there is no way to know if an overfit of the data has occurred. A possibility is to use the PRESS statistic to assess the quality of the fit, but this can be very time consuming for moderately large samples. The preceding complication is avoided in this study by using the option of employing smoothing splines as the scatterplot smoother in *ppr* with a degrees of freedom δ_k specified for each smoothing operation (see Table 4). With this option, the resultant degrees of freedom associated with the smoothing operations can be obtained directly from *ppr* rather than approximated from the δ_k 's as is done in the stepwise procedure from the construction of GAMs (see Table 3).

As iterative smoothing operations are applied, the possibility exists that the degrees of freedom will decrease when a variable is added to a model. This situation occurs when the successor model is less complex (i.e., involves less smoothing) than the predecessor model. When this occurs, the procedure described for use in the same situation with LOESS is applied (Sect. 4.3).

4.6 Recursive Partitioning Regression (RP_REG)

As for LOESS (Table 2), GAMs (Table 3) and PP_REG (Table 4), the forward stepwise procedure with RP_REG can be implemented with a two stage variable selection process at each step (Table 5). First, multiple RP_REG models are constructed for each candidate variable, and the “best” of these models is selected for each candidate variable. Second, the “best” of the selected models for the individual variables is identified with the approximate hypothesis test indicated in conjunction with Eq. (3.54). Specifically, the model with the smallest p -value is identified, and the associated candidate variable is the variable selected at that step. The process terminates with no variable selected if all p -values exceed $\alpha = 0.02$.

For each candidate variable at each step in the partitioning process, it is necessary to investigate a large number of possible split points (Sect. 3.3.4). Each possible split point requires the construction of a regression model. For example, if the partitioning process has reached the point that five variables are under consideration, then each possible split point requires the construction of a regression model with six parameters. To reduce the number of required regression constructions, every observation for a variable is not investigated as a possible split point. Instead, for every variable, a split point is considered at the smallest sampled value possible for use in splitting and then at every k^{th} observed value after that up to the largest sampled value possible for use in splitting. For example, if a sample of size $nS = 300$ is under consideration, $k = 3$ and the partitioning process has reached the point at which five independent variables are under consideration in the regression model construction, then the possible split points for the variable x_j would be $x_{(6)j}, x_{(9)j}, \dots, x_{(291)j}, x_{(294)j}$, where $x_{(i)j}$, $i = 1, 2, \dots, 300$, denotes a rank ordering of the observed values for variable x_j . In this example, the smallest possible value for splitting is $x_{(6)j}$ because at least six observations are required to estimate the six parameters in the associated regression model. In the examples presented in Sect. 5, $k = 2$ is used when $nS = 100$, and $k = 3$ is used when $nS = 300$.

As indicated in Table 5, the split point that results in the largest increase in R^2 defines the split point to be used (see Eqs. (3.47) – (3.48)). If the adjusted PRESS value PRS_A is smaller after the split than before, the split is kept and the search continues for the next possible split point. The construction process continues in this manner until PRS_A increases after a split, at which point the split is not kept and the model is completed for that step and the particular candidate variable under consideration. Then, the F -static and the associated p -value are determined for each model constructed at this step in a comparison with the model retained at the preceding step. As with the other procedures, a cutoff of $\alpha = 0.02$ for the approximate p -value is used in the stepwise variable selection procedure (Sect. 4.1) to determine whether or not a new variable should be retained in the stepwise procedure.

5. Summary

Sampling-based approaches to uncertainty and sensitivity analysis are very popular. With such approaches, a probabilistically-based sample is generated from the distributions that characterize the uncertainty in analysis inputs and then the elements of this sample are propagated through the analysis. The resultant distributions of analysis outcomes provide the desired uncertainty analysis as these distributions summarize the uncertainty in analysis outcomes that derives from uncertainty in analysis inputs. Further, this propagation provides a mapping between uncertain analysis inputs and analysis outcomes that can be explored with a variety of sensitivity analysis procedures. Among the most popular of these procedures is stepwise regression analysis with raw or rank-transformed data. Unfortunately, regression analyses with raw data are ineffective when the relationships between analysis inputs and analysis outcomes are nonlinear, and regression analyses with rank-transformed data are ineffective when the relationships between analysis inputs and analysis outcomes are both nonlinear and nonmonotonic.

Nonparametric regression procedures constitute a promising alternative to more traditional parametric regression procedures for use in sampling-based sensitivity analyses that involve relationships between analysis inputs and analysis outcomes that are both nonlinear and nonmonotonic. This promise derives from the local manner in which nonparametric regression procedures attempt to match the relationships between analysis inputs and analysis outcomes. In contrast, parametric regression procedures attempt to match global relationships between analysis inputs and analysis outcomes. As a result, nonparametric regression procedures can match relationships between analysis inputs and analysis outcomes on a smaller scale than is possible with parametric regression procedures.

The following nonparametric regression procedures with potential for use in sampling-based sensitivity analyses have been introduced and briefly summarized: (i) locally weighted regression (LOESS), (ii) additive models, (iii) projection pursuit regression, and (iv) recursive partitioning regression. Further, algorithms for the stepwise implementation of these procedures in the R language as part of a sensitivity analysis have been described.

The second part of this presentation¹⁰⁷ illustrates the stepwise implementation of these procedures as parts of sampling-based sensitivity analyses. Specifically, the procedures are illustrated with both simple test problems and results from a performance assessment for the Waste Isolation Pilot Plant (WIPP).^{56, 57} As shown by the example illustrations, nonparametric regression procedures can yield more informative sensitivity analysis results than can be obtained with more traditional parametric regression procedures when nonlinear relationships between analysis inputs and analysis outcomes are present.

Acknowledgements

Work performed for Sandia National Laboratories (SNL), which is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Security Administration under contract DE-AC04-94AL-85000. Review at SNL provided by L. Swiler and C. Sallaberry. Editorial support provided by F. Puffer, K. Best, M. Spielman, and J. Ripple of Tech Repts, a division of Ktech Corporation.

References

1. Christie, M.A., J. Glimm, J.W. Grove, D.M. Higdon, D.H. Sharp, and M.M. Wood-Schultz. 2005. "Error Analysis and Simulations of Complex Phenomena," *Los Alamos Science*. Vol. 29, pp. 6-25.
2. Sharp, D.H. and M.M. Wood-Schultz. 2003. "QMU and Nuclear Weapons Certification: What's Under the Hood?," *Los Alamos Science*. Vol. 28, pp. 47-53.
3. Wagner, R.L. 2003. "Science, Uncertainty and Risk: The Problem of Complex Phenomena," *APS News*. Vol. 12, no. 1, pp. 8.

4. Oberkampf, W.L., S.M. DeLand, B.M. Rutherford, K.V. Diegert, and K.F. Alvin. 2002. "Error and Uncertainty in Modeling and Simulation," *Reliability Engineering and System Safety*. Vol. 75, no. 3, pp. 333-357.
5. Risk Assessment Forum. 1997. *Guiding Principles for Monte Carlo Analysis*, EPA/630/R-97/001. Washington DC: U.S. Environmental Protection Agency. (Available from the NTIS as PB97-188106/XAB.).
6. NCRP (National Council on Radiation Protection and Measurements). 1996. *A Guide for Uncertainty Analysis in Dose and Risk Assessments Related to Environmental Contamination*, NCRP Commentary No. 14. Bethesda, MD: National Council on Radiation Protection and Measurements.
7. NRC (National Research Council). 1994. *Science and Judgment in Risk Assessment*, Washington, DC: National Academy Press.
8. NRC (National Research Council). 1993. *Issues in Risk Assessment*. Washington, DC: National Academy Press.
9. U.S. EPA (U.S. Environmental Protection Agency). 1993. *An SAB Report: Multi-Media Risk Assessment for Radon, Review of Uncertainty Analysis of Risks Associated with Exposure to Radon*, EPA-SAB-RAC-93-014. Washington, DC: U.S. Environmental Protection Agency.
10. IAEA (International Atomic Energy Agency). 1989. *Evaluating the Reliability of Predictions Made Using Environmental Transfer Models*, Safety Series No. 100. Vienna: International Atomic Energy Agency.
11. Beck, M.B. 1987. "Water-Quality Modeling: A Review of the Analysis of Uncertainty," *Water Resources Research*. Vol. 23, no. 8, pp. 1393-1442.
12. Cacuci, D.G. 2003. *Sensitivity and Uncertainty Analysis, Vol. 1: Theory*. Boca Raton, FL: Chapman and Hall/CRC Press.
13. Turányi, T. 1990. "Sensitivity Analysis of Complex Kinetic Systems. Tools and Applications," *Journal of Mathematical Chemistry*. Vol. 5, no. 3, pp. 203-248.
14. Rabitz, H., M. Kramer, and D. Dacol. 1983. "Sensitivity Analysis in Chemical Kinetics," *Annual Review of Physical Chemistry*. Vol. 34. Eds. B.S. Rabinovitch, J.M. Schurr, and H.L. Strauss. Palo Alto, CA: Annual Reviews Inc, pp. 419-461.
15. Lewins, J. and M. Becker, eds. 1982. *Sensitivity and Uncertainty Analysis of Reactor Performance Parameters*. Vol. 14. New York, NY: Plenum Press.
16. Frank, P.M. 1978. *Introduction to System Sensitivity Theory*. New York, NY: Academic Press.

17. Tomovic, R. and M. Vukobratovic. 1972. *General Sensitivity Theory*. New York, NY: Elsevier.
18. Myers, R.H., D.C. Montgomery, G.G. Vining, C.M. Borror, and S.M. Kowalski. 2004. "Response Surface Methodology: A Retrospective and Literature Review," *Journal of Quality Technology*. Vol. 36, no. 1, pp. 53-77.
19. Myers, R.H. 1999. "Response Surface Methodology - Current Status and Future Directions," *Journal of Quality Technology*. Vol. 31, no. 1, pp. 30-44.
20. Andres, T.H. 1997. "Sampling Methods and Sensitivity Analysis for Large Parameter Sets," *Journal of Statistical Computation and Simulation*. Vol. 57, no. 1-4, pp. 77-110.
21. Kleijnen, J.P.C. 1997. "Sensitivity Analysis and Related Analyses: A Review of Some Statistical Techniques," *Journal of Statistical Computation and Simulation*. Vol. 57, no. 1-4, pp. 111-142.
22. Kleijnen, J.P.C. 1992. "Sensitivity Analysis of Simulation Experiments: Regression Analysis and Statistical Design," *Mathematics and Computers in Simulation*. Vol. 34, no. 3-4, pp. 297-315.
23. Sacks, J., W.J. Welch, T.J. Mitchel, and H.P. Wynn. 1989. "Design and Analysis of Computer Experiments," *Statistical Science*. Vol. 4, no. 4, pp. 409-435.
24. Morton, R.H. 1983. "Response Surface Methodology," *Mathematical Scientist*. Vol. 8, pp. 31-52.
25. Mead, R. and D.J. Pike. 1975. "A Review of Response Surface Methodology from a Biometric Viewpoint," *Biometrics*. Vol. 31, pp. 803-851.
26. Myers, R.H. 1971. *Response Surface Methodology*. Boston, MA: Allyn and Bacon.
27. Helton, J.C. and F.J. Davis. 2003. "Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems," *Reliability Engineering and System Safety*. Vol. 81, no. 1, pp. 23-69.
28. Helton, J.C. and F.J. Davis. 2002. "Illustration of Sampling-Based Methods for Uncertainty and Sensitivity Analysis," *Risk Analysis*. Vol. 22, no. 3, pp. 591-622.
29. Helton, J.C. and F.J. Davis. 2000. "Sampling-Based Methods," *Sensitivity Analysis*. Ed. A. Saltelli, K. Chan, and E.M. Scott. New York, NY: Wiley. pp. 101-153.
30. Kleijnen, J.P.C. and J.C. Helton. 1999. "Statistical Analyses of Scatterplots to Identify Important Factors in Large-Scale Simulations, 1: Review and Comparison of Techniques," *Reliability Engineering and System Safety*. Vol. 65, no. 2, pp. 147-185.

31. Blower, S.M. and H. Dowlatabadi. 1994. "Sensitivity and Uncertainty Analysis of Complex Models of Disease Transmission: an HIV Model, as an Example," *International Statistical Review*. Vol. 62, no. 2, pp. 229-243.
32. Saltelli, A., T.H. Andres, and T. Homma. 1993. "Sensitivity Analysis of Model Output. An Investigation of New Techniques," *Computational Statistics and Data Analysis*. Vol. 15, no. 2, pp. 445-460.
33. Iman, R.L. 1992. "Uncertainty and Sensitivity Analysis for Computer Modeling Applications," *Reliability Technology - 1992, The Winter Annual Meeting of the American Society of Mechanical Engineers, Anaheim, California, November 8-13, 1992*. Eds. T.A. Cruse. Vol. 28, pp. 153-168. New York, NY: American Society of Mechanical Engineers, Aerospace Division.
34. Saltelli, A. and J. Marivoet. 1990. "Non-Parametric Statistics in Sensitivity Analysis for Model Output. A Comparison of Selected Techniques," *Reliability Engineering and System Safety*. Vol. 28, no. 2, pp. 229-253.
35. Iman, R.L., J.C. Helton, and J.E. Campbell. 1981. "An Approach to Sensitivity Analysis of Computer Models, Part 1. Introduction, Input Variable Selection and Preliminary Variable Assessment," *Journal of Quality Technology*. Vol. 13, no. 3, pp. 174-183.
36. Iman, R.L., J.C. Helton, and J.E. Campbell. 1981. "An Approach to Sensitivity Analysis of Computer Models, Part 2. Ranking of Input Variables, Response Surface Validation, Distribution Effect and Technique Synopsis," *Journal of Quality Technology*. Vol. 13, no. 4, pp. 232-240.
37. Iman, R.L. and W.J. Conover. 1980. "Small Sample Sensitivity Analysis Techniques for Computer Models, with an Application to Risk Assessment," *Communications in Statistics: Theory and Methods*. Vol. A9, no. 17, pp. 1749-1842.
38. McKay, M.D., R.J. Beckman, and W.J. Conover. 1979. "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*. Vol. 21, no. 2, pp. 239-245.
39. Li, G., C. Rosenthal, and H. Rabitz. 2001. "High-Dimensional Model Representations," *The Journal of Physical Chemistry*. Vol. 105, no. 33, pp. 7765-7777.
40. Rabitz, H. and O.F. Alis. 1999. "General Foundations of High-Dimensional Model Representations," *Journal of Mathematical Chemistry*. Vol. 25, no. 2-3, pp. 197-233.
41. Saltelli, A., S. Tarantola, and K.P.-S. Chan. 1999. "A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output," *Technometrics*. Vol. 41, no. 1, pp. 39-56.

42. Sobol', I.M. 1993. "Sensitivity Estimates for Nonlinear Mathematical Models," *Mathematical Modeling & Computational Experiment*. Vol. 1, no. 4, pp. 407-414.
43. Cukier, R.I., H.B. Levine, and K.E. Shuler. 1978. "Nonlinear Sensitivity Analysis of Multiparameter Model Systems," *Journal of Computational Physics*. Vol. 26, no. 1, pp. 1-42.
44. Saltelli, A., M. Ratto, S. Tarantola, and F. Campolongo. 2005. "Sensitivity Analysis for Chemical Models," *Chemical Reviews*. Vol. 105, no. 7, pp. 2811-2828.
45. Ionescu-Bujor, M. and D.G. Cacuci. 2004. "A Comparative Review of Sensitivity and Uncertainty Analysis of Large-Scale Systems--I: Deterministic Methods," *Nuclear Science and Engineering*. Vol. 147, no. 3, pp. 189-2003.
46. Cacuci, D.G. and M. Ionescu-Bujor. 2004. "A Comparative Review of Sensitivity and Uncertainty Analysis of Large-Scale Systems--II: Statistical Methods," *Nuclear Science and Engineering*. Vol. 147, no. 3, pp. 204-217.
47. Frey, H.C. and S.R. Patil. 2002. "Identification and Review of Sensitivity Analysis Methods," *Risk Analysis*. Vol. 22, no. 3, pp. 553-578.
48. Saltelli, A., K. Chan, and E.M. Scott (eds). 2000. *Sensitivity Analysis*. New York, NY: Wiley.
49. Hamby, D.M. 1994. "A Review of Techniques for Parameter Sensitivity Analysis of Environmental Models," *Environmental Monitoring and Assessment*. Vol. 32, no. 2, pp. 135-154.
50. Helton, J.C. 1993. "Uncertainty and Sensitivity Analysis Techniques for Use in Performance Assessment for Radioactive Waste Disposal," *Reliability Engineering and System Safety*. Vol. 42, no. 2-3, pp. 327-367.
51. Ronen, Y. 1988. *Uncertainty Analysis*. Boca Raton, FL: CRC Press, Inc.
52. Iman, R.L. and J.C. Helton. 1988. "An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models," *Risk Analysis*. Vol. 8, no. 1, pp. 71-90.
53. Vaughn, P., J.E. Bean, J.C. Helton, M.E. Lord, R.J. MacKinnon, and J.D. Schreiber. 2000. "Representation of Two-Phase Flow in the Vicinity of the Repository in the 1996 Performance Assessment for the Waste Isolation Pilot Plant," *Reliability Engineering and System Safety*. Vol. 69, no. 1-3, pp. 205-226.
54. Helton, J.C. and R.J. Breeding. 1993. "Calculation of Reactor Accident Safety Goals," *Reliability Engineering and System Safety*. Vol. 39, no. 2, pp. 129-158.

55. Breeding, R.J., J.C. Helton, E.D. Gorham, and F.T. Harper. 1992. "Summary Description of the Methods Used in the Probabilistic Risk Assessments for NUREG-1150," *Nuclear Engineering and Design*. Vol. 135, no. 1, pp. 1-27.
56. Helton, J.C. and M.G. Marietta. 2000. "Special Issue: The 1996 Performance Assessment for the Waste Isolation Pilot Plant," *Reliability Engineering and System Safety*. Vol. 69, no. 1-3, pp. 1-451.
57. Helton, J.C., D.R. Anderson, H.-N. Jow, M.G. Marietta, and G. Basabilvazo. 1999. "Performance Assessment in Support of the 1996 Compliance Certification Application for the Waste Isolation Pilot Plant," *Risk Analysis*. Vol. 19, no. 5, pp. 959 - 986.
58. Garthwaite, P.H., J.B. Kadane, and A. O'Hagan. 2005. "Statistical Methods for Eliciting Probability Distributions," *Journal of the American Statistical Association*. Vol. 100, no. 470, pp. 680-700.
59. Cooke, R.M. and L.H.J. Goossens. 2004. "Expert Judgement Elicitation for Risk Assessment of Critical Infrastructures," *Journal of Risk Research*. Vol. 7, no. 6, pp. 643-656.
60. Ayyub, B.M. 2001. *Elicitation of Expert Opinions for Uncertainty and Risks*, Boca Raton, FL: CRC Press.
61. McKay, M. and M. Meyer. 2000. "Critique of and Limitations on the use of Expert Judgements in Accident Consequence Uncertainty Analysis," *Radiation Protection Dosimetry*. Vol. 90, no. 3, pp. 325-330.
62. Goossens, L.H.J., F.T. Harper, B.C.P. Kraan, and H. Metivier. 2000. "Expert Judgement for a Probabilistic Accident Consequence Uncertainty Analysis," *Radiation Protection Dosimetry*. Vol. 90, no. 3, pp. 295-301.
63. Budnitz, R.J., G. Apostolakis, D.M. Boore, L.S. Cluff, K.J. Coppersmith, C.A. Cornell, and P.A. Morris. 1998. "Use of Technical Expert Panels: Applications to Probabilistic Seismic Hazard Analysis," *Risk Analysis*. Vol. 18, no. 4, pp. 463-469.
64. Goossens, L.H.J. and F.T. Harper. 1998. "Joint EC/USNRC Expert Judgement Driven Radiological Protection Uncertainty Analysis," *Journal of Radiological Protection*. Vol. 18, no. 4, pp. 249-264.
65. Siu, N.O. and D.L. Kelly. 1998. "Bayesian Parameter Estimation in Probabilistic Risk Assessment," *Reliability Engineering and System Safety*. Vol. 62, no. 1-2, pp. 89-116.
66. Evans, J.S., G.M. Gray, R.L. Sielken Jr., A.E. Smith, C. Valdez-Flores, and J.D. Graham. 1994. "Use of Probabilistic Expert Judgement in Uncertainty Analysis of Carcinogenic Potency," *Regulatory Toxicology and Pharmacology*. Vol. 20, no. 1, pt. 1, pp. 15-36.

67. Thorne, M.C. 1993. "The Use of Expert Opinion in Formulating Conceptual Models of Underground Disposal Systems and the Treatment of Associated Bias," *Reliability Engineering and System Safety*. Vol. 42, no. 2-3, pp. 161-180.
68. Chhibber, S., G. Apostolakis, and D. Okrent. 1992. "A Taxonomy of Issues Related to the Use of Expert Judgments in Probabilistic Safety Studies," *Reliability Engineering and System Safety*. Vol. 38, no. 1-2, pp. 27-45.
69. Otway, H. and D.V. Winterfeldt. 1992. "Expert Judgement in Risk Analysis and Management: Process, Context, and Pitfalls," *Risk Analysis*. Vol. 12, no. 1, pp. 83-93.
70. Thorne, M.C. and M.M.R. Williams. 1992. "A Review of Expert Judgement Techniques with Reference to Nuclear Safety," *Progress in Nuclear Safety*. Vol. 27, no. 2-3, pp. 83-254.
71. Cooke, R.M. 1991. *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford; New York: Oxford University Press.
72. Meyer, M.A. and J.M. Booker. 1991. *Eliciting and Analyzing Expert Judgment: A Practical Guide*. New York, NY: Academic Press.
73. Hora, S.C. and R.L. Iman. 1989. "Expert Opinion in Risk Analysis: The NUREG-1150 Methodology," *Nuclear Science and Engineering*. Vol. 102, no. 4, pp. 323-331.
74. Helton, J.C. 1997. "Uncertainty and Sensitivity Analysis in the Presence of Stochastic and Subjective Uncertainty," *Journal of Statistical Computation and Simulation*. Vol. 57, no. 1-4, pp. 3-76.
75. Helton, J.C. and D.E. Burmaster. 1996. "Guest Editorial: Treatment of Aleatory and Epistemic Uncertainty in Performance Assessments for Complex Systems," *Reliability Engineering and System Safety*. Vol. 54, no. 2-3, pp. 91-94.
76. Paté-Cornell, M.E. 1996. "Uncertainties in Risk Analysis: Six Levels of Treatment," *Reliability Engineering and System Safety*. Vol. 54, no. 2-3, pp. 95-111.
77. Winkler, R.L. 1996. "Uncertainty in Probabilistic Risk Assessment," *Reliability Engineering and System Safety*. Vol. 54, no. 2-3, pp. 127-132.
78. Hoffman, F.O. and J.S. Hammonds. 1994. "Propagation of Uncertainty in Risk Assessments: The Need to Distinguish Between Uncertainty Due to Lack of Knowledge and Uncertainty Due to Variability," *Risk Analysis*. Vol. 14, no. 5, pp. 707-712.

79. Helton, J.C. 1994. "Treatment of Uncertainty in Performance Assessments for Complex Systems," *Risk Analysis*. Vol. 14, no. 4, pp. 483-511.
80. Apostolakis, G. 1990. "The Concept of Probability in Safety Assessments of Technological Systems," *Science*. Vol. 250, no. 4986, pp. 1359-1364.
81. Haan, C.T. 1989. "Parametric Uncertainty in Hydrologic Modeling," *Transactions of the ASAE*. Vol. 32, no. 1, pp. 137-146.
82. Parry, G.W. and P.W. Winter. 1981. "Characterization and Evaluation of Uncertainty in Probabilistic Risk Analysis," *Nuclear Safety*. Vol. 22, no. 1, pp. 28-42.
83. Kaplan, S. and B.J. Garrick. 1981. "On the Quantitative Definition of Risk," *Risk Analysis*. Vol. 1, no. 1, pp. 11-27.
84. Kleijnen, J.P.C. and J.C. Helton. 1999. "Statistical Analyses of Scatterplots to Identify Important Factors in Large-Scale Simulations, 2: Robustness of Techniques," *Reliability Engineering and System Safety*. Vol. 65, no. 2, pp. 187-197.
85. Assunção, R. 1994. "Testing Spatial Randomness by Means of Angles," *Biometrics*. Vol. 50, pp. 531-537.
86. Ripley, B.D. 1987. "Spatial Point Pattern Analysis in Ecology," *Developments in Numerical Ecology. NATO ASI Series, Series G: Ecological Sciences*. Vol. 14. Eds. P. Legendre and L. Legendre. Berlin; New York: Springer-Verlag, pp. 407-430.
87. Zeng, G. and R.C. Dubes. 1985. "A Comparison of Tests for Randomness," *Pattern Recognition*. Vol. 18, no. 2, pp. 191-198.
88. Diggle, P.J. and T.F. Cox. 1983. "Some Distance-Based Tests of Independence for Sparsely-Sampled Multivariate Spatial Point Patterns," *International Statistical Review*. Vol. 51, no. 1, pp. 11-23.
89. Byth, K. 1982. "On Robust Distance-Based Intensity Estimators," *Biometrics*. Vol. 38, no. 1, pp. 127-135.
90. Byth, K. and B.D. Ripley. 1980. "On Sampling Spatial Patterns by Distance Methods," *Biometrics*. Vol. 36, no. 2, pp. 279-284.
91. Ripley, B.D. 1979. "Tests of Randomness for Spatial Point Patterns," *Journal of the Royal Statistical Society*. Vol. 41, no. 3, pp. 368-374.
92. Diggle, P.J. 1979. "Statistical Methods for Spatial Point Patterns in Ecology," *Spatial and Temporal Analysis in Ecology*. Ed. R.M. Cormack and J.K. Ord. Fairfield, MD: International Co-operative Pub. House. 95-150.

93. Diggle, P.J. 1979. "On Parameter Estimation and Goodness-of-Fit Testing for Spatial Point Patterns," *Biometrics*. Vol. 35, no. 1, pp. 87-101.
94. Besag, J. and P.J. Diggle. 1977. "Simple Monte Carlo Tests for Spatial Pattern," *Applied Statistics*. Vol. 26, no. 3, pp. 327-333.
95. Diggle, P.J., J. Besag, and J.T. Gleaves. 1976. "Statistical Analysis of Spatial Point Patterns by Means of Distance Methods," *Biometrics*. Vol. 32, pp. 659-667.
96. Cox, T.F. and T. Lewis. 1976. "A Conditioned Distance Ratio Method for Analyzing Spatial Patterns," *Biometrika*. Vol. 63, no. 3, pp. 483-491.
97. Holgate, P. 1972. "The Use of Distance Methods for the Analysis of Spatial Distribution of Points," *Stochastic Point Processes: Statistical Analysis, Theory, and Applications*. Eds. P.A.W. Lewis. New York, NY: Wiley-Interscience. 122-135.
98. Holgate, P. 1965. "Tests of Randomness Based on Distance Methods," *Biometrika*. Vol. 52, no. 3-4, pp. 345-353.
99. Garvey, J.E., E.A. Marschall, and R.A. Wright. 1998. "From Star Charts to Stoneflies: Detecting Relationships in Continuous Bivariate Data," *Ecology*. Vol. 79, no. 2, pp. 442-447.
100. Fasano, G. and A. Franceschini. 1987. "A Multidimensional Version of the Kolmogorov-Smirnov Test," *Monthly Notices of the Royal Astronomical Society*. Vol. 225, no. 1, pp. 155-170.
101. Gosset, E. 1987. "A 3-Dimensional Extended Kolmogorov-Smirnov Test as a Useful Tool in Astronomy," *Astronomy and Astrophysics*. Vol. 188, no. 1, pp. 258-264.
102. Peacock, J.A. 1983. "Two-Dimensional Goodness-Of-Fit Testing in Astronomy," *Monthly Notices of the Royal Astronomical Society*. Vol. 202, no. 2, pp. 615-627.
103. Helton, J.C., F.J. Davis, and J.D. Johnson. 2005. "A Comparison of Uncertainty and Sensitivity Analysis Results Obtained with Random and Latin Hypercube Sampling," *Reliability Engineering and System Safety*. Vol. 89, no. 3, pp. 305-330.
104. Hora, S.C. and J.C. Helton. 2003. "A Distribution-Free Test for the Relationship Between Model Input and Output when Using Latin Hypercube Sampling," *Reliability Engineering and System Safety*. Vol. 79, no. 3, pp. 333-339.

105. Mokhtari, A., H.C. Frey, and L.A. Jaykus. 2006. "Application of Classification and Regression Trees for Sensitivity Analysis of the Escherichia coli O157: H7 Food Safety Process Risk Model," *Journal of Food Protection*. Vol. 69, no. 3, pp. 609-618.
106. Mishra, S., N.E. Deeds, and B.S. RamaRao. 2003. "Application of Classification Trees in the Sensitivity Analysis of Probabilistic Model Results," *Reliability Engineering and System Safety*. Vol. 79, no. 2, pp. 123-129.
107. Storlie, C.B. and J.C. Helton. To Appear. "Multiple Predictor Smoothing Methods for Sensitivity Analysis: Example Results," *Reliability Engineering and System Safety*.
108. Maindonald, J. and J. Braun. 2003. *Data Analysis and Graphics Using R*. Cambridge: Cambridge University Press.
109. Insightful Corporation. 2001. *S-PLUS 6 for Windows Guide to Statistics, Volume 1*, Seattle, WA: Insightful Corporation.
110. Myers, R.H. 1990. *Classical and Modern Regression with Applications*. 2nd ed. Boston: Duxbury Press.
111. Draper, N.R. and H. Smith. 1981. *Applied Regression Analysis*. 2nd ed. New York, NY: John Wiley & Sons.
112. Daniel, C., F.S. Wood, and J.W. Gorman. 1980. *Fitting Equations to Data: Computer Analysis of Multifactor Data*. 2nd ed. New York, NY: John Wiley & Sons.
113. Seber, G.A. 1977. *Linear Regression Analysis*. New York, NY: John Wiley & Sons.
114. Neter, J. and W. Wasserman. 1974. *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs*. Homewood, IL: Richard D. Irwin.
115. Iman, R.L. and W.J. Conover. 1979. "The Use of the Rank Transform in Regression," *Technometrics*. Vol. 21, no. 4, pp. 499-509.
116. Helton, J.C. 1999. "Uncertainty and Sensitivity Analysis in Performance Assessment for the Waste Isolation Pilot Plant," *Computer Physics Communications*. Vol. 117, no. 1-2, pp. 156-180.
117. Helton, J.C., J.D. Johnson, M.D. McKay, A.W. Shiver, and J.L. Sprung. 1995. "Robustness of an Uncertainty and Sensitivity Analysis of Early Exposure Results with the MACCS Reactor Accident Consequence Model," *Reliability Engineering and System Safety*. Vol. 48, no. 2, pp. 129-148.
118. Iman, R.L. and J.C. Helton. 1991. "The Repeatability of Uncertainty and Sensitivity Analyses for Complex Probabilistic Risk Assessments," *Risk Analysis*. Vol. 11, no. 4, pp. 591-606.

119. Myers, R.H. and D.C. Montgomery. 1995. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. New York, NY: Wiley.
120. Huet, S., A. Bouvier, M.-A. Gruet, and E. Jolivet. 2003. *Statistical Tools for Nonlinear Regression: A Practical Guide with S-PLUS and R Examples*. New York, NY: Springer-Verlag.
121. Hastie, T.J. and R.J. Tibshirani. 1990. *Generalized Additive Models*. London: Chapman & Hall.
122. Simonoff, J.S. 1996. *Smoothing Methods in Statistics*. New York, NY: Springer-Verlag.
123. Bowman, A.W. and A. Azzalini. 1997. *Applied Smoothing Techniques for Data Analysis*. Oxford: Clarendon.
124. Ruppert, D., M.P. Ward, and R.J. Carroll. 2003. *Semiparametric Regression*. Cambridge: Cambridge University Press.
125. Cleveland, W.S. 1979. "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*. Vol. 14, no. 368, pp. 829-836.
126. Schimek, M.G. 2000. *Smoothing and Regression: Approaches, Computation, and Application*. New York, NY: Wiley.
127. Chambers, J.M. and T.J. Hastie. 1992. *Statistical Models in S*. Pacific Grove, CA: Wadsworth & Brooks.
128. Hocking, R.R. 1996. *Methods and Applications of Linear Models*. New York, NY: Wiley.
129. Christensen, R. 1996. *Plane Answers to Complex Questions*. New York, NY: Springer-Verlag.
130. Allen, D.M. 1971. *The Prediction Sum of Squares as a Criterion for Selecting Predictor Variables*, 23. Lexington: University of Kentucky, Department of Statistics.
131. Friedman, J.H. and W. Stuetzle. 1981. "Projection Pursuit Regression," *Journal of the American Statistical Association*. Vol. 76, no. 376, pp. 817-823.
132. Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone. 1984. *Classification and Regression Trees*. Belmont, CA: Wadsworth Intl.
133. Rousseeuw, P.J. and A.M. Leroy. 1987. *Robust Regression and Outlier Detection*. New York, NY: Wiley.
134. Hastie, T., R. Tibshirani, and J. Friedman. 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer.

135. Press, W.H., S.A. Teukolsky, W.T. Vetterlings, and B.P. Flannery. 1992. *Numerical Recipes: the Art of Scientific Computing (FORTRAN Version)*. 2nd ed. Cambridge: Cambridge University Press.
136. Hogg, R.V. and A.T. Craig. 1978. *Introduction to Mathematical Statistics, 4th Ed.* New York, NY: Macmillan.
137. Breeding, R.J., J.C. Helton, W.B. Murfin, L.N. Smith, J.D. Johnson, H.-N. Jow, and A.W. Shiver. 1992. "The NUREG-1150 Probabilistic Risk Assessment for the Surry Nuclear Power Station," *Nuclear Engineering and Design*. Vol. 135, no. 1, pp. 29-59.
138. Payne, A.C., Jr., R.J. Breeding, J.C. Helton, L.N. Smith, J.D. Johnson, H.-N. Jow, and A.W. Shiver. 1992. "The NUREG-1150 Probabilistic Risk Assessment for the Peach Bottom Atomic Power Station," *Nuclear Engineering and Design*. Vol. 135, no. 1, pp. 61-94.
139. Gregory, J.J., R.J. Breeding, J.C. Helton, W.B. Murfin, S.J. Higgins, and A.W. Shiver. 1992. "The NUREG-1150 Probabilistic Risk Assessment for the Sequoyah Nuclear Plant," *Nuclear Engineering and Design*. Vol. 135, no. 1, pp. 92-115.
140. Brown, T.D., R.J. Breeding, J.C. Helton, H.-N. Jow, S.J. Higgins, and A.W. Shiver. 1992. "The NUREG-1150 Probabilistic Risk Assessment for the Grand Gulf Nuclear Station," *Nuclear Engineering and Design*. Vol. 135, no. 1, pp. 117-137.
141. Helton, J.C., M.-A. Martell, and M.S. Tierney. 2000. "Characterization of Subjective Uncertainty in the 1996 Performance Assessment for the Waste Isolation Pilot Plant," *Reliability Engineering and System Safety*. Vol. 69, no. 1-3, pp. 191-204.
142. CRWMS M&O (Civilian Radioactive Waste Management System Management and Operating Contractor). 2000. *Total System Performance Assessment for the Site Recommendation*, TDR-WIS-PA-000001 REV 00. Las Vegas, NV: CRWMS M&O.
143. CRWMS M&O (Civilian Radioactive Waste Management System Management and Operating Contractor). 2000. *Total System Performance Assessment (TSPA) Model for Site Recommendation*, MDL-WIS-PA-000002 REV 00. Las Vegas, NV: CRWMS M&O.

Figure Captions

- Fig. 1. Linear regression on results generated in a sensitivity analysis of a two-phase fluid flow model.
- Fig. 2. Rank regression on an example monotonic relationship.
- Fig. 3. Rank regression on a nonlinear and nonmonotonic relationship generated in a sensitivity analysis of a two-phase fluid flow model.
- Fig. 4. Quadratic regression on a nonlinear and nonmonotonic relationship generated in a sensitivity analysis of a two-phase fluid flow model.
- Fig. 5. Running means with $r = 20$ on results generated in a sensitivity analysis of a two-phase fluid flow model.
- Fig. 6. Locally weighted means with kernel function $k(z; h)$ in Eq. (3.5) and bandwidth $h = 0.6$ on results generated in a sensitivity analysis of a two-phase fluid flow model.
- Fig. 7. Analysis with LOESS for kernel function $k(z; h)$ in Eq. (3.11) and $r = 60$ (i.e., a span of 0.20) on results generated in a sensitivity analysis of a two-phase fluid flow model.
- Fig. 8. Analysis with smoothing spline with $a = x_{(1)}$, $b = x_{(nS)}$ and $df = 8$ (see Eq. (3.13)) on results generated in a sensitivity analysis of a two-phase fluid flow model.
- Fig. 9. Example of LOESS surface constructed for $y = f(x_1, x_2) = (1/2\pi) \exp\{-[(x_1 - 5)^2 + (x_2 - 5)^2]/2\}$; see Eq. (3.27).
- Fig. 10. Example of additive model surface constructed for $y = f(x_1, x_2) = \sin(x_1) + (x_2 - 5)^2$; see Eq. (3.35).
- Fig. 11. Recursive partitioning regression on results generated in a sensitivity analysis of a two-phase fluid flow model: (a) Individual regression lines generated with traditional least squares regression, and (b) Individual regression lines generated with robust regression in which the sum of squares is minimized over the middle two quartiles of the deviations from the regression line.
- Fig. 12. Recursive partitioning regression on results generated in a sensitivity analysis of a two-phase fluid flow model with individual regression lines constrained to meet continuously: (a) Individual regression lines generated with traditional least squares regression, and (b) Individual regression lines generated with robust regression in which the sum of squares is minimized over the middle two quartiles of the deviations from the regression line.
- Fig. 13. Recursive partitioning regression constructed for $y = f(x_1, x_2) = \sin x_1 + (x_2 - 5)^2$ with individual regression surfaces constrained to meet continuously; see Eq. (3.35).

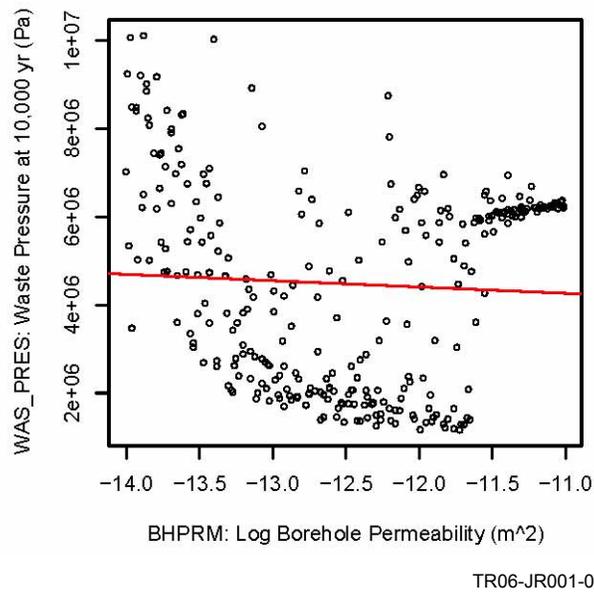
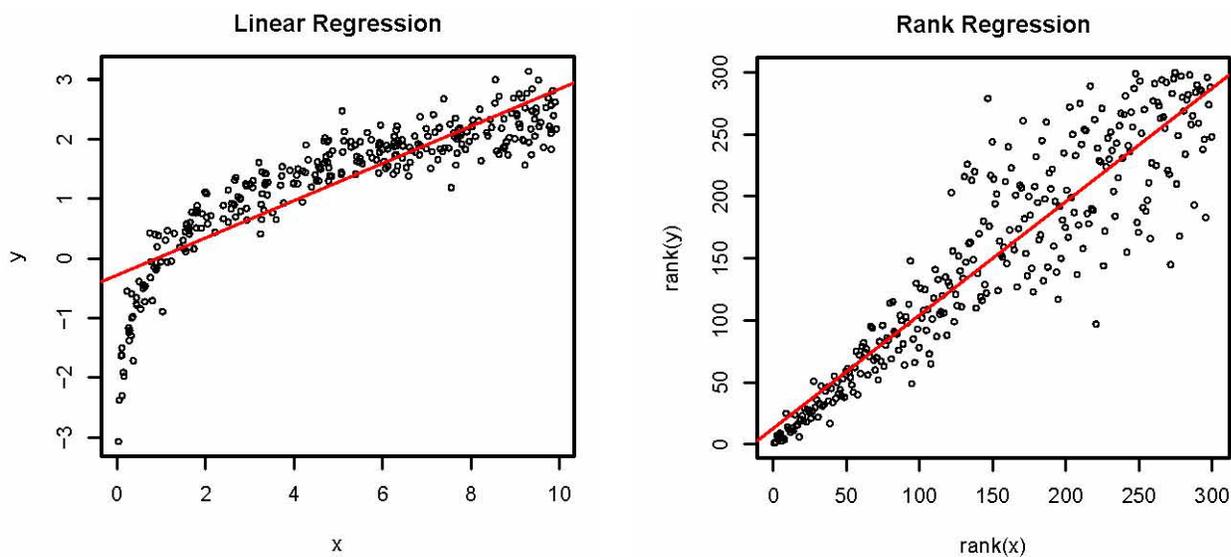
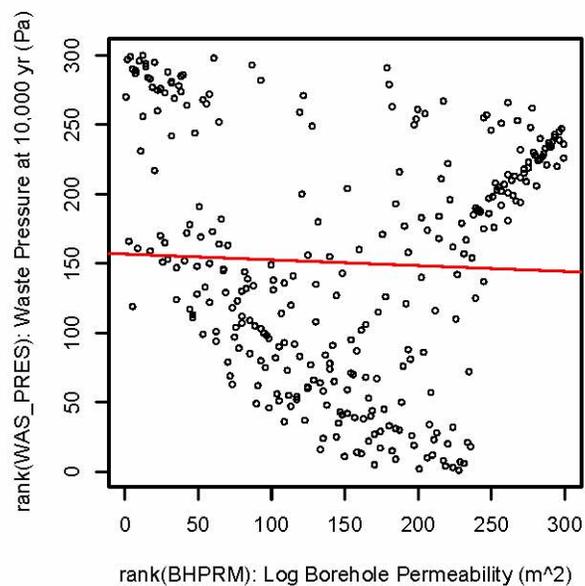


Fig. 1. Linear regression on results generated in a sensitivity analysis of a two-phase fluid flow model.



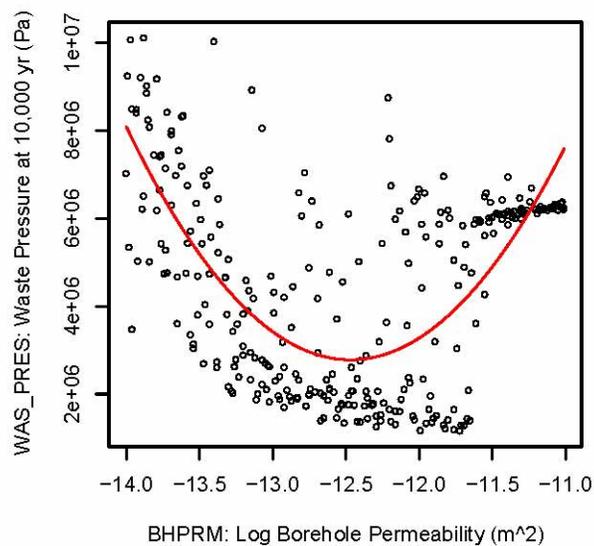
TR06-JR002-0

Fig. 2. Rank regression on an example monotonic relationship.



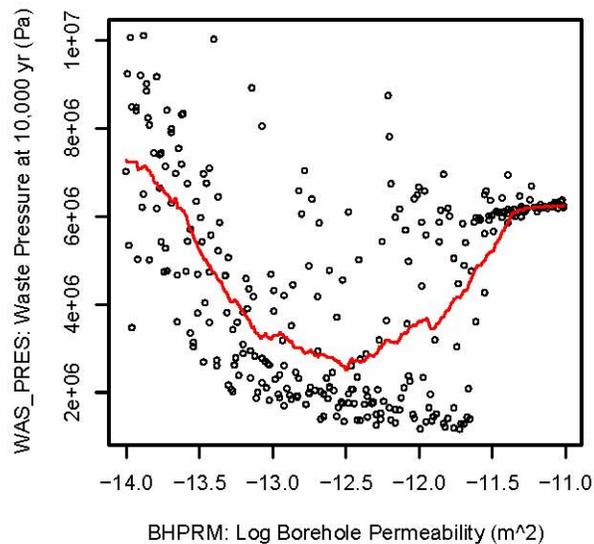
TR06-JR003-0

Fig. 3. Rank regression on a nonlinear and nonmonotonic relationship generated in a sensitivity analysis of a two-phase fluid flow model.



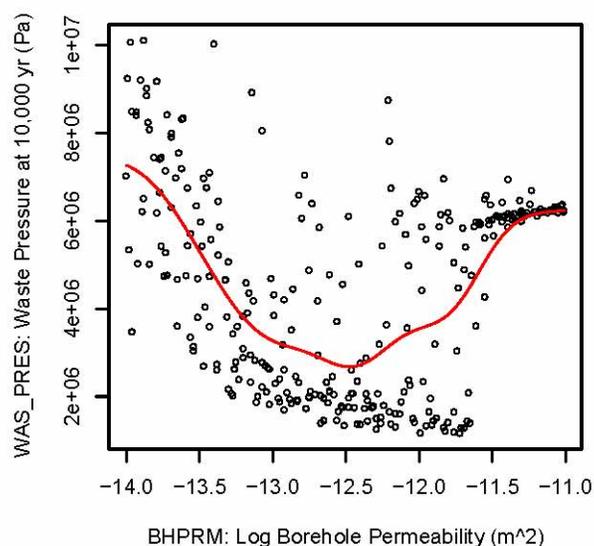
TR06-JR004-0

Fig. 4. Quadratic regression on a nonlinear and nonmonotonic relationship generated in a sensitivity analysis of a two-phase fluid flow model.



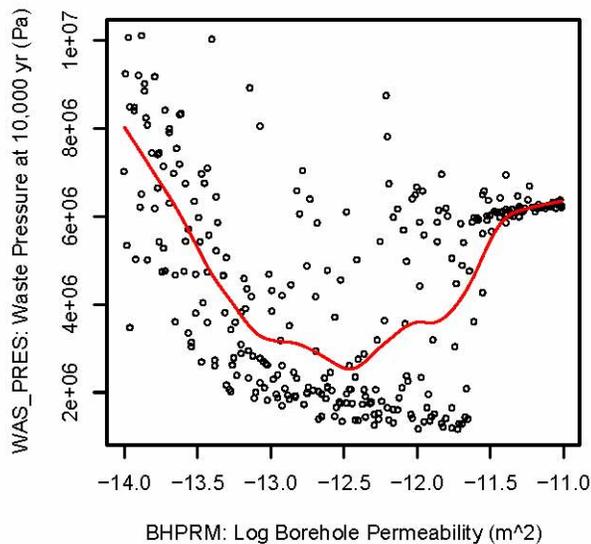
TR06-JR005-0

Fig. 5. Running means with $r = 20$ on results generated in a sensitivity analysis of a two-phase fluid flow model.



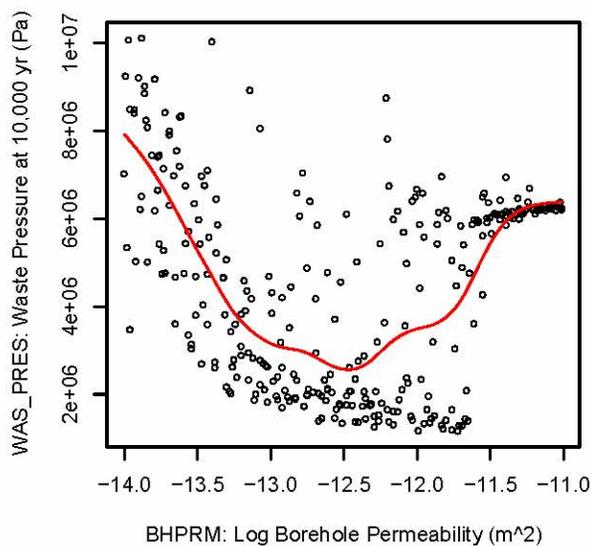
TR06-JR006-0

Fig. 6. Locally weighted means with kernel function $k(z; h)$ in Eq. (3.5) and bandwidth $h = 0.6$ on results generated in a sensitivity analysis of a two-phase fluid flow model.



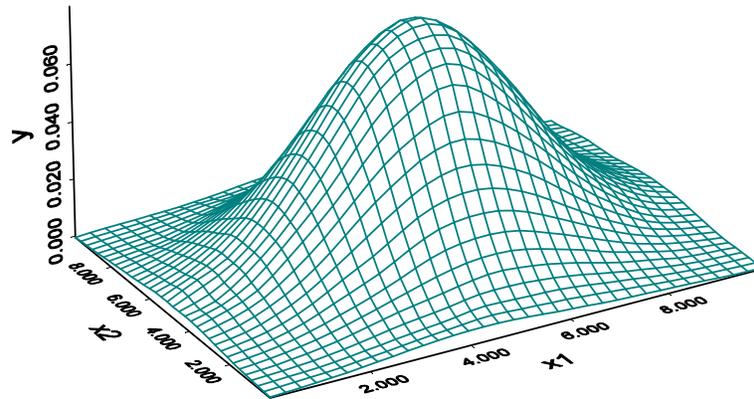
TR06-JR007-0

Fig. 7. Analysis with LOESS for kernel function $k(z; h)$ in Eq. (3.11) and $r = 60$ (i.e., a span of 0.20) on results generated in a sensitivity analysis of a two-phase fluid flow model.



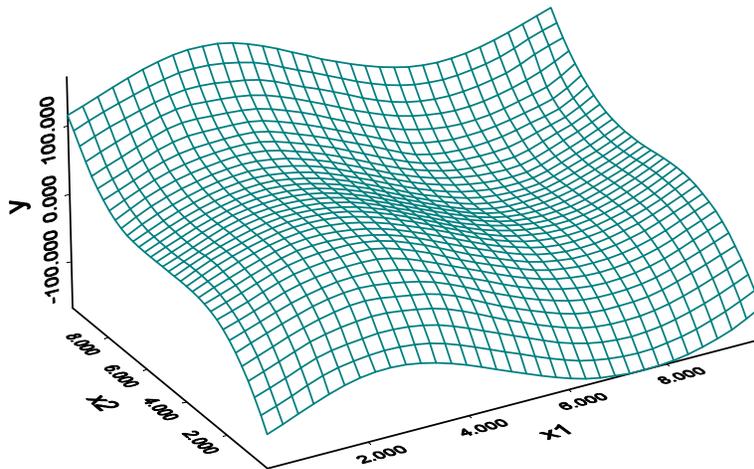
TR06-JR008-0

Fig. 8. Analysis with smoothing spline with $a = x_{(1)}$, $b = x_{(n)}$ and $df = 8$ (see Eq. (3.13)) on results generated in a sensitivity analysis of a two-phase fluid flow model.



TR06-JR009-0

Fig. 9. Example of LOESS surface constructed for $y = f(x_1, x_2) = (1/2\pi) \exp\{-(x_1 - 5)^2 + (x_2 - 5)^2/2\}$; see Eq. (3.27).



TR06-JR010-0

Fig. 10. Example of additive model surface constructed for $y = f(x_1, x_2) = \sin(x_1) + (x_2 - 5)^2$; see Eq. (3.35).

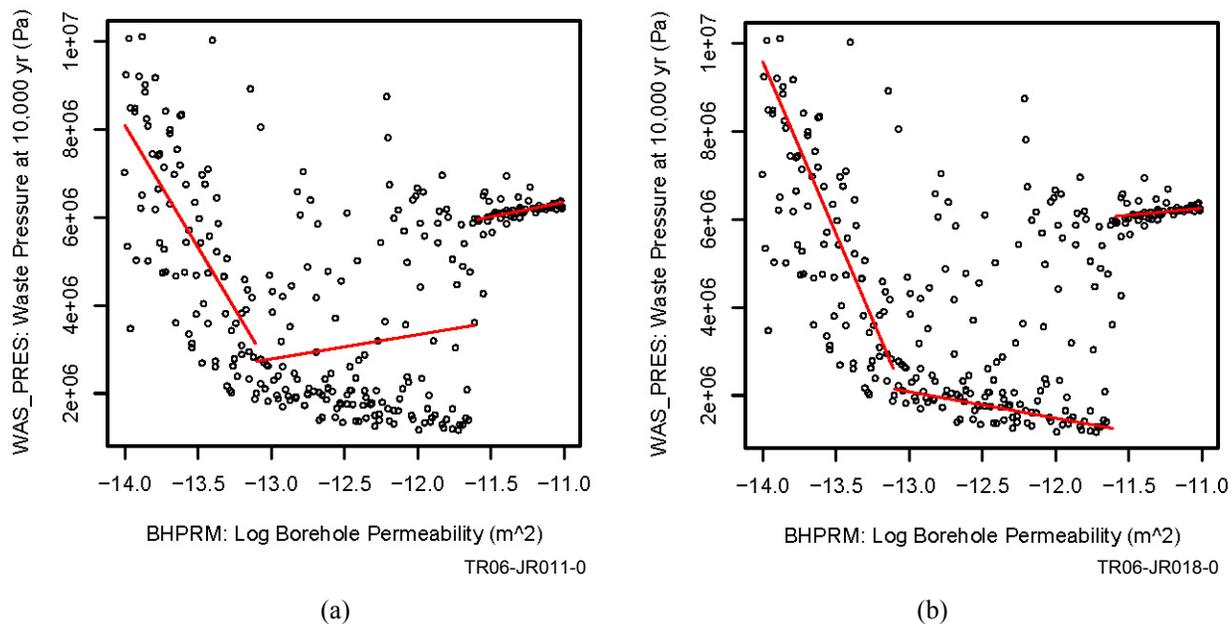


Fig. 11. Recursive partitioning regression on results generated in a sensitivity analysis of a two-phase fluid flow model: (a) Individual regression lines generated with traditional least squares regression, and (b) Individual regression lines generated with robust regression in which the sum of squares is minimized over the middle two quartiles of the deviations from the regression line.

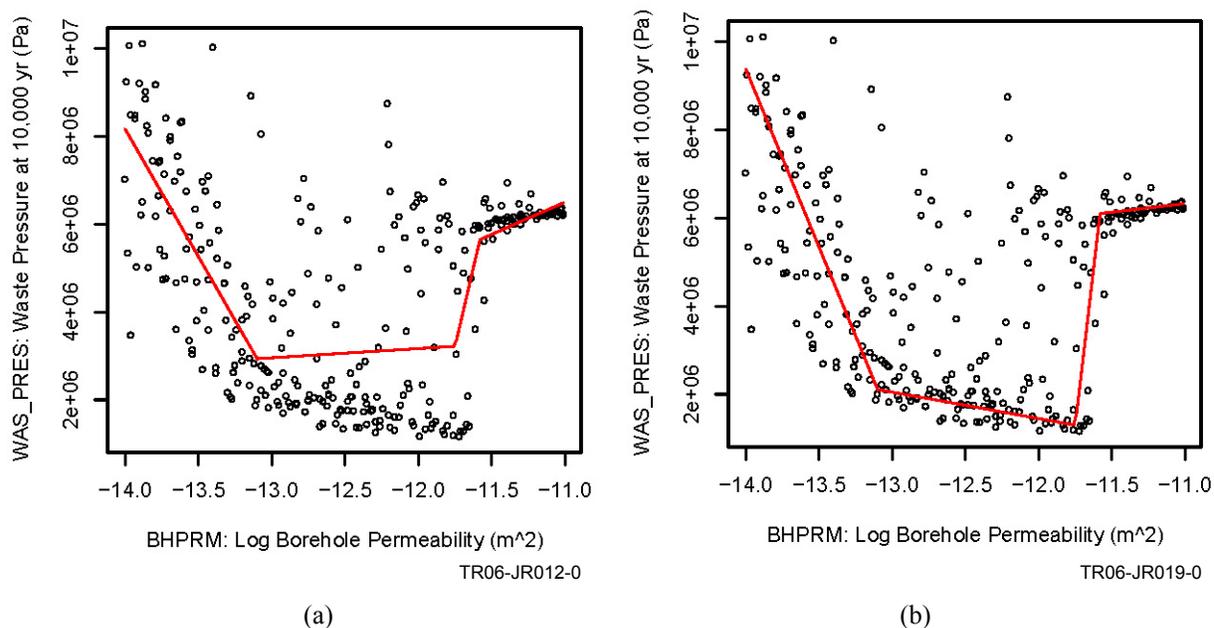
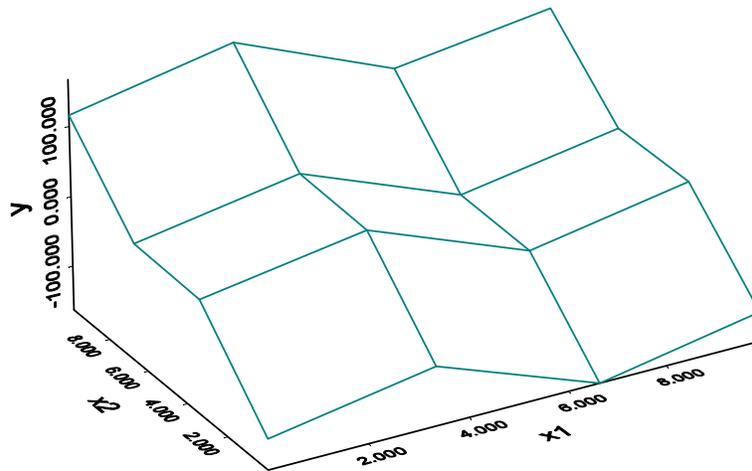


Fig. 12. Recursive partitioning regression on results generated in a sensitivity analysis of a two-phase fluid flow model with individual regression lines constrained to meet continuously: (a) Individual regression lines generated with traditional least squares regression, and (b) Individual regression lines generated with robust regression in which the sum of squares is minimized over the middle two quartiles of the deviations from the regression line.



TR06-JR013-0

Fig. 13. Recursive partitioning regression constructed for $y = f(x_1, x_2) = \sin x_1 + (x_2 - 5)^2$ with individual regression surfaces constrained to meet continuously; see Eq. (3.35).

Table 1. Forward Stepwise Variable Selection Algorithm for Sensitivity Analysis with LIN_REG, RANK_REG and QUAD_REG

Step 1. Estimate $y \cong f_{1j}(x_j)$ with regression procedure in use (i.e., LIN_REG, RANK_REG or QUAD_REG) for $j = 1, 2, \dots, nX$. For each of the models $y \cong f_{1j}(x_j)$, determine (i) degrees of freedom df_j (i.e., $df_j = 2$ for LIN_REG and RANK_REG and $df_j = 3$ for QUAD_REG), (ii) F -statistic F_j for comparison against mean only model, and (iii) resultant p -value p_j (see Eq. (3.54)). Variable \tilde{x}_1 with smallest p -value is selected as most important variable at Step 1; corresponding model and degrees of freedom are represented by $y \cong f_1(\tilde{x}_1)$ and \tilde{df}_1 , respectively. The process terminates with no variable selected if all p_j are greater than a specified cutoff (e.g., $\alpha = 0.02$).

Step 2. Estimate $y \cong f_{2j}(\tilde{x}_1, x_j)$ with regression procedure in use (i.e., LIN_REG, RANK_REG or QUAD_REG) for $j = 1, 2, \dots, nX$ and $x_j \neq \tilde{x}_1$. For each of the models $y \cong f_{2j}(\tilde{x}_1, x_j)$, determine (i) degrees of freedom df_j (i.e., $df_j = 3$ for LIN_REG and RANK_REG and $df_j = 5$ for QUAD_REG), (ii) F -statistic F_j for comparison against model $y \cong f_1(\tilde{x}_1)$ selected at Step 1, and (iii) resultant p -value p_j (see Eq. (3.54)). Variable \tilde{x}_2 with smallest p -value is selected as most important variable at Step 2; corresponding model and degrees of freedom are represented by $y \cong f_2(\tilde{x}_1, \tilde{x}_2)$ and \tilde{df}_2 , respectively. The process terminates with no variable selected at Step 2 if all p_j are greater than a specified cutoff (e.g., $\alpha = 0.02$).

Step 3. Estimate $y \cong f_{3j}(\tilde{x}_1, \tilde{x}_2, x_j)$ with regression procedure in use (i.e., LIN_REG, RANK_REG or QUAD_REG) for $j = 1, 2, \dots, nX$, $x_j \neq \tilde{x}_1$ and $x_j \neq \tilde{x}_2$. Continue as in Step 2.

...

Step N. Terminate process when no variable satisfies specified cutoff.

Table 2. Forward Stepwise Variable Selection Algorithm for Sensitivity Analysis with LOESS

Notation: Variables s_k , $k = 1, 2, \dots, 5$, represent the candidate spans 10, 0.7, 0.3, 0.1, 0.07 and 0.05 described in Sect. 4.3 with $s_1 = 10$ designating the use of linear regression.

Step 1. Estimate $y \cong f_{1jk}(x_j | s_k)$ with LOESS for $j = 1, 2, \dots, nX$ and $k = 1, 2, \dots, 6$. For each x_j , identify span \tilde{s}_j that results in the smallest value for the adjusted PRESS statistic PRS_A (see Eqs. (3.22) and (3.25)) for the models $y \cong f_{1jk}(x_j | s_k)$, $k = 1, 2, \dots, 6$. For each of the models $y \cong f_{1j}(x_j | \tilde{s}_j)$, determine (i) degrees of freedom df_j (i.e., $df_j = \text{tr}(\mathbf{S}_{F_j})$, where \mathbf{S}_{F_j} is the smoother matrix associated with the selected span \tilde{s}_j for x_j (see Eq. (3.54) and associated discussion in Sect. 3.4), (ii) F -statistic F_j for comparison against mean only model, and (iii) resultant p -value p_j (see Eq. (3.54) and associated discussion in Sect. 4.3). Variable \tilde{x}_1 with smallest p -value is selected as most important variable at Step 1; corresponding model and degrees of freedom are represented by $y \cong f_1(\tilde{x}_1 | \tilde{s}_1)$ and \tilde{df}_1 , respectively. The process terminates with no variable selected if all p_j are greater than a specified cutoff (e.g., $\alpha = 0.02$).

Step 2. Estimate $y \cong f_{2j}(\tilde{x}_1, x_j | s_k)$ with LOESS for $j = 1, 2, \dots, nX$, $x_j \neq \tilde{x}_1$, and $k = 1, 2, \dots, 6$. For each x_j , identify span \tilde{s}_j that results in smallest value for the adjusted PRESS statistic PRS_A (see Eqs. (3.22) and (3.25)) for the models $y \cong f_{2jk}(\tilde{x}_1, x_j | s_k)$, $k = 1, 2, \dots, 6$. For each of the models $y \cong f_{2j}(\tilde{x}_1, x_j | \tilde{s}_j)$, determine (i) degrees of freedom df_j , (ii) F -statistic F_j for comparison against model $y \cong f_1(\tilde{x}_1 | \tilde{s}_1)$ selected in Step 1, and (iii) resultant p -value p_j (see Eq. (3.54) and associated discussion in Sect. 4.3). Variable \tilde{x}_2 with smallest p -value is selected as most important variable at Step 2; corresponding model and degrees of freedom are represented by $y \cong f_2(\tilde{x}_1, \tilde{x}_2 | \tilde{s}_2)$ and \tilde{df}_2 , respectively. The process terminates with no variable selected at Step 2 if all p_j are greater than a specified cutoff (e.g., $\alpha = 0.02$).

Step 3. Estimate $y \cong f_{2jk}(\tilde{x}_1, \tilde{x}_2, x_j | s_k)$ with LOESS for $j = 1, 2, \dots, nX$, $x_j \neq \tilde{x}_1$, $x_j \neq \tilde{x}_2$, and $k = 1, 2, \dots, 6$. Continue as in Step 2.

...

Step N. Terminate process when no variable satisfies specified cutoff.

Table 3. Forward Stepwise Variable Selection Algorithm for Sensitivity Analysis with GAMs

Notation. Variables $\lambda_k, k = 1, 2, \dots, 6$, represent candidate smoothing parameters used in smoothing splines (see Eq. (3.13)) in the sequential construction of GAMs, with λ_k resulting in a smoothing process with approximately δ_k degrees of freedom. Specifically, $\lambda_1 \sim \delta_1 = 1$, $\lambda_2 \sim \delta_2 = 2$, $\lambda_3 \sim \delta_3 = 4$, $\lambda_4 \sim \delta_4 = 7$, $\lambda_5 \sim \delta_5 = 10$, and $\lambda_6 \sim \delta_6 = 15$. The actual value used for λ_k in Eq. (3.13) is determined from the specified value for δ_k (see Sect. 3.5, Ref. [121] and Sect. 7.4.1, Ref. [127]).

Step 1. Estimate $y \cong f_{1jk}(x_j | \lambda_k)$ with a smoothing spline on $(x_{ij}, y_i), i = 1, 2, \dots, nS$, for $j = 1, 2, \dots, nX$ and $k = 1, 2, \dots, 6$. For each x_j , select model $f_{1j}(x_j | \tilde{\lambda}_j)$ with smoothing parameter $\tilde{\lambda}_j$ from the models $y \cong f_{1jk}(x_j | \lambda_k), k = 1, 2, \dots, 5$, that results in the smallest value for the adjusted PRESS statistic PRS_A (see Eqs. (3.22) and (3.25) and discussion in Sect. 4.4). For each of the models $y \cong f_{1j}(x_j | \tilde{\lambda}_j)$, determine (i) degrees of freedom (i.e., $df_j = \tilde{\delta}_j$, with $\tilde{\lambda}_j \sim \tilde{\delta}_j$; see discussion in Sect. 4.4), (ii) F -statistic F_j for comparison against mean only model, and (iii) resultant p -value p_j (see Eq. (3.54)). Variable \tilde{x}_1 with smallest p -value is selected as the most important variable at Step 1; corresponding model, smoothing parameter, and degrees of freedom are represented by $y \cong f_1(\tilde{x}_1 | \tilde{\lambda}_1)$, $\tilde{\lambda}_1$ and \tilde{df}_1 , respectively. The process terminates with no variable selected if all p_j are greater than a specified cutoff (e.g., $\alpha = 0.02$).

Step 2. Estimate $y \cong f_{2jk}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \lambda_k)$ through a sequence of smoothing operations for $j = 1, 2, \dots, nX$, $x_j \neq \tilde{x}_1$ and $k = 1, 2, \dots, 6$; see Step 2' for details. For each x_j , select model $f_{2j}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \tilde{\lambda}_j)$ with smoothing parameter $\tilde{\lambda}_j$ from the models $y \cong f_{2jk}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \lambda_k), k = 1, 2, \dots, 6$, that results in the smallest value for the adjusted PRESS statistic PRS_A (see Eqs. (3.22) and (3.25) and discussion in Sect. 4.4). For each of the models $y \cong f_{2j}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \tilde{\lambda}_j)$, determine (i) degrees of freedom (i.e., $df_j = \tilde{\delta}_1 + \tilde{\delta}_j$ with $\tilde{\lambda}_1 \sim \tilde{\delta}_1$ and $\tilde{\lambda}_j \sim \tilde{\delta}_j$; see discussion in Sect. 4.4), (ii) F -statistic F_j for comparison against model $y \cong f_1(\tilde{x}_1 | \tilde{\lambda}_1)$ constructed in Step 1, and (iii) resultant p -value p_j (see Eq. (3.54)). Variable \tilde{x}_2 with smallest p -value is selected as most important variable at Step 2; corresponding model, smoothing parameter and degrees of freedom are represented by $y \cong f_2(\tilde{x}_1, \tilde{x}_2 | \tilde{\lambda}_1, \tilde{\lambda}_2) = f_{21}(\tilde{x}_1 | \tilde{\lambda}_1) + f_{22}(\tilde{x}_2 | \tilde{\lambda}_2)$, $\tilde{\lambda}_2$ and \tilde{df}_2 , respectively, where $f_{21}(\tilde{x}_1 | \tilde{\lambda}_1)$ is a smoothed estimate of y as function of \tilde{x}_1 (i.e., $f_{21}(\tilde{x}_1 | \tilde{\lambda}_1)$ corresponds to $F_{jk,2l}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \lambda_j)$ in Step 2.4' for the selected values for \tilde{x}_2 and $\tilde{\lambda}_2$) and $f_{22}(\tilde{x}_2 | \tilde{\lambda}_2)$ is a smoothed estimate of y as a function of \tilde{x}_2 (i.e., $f_{22}(\tilde{x}_2 | \tilde{\lambda}_2)$ corresponds to $F_{jk,2l+1}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \lambda_j)$ in Step 2.4' for the selected values of \tilde{x}_2 and $\tilde{\lambda}_2$). The process terminates with no variable selected at Step 2 if all p_j are greater than a specified cutoff (e.g., $\alpha = 0.02$).

Step 2'. Procedure for obtaining smoothed model $f_{2jk}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \lambda_k)$ on the basis of a relative error criterion for variable x_j , $x_j \neq \tilde{x}_1$, and smoothing parameter λ_k in Step 2.

Step 2.1'. Estimate $F_{jk1}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \lambda_k) \cong y - f_1(\tilde{x}_1 | \tilde{\lambda}_1)$ by smoothing on $(x_{ij}, y_i - f_1(\tilde{x}_{i1} | \tilde{\lambda}_1)), i = 1, 2, \dots, nS$, with a smoothing spline and smoothing parameter λ_k . Result is estimate $y_i \cong G_{jk1}(\tilde{x}_{i1}, x_{ij} | \tilde{\lambda}_1, \lambda_k) = f_1(\tilde{x}_{i1} | \tilde{\lambda}_1) + F_{jk1}(\tilde{x}_{i1}, x_{ij} | \tilde{\lambda}_1, \lambda_k)$.

Table 3. Forward Stepwise Variable Selection Algorithm for Sensitivity Analysis with GAMs (Continued)

Step 2.2'. Estimate $F_{jk2}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \lambda_k) \equiv y - F_{jk1}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \lambda_k)$ by smoothing on $(\tilde{x}_{i1}, y_i - F_{jk1}(\tilde{x}_{i1}, x_{ij} | \tilde{\lambda}_1, \lambda_k))$, $i = 1, 2, \dots, nS$, with a smoothing spline and smoothing parameter $\tilde{\lambda}_1$. Then, estimate $F_{jk3}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \lambda_k) \equiv y - F_{jk2}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \lambda_k)$ by smoothing on $(x_{ij}, y_i - F_{jk2}(\tilde{x}_1, x_{ij} | \tilde{\lambda}_1, \lambda_k))$, $i = 1, 2, \dots, nS$, with a smoothing spline and smoothing parameter λ_k . Result is estimate $y_i \equiv G_{jk2}(\tilde{x}_{i1}, x_{ij} | \tilde{\lambda}_1, \lambda_k) = F_{jk2}(\tilde{x}_{i1}, x_{ij} | \tilde{\lambda}_1, \lambda_k) + F_{jk3}(\tilde{x}_{i1}, x_{ij} | \tilde{\lambda}_1, \lambda_k)$.

Step 2.3'. Similar to Step 2.2'. First, estimate $F_{jk4}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \lambda_k) \equiv y - F_{jk3}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \lambda_k)$ by smoothing on $(\tilde{x}_{i1}, y_i - F_{jk3}(\tilde{x}_{i1}, x_{ij} | \tilde{\lambda}_1, \lambda_k))$, $i = 1, 2, \dots, nS$, with a smoothing spline and associated smoothing parameter $\tilde{\lambda}_1$. Then, estimate $F_{jk5}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \lambda_k) \equiv y_2 - F_{jk4}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \lambda_k)$ by smoothing on $(x_{ij}, y_i - F_{jk4}(\tilde{x}_{i1}, x_{ij} | \tilde{\lambda}_1, \lambda_k))$, $i = 1, 2, \dots, nS$, with a smoothing spline and associated smoothing parameter λ_k . Result is estimate $y_i \equiv G_{jk3}(\tilde{x}_{i1}, x_{ij} | \tilde{\lambda}_1, \lambda_k) = F_{jk4}(\tilde{x}_{i1}, x_{ij} | \tilde{\lambda}_1, \lambda_k) + F_{jk5}(\tilde{x}_{i1}, x_{ij} | \tilde{\lambda}_1, \lambda_k)$.

Step 2.4'. Continue as in Step 2.3' until the relative error criterion $\|\mathbf{G}_{jk,l+1} - \mathbf{G}_{jkl}\| \leq \text{rerr} \|\mathbf{G}_{jkl}\|$ is satisfied for $\mathbf{G}_{jkr} = [G_{jkr}(\tilde{x}_{11}, x_{1j} | \tilde{\lambda}_1, \lambda_k), G_{jkr}(\tilde{x}_{21}, x_{2j} | \tilde{\lambda}_1, \lambda_k), \dots, G_{jkr}(\tilde{x}_{nS,1}, x_{nS,j} | \tilde{\lambda}_1, \lambda_k)]$, $r = l, l+1$, and $\text{rerr} = 10^{-7}$. At this point, the construction process stops; f_{2jk} is defined by $f_{2jk}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \lambda_k) = G_{jk,l+1}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \lambda_k) = F_{jk,2l}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \lambda_k) + F_{jk,2l+1}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \lambda_k)$, where $F_{jk,2l}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \lambda_k)$ is a smoothed estimate of y as a function of \tilde{x}_1 and $F_{jk,2l+1}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \lambda_k)$ is a smoothed estimate of y as a function of x_j ; and the adjusted PRESS statistic PRS_{Ajk} is determined for the approximation to y defined by f_{2jk} .

Step 3. Similar to Step 2 with $y \equiv f_{3jk}(\tilde{x}_1, \tilde{x}_2, x_j | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k)$ being estimated through a sequence of smoothing operations for $j = 1, 2, \dots, nX$, $x_j \neq \tilde{x}_1$, $x_j \neq \tilde{x}_2$, and $k = 1, 2, \dots, 6$; details of the estimation of $f_{3jk}(\tilde{x}_1, \tilde{x}_2, x_j | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k)$ are described in Step 3' and are similar to those described in Step 2' for the estimation of $f_{2jk}(\tilde{x}_1, x_j | \tilde{\lambda}_1, \lambda_k)$ with the addition that the intermediate smoothings indicated in Steps 2.2' and 2.3' now involve \tilde{x}_1 , \tilde{x}_2 and x_j rather than \tilde{x}_1 and x_j . Remainder of Step 3 is the same as in Step 2 and results in the selection of \tilde{x}_3 as the most important variable at Step 3.

Step 3'. Procedure for obtaining smoothed model $f_{3jk}(\tilde{x}_1, \tilde{x}_2, x_j | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k)$ on the basis of a relative error criterion for variable x_j , $x_j \neq \tilde{x}_1$, $x_j \neq \tilde{x}_2$, and smoothing parameter λ_k in Step 3.

Step 3.1'. Estimate $F_{jkl}(\tilde{x}_1, \tilde{x}_2 | \tilde{\lambda}_1, \tilde{\lambda}_2) \equiv y - f_1(\tilde{x}_1 | \tilde{\lambda}_1)$ by smoothing on $(\tilde{x}_{i2}, y_i - f_1(\tilde{x}_{i1} | \tilde{\lambda}_1))$, $i = 1, 2, \dots, nS$, with a smoothing spline and smoothing parameter $\tilde{\lambda}_2$ (Note: $F_{jkl}(\tilde{x}_1, \tilde{x}_2 | \tilde{\lambda}_1, \tilde{\lambda}_2)$ was previously determined in Step 2.1' for $\tilde{x}_2 = x_j$ and $\tilde{\lambda}_2 = \lambda_k$). Estimate $f_{jk2}(\tilde{x}_1, \tilde{x}_2, x_j | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k) \equiv y - f_1(\tilde{x}_1 | \tilde{\lambda}_1) - F_{jkl}(\tilde{x}_1, \tilde{x}_2 | \tilde{\lambda}_1, \tilde{\lambda}_2)$ by smoothing on $(x_{ij}, y_i - f_1(\tilde{x}_{i1} | \tilde{\lambda}_1) - F_{jkl}(\tilde{x}_{i1}, \tilde{x}_{i2} | \tilde{\lambda}_1, \tilde{\lambda}_2))$, $i = 1, 2, \dots, nS$, with a smoothing spline and smoothing parameter λ_k . Result is estimate $y_i \equiv G_{jkl}(\tilde{x}_{i1}, \tilde{x}_{i2}, x_{ij} | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k) = f_1(\tilde{x}_{i1} | \tilde{\lambda}_1) + F_{jkl}(\tilde{x}_{i1}, \tilde{x}_{i2} | \tilde{\lambda}_1, \tilde{\lambda}_2) + F_{jk2}(\tilde{x}_{i1}, \tilde{x}_{i2}, x_{ij} | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k)$.

Table 3. Forward Stepwise Variable Selection Algorithm for Sensitivity Analysis with GAMs (Continued)

Step 3.2'. Estimate $F_{jk3}(\tilde{x}_1, \tilde{x}_2, x_j | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k) \cong y - F_{jk1}(\tilde{x}_1, \tilde{x}_2 | \tilde{\lambda}_1, \tilde{\lambda}_2) - F_{jk2}(\tilde{x}_1, \tilde{x}_2, x_j | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k)$ with a smoothing spline on \tilde{x}_1 and smoothing parameter $\tilde{\lambda}_1$. Then, estimate $F_{jk4}(\tilde{x}_1, \tilde{x}_2, x_j | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k) \cong y - F_{jk2}(\tilde{x}_1, \tilde{x}_2, x_j | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k) - F_{jk3}(\tilde{x}_1, \tilde{x}_2, x_j | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k)$ with a smoothing spline on \tilde{x}_2 and smoothing parameter $\tilde{\lambda}_2$, and estimate $F_{jk5}(\tilde{x}_1, \tilde{x}_2, x_j | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k) \cong y - F_{jk3}(\tilde{x}_1, \tilde{x}_2, x_j | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k) - F_{jk4}(\tilde{x}_1, \tilde{x}_2, x_j | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k)$ with a smoothing spline on x_j and smoothing parameter λ_k . Result is estimate $y_i \cong G_{jk2}(\tilde{x}_{i1}, \tilde{x}_{i2}, x_{ij} | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k) = \sum_{r=3}^5 F_{jkr}(\tilde{x}_{i1}, \tilde{x}_{i2}, x_{ij} | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k)$.

Step 3.3'. Similar to Step 3.2' with $F_{jkr}(\tilde{x}_1, \tilde{x}_2, x_j | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k) \cong y - F_{jk,r-2}(\tilde{x}_1, \tilde{x}_2, x_j | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k) - F_{jk,r-1}(\tilde{x}_1, \tilde{x}_2, x_j | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k)$ being estimated for $r = 6, 7$ and 8 by smoothing on \tilde{x}_1 , \tilde{x}_2 and x_j , respectively, with corresponding smoothing parameters $\tilde{\lambda}_1$, $\tilde{\lambda}_2$ and λ_k . Result is estimate $y_i \cong G_{jk3}(\tilde{x}_{i1}, \tilde{x}_{i2}, x_{ij} | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k) = \sum_{r=6}^8 F_{jkr}(\tilde{x}_{i1}, \tilde{x}_{i2}, x_{ij} | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k)$.

Step 3.4'. Continue as in Step 3.4' until the relative error criterion $\|\mathbf{G}_{jk,l+1} - \mathbf{G}_{jkl}\| \leq \text{rerr} \|\mathbf{G}_{jkl}\|$ is satisfied for $\mathbf{G}_{jkr} = [G_{jkr}(\tilde{x}_{i1}, \tilde{x}_{i2}, \tilde{x}_{ij} | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k), G_{jkr}(\tilde{x}_{21}, \tilde{x}_{22}, \tilde{x}_{2j} | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k), \dots, G_{jkr}(\tilde{x}_{nS,1}, \tilde{x}_{nS,2}, x_{nS,j} | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k)]$, $r = l, l + 1$, and $\text{rerr} = 10^{-7}$. At this point, the construction process stops; f_{3jk} is defined by $f_{3jk}(\tilde{x}_1, \tilde{x}_2, x_j | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k) = G_{jk,l+1}(\tilde{x}_1, \tilde{x}_2, x_j | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k) = \sum_{r=3}^{l+2} F_{jkr}(\tilde{x}_1, \tilde{x}_2, x_j | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k)$, where $F_{jkr}(\tilde{x}_1, \tilde{x}_2, x_j | \tilde{\lambda}_1, \tilde{\lambda}_2, \lambda_k)$ is a smoothed estimate of y as function of \tilde{x}_1 , \tilde{x}_2 and x_j for $\tilde{\lambda}_1$, $\tilde{\lambda}_2$ and λ_k , respectively; and the adjusted PRESS statistic PRS_{Ajk} is determined for the approximation to y defined by f_{3jk} .

...

Step N. Terminate process when no variable satisfies specified cutoff.

Table 4. Forward Stepwise Variable Selection Algorithm for Sensitivity Analysis with PP_REG

Notation. Variables $\lambda_k, k = 1, 2, \dots, 6$, represent candidate smoothing parameters used in smoothing splines (see Eq. (3.13)) in the sequential construction of PP_REG models, with λ_k resulting in a smoothing process with approximately δ_k degrees of freedom. Specifically, $\lambda_1 \sim \delta_1 = 1, \lambda_2 \sim \delta_2 = 2, \lambda_3 \sim \delta_3 = 4, \lambda_4 \sim \delta_4 = 7, \lambda_5 \sim \delta_5 = 10$, and $\lambda_6 \sim \delta_6 = 15$. However, unlike the stepwise construction procedure for GAMs described in Table 3, the degrees of freedom associated with smoothing splines in the stepwise construction of PP_REG models is obtained directly from the *ppr* subroutine in R rather than approximated from the δ_k 's.

Step 1. Estimate $y \cong f_{1jk}(x_j | \lambda_k)$ with a smoothing spline on $(x_{ij}, y_i), i = 1, 2, \dots, nS$, for $j = 1, 2, \dots, nX$ and $k = 1, 2, \dots, 6$. For each x_j , select model $f_{1j}(x_j | \tilde{\lambda}_j)$ with smoothing parameter $\tilde{\lambda}_j$ from the models $y \cong f_{1jk}(x_j | \lambda_k), k = 1, 2, \dots, 6$, that results in the smallest value for the adjusted PRESS statistic PRS_A (see Eqs. (3.22) and (3.25) and discussion in Sect. 4.4). For each of the models $y \cong f_{1j}(x_j | \tilde{\lambda}_j)$, determine (i) degrees of freedom, (ii) F -statistic F_j for comparison against mean only model, and (iii) resultant p -value p_j (see Eq. (3.54)). Variable \tilde{x}_1 with smallest p -value is selected as the most important variable at Step 1; corresponding model, smoothing parameter, and degrees of freedom are represented by $y \cong f_1(\tilde{x}_1 | \tilde{\lambda}_1), \tilde{\lambda}_1$ and \tilde{df}_1 , respectively. The process terminates with no variable selected if all p_j are greater than a specified cutoff (e.g., $\alpha = 0.02$).

Step 2. Estimate $y \cong f_{2j}(\tilde{x}_1, x_j | \lambda_{j1}, \lambda_{j2}) = F_{j1}(\tilde{x}_1, x_j | \lambda_{j1}) + F_{j2}(\tilde{x}_1, x_j | \lambda_{j2})$ for $j = 1, 2, \dots, nX$ and $x_j \neq \tilde{x}_1$ through a sequential application of PP_REG described in Steps 2.1' – 2.3'. For each of the models $y \cong f_{2j}(\tilde{x}_1, x_j | \lambda_{j1}, \lambda_{j2})$, determine (i) degrees of freedom, (ii) F statistic for comparison against model $y \cong f_1(\tilde{x}_1 | \tilde{\lambda}_{11})$ constructed in Step 1, and (iii) resultant p -value p_j (see Eq. (3.54)). Variable \tilde{x}_2 with smallest p -value is selected as most important variable at Step 2; corresponding model, smoothing parameters, and degrees of freedom are represented by $y \cong f_2(\tilde{x}_1, \tilde{x}_2 | \tilde{\lambda}_{21}, \tilde{\lambda}_{22}), \tilde{\lambda}_{21}, \tilde{\lambda}_{22}$, and \tilde{df}_2 , respectively. The process terminates with no variable selected at Step 2 if all p_j are greater than a specified cutoff (e.g., $\alpha = 0.02$).

Step 2'. Procedure for obtaining model $f_{2j}(\tilde{x}_1, x_j | \lambda_{j1}, \lambda_{j2})$ and smoothing parameters λ_{j1} and λ_{j2} for variable $x_j, x_j \neq \tilde{x}_1$, in Step 2 through a sequential application of PP_REG.

Step 2.1'. Estimate $y \cong F_{1jk}(\tilde{x}_1, x_j | \lambda_k)$ for $k = 1, 2, \dots, 6$ from the observations $([\tilde{x}_{i1}, x_{ij}], y_i), i = 1, 2, \dots, nS$, with PP_REG as indicated in conjunction with Eq. (3.39). Determine adjusted PRESS value PRS_A (see Eqs. (3.22) and (3.25)) for each of the six models and select model $y \cong F_{j1}(\tilde{x}_1, x_j | \lambda_{j1})$ and associated smoothing parameter λ_{j1} with smallest value for PRS_A .

Step 2.2'. Estimate $F_{2jk}(\tilde{x}_1, x_j | \lambda_k)$ for $k = 1, 2, \dots, 6$ from the observations $([\tilde{x}_{i1}, x_{ij}], y_i - F_{j1}(\tilde{x}_{i1}, x_{ij} | \lambda_{j1})), i = 1, 2, \dots, nS$, with PP_REG as indicated in conjunction with Eq. (3.39); this corresponds to the second step in a PP_REG as indicated in Eq. (3.40).

Table 4. Forward Stepwise Variable Selection Algorithm for Sensitivity Analysis with PP_REG (Continued)

Step 2.3'. Approximations to y are given by $F_{j1}(\tilde{x}_1, x_j | \lambda_{j1}) + F_{2jk}(\tilde{x}_1, x_j | \lambda_k)$ for $k = 1, 2, \dots, 6$. For each of these six models, determine the adjusted PRESS value PRS_A (see Eqs. (3.22) and (3.25) and select the model with the smallest value for PRS_A . Specifically, with $F_{j2}(\tilde{x}_1, x_j | \lambda_{j2})$ and λ_{j2} representing the selected model and smoothing parameter, the desired approximation to y for x_j is given by $y \cong f_{2j}(\tilde{x}_1, x_j | \lambda_{j1}, \lambda_{j2}) = F_{j1}(\tilde{x}_1, x_j | \lambda_{j1}) + F_{2j}(\tilde{x}_1, x_j | \lambda_{j2})$ as indicated at the beginning of Step 2.

Step 3. Same as Step 2 but starting with estimate $y \cong f_{3j}(\tilde{x}_1, \tilde{x}_2, x_j | \lambda_{j1}, \lambda_{j2}, \lambda_{j3}) = \sum_{s=1}^3 F_{js}(\tilde{x}_1, \tilde{x}_2, x_j | \lambda_{js})$ for $j = 1, 2, \dots, nX, x_j \neq \tilde{x}_1$ and $x_j \neq \tilde{x}_2$ developed through a sequential application of PP_REG analogous to that described in Steps 2.1' – 2.3'.

...

Step N. Terminate process when no variable satisfies specified cutoff.

Table 5. Forward Stepwise Variable Selection Algorithm for Sensitivity Analysis with RP_REG

Step 1. Estimate $y \cong f_{1j}(x_j)$ by performing RP_REG on (x_{ij}, y_i) , $i = 1, 2, \dots, nS$, for $j = 1, 2, \dots, nX$ as described in Sect. 3.3.4. This results in partition sets \mathcal{A}_{jk} , $k = 1, 2, \dots, nP_j$, of the values x_{ij} , $i = 1, 2, \dots, nS$, and associated regressions $\hat{y}_{jk} = \hat{\beta}_{0jk} + \hat{\beta}_{1jk}x_j$, $k = 1, 2, \dots, nP_j$, for each x_j . At each step in the partitioning process for x_j (i.e., for $nP_j = 2$, then $nP_j = 3$, and so on), the partition that results in the highest R^2 value is retained (see Eqs. (3.47) – (3.48) and associated discussion); the partitioning process for x_j is stopped when the partitioning of \mathcal{A}_{jk} , $k = 1, 2, \dots, nP_j$, into $\tilde{\mathcal{A}}_{jk}$, $k = 1, 2, \dots, nP_j + 1$, results in the model associated with the partitions $\tilde{\mathcal{A}}_{jk}$ that having a higher adjusted PRESS value $PRSA$ (see Eqs. (3.22) and (3.25)) than the model associated with the partitions \mathcal{A}_{jk} (Note: Because of the sequential partitioning process, only two of the partitions in the sequence $\tilde{\mathcal{A}}_{jk}$, $k = 1, 2, \dots, nP_j + 1$, differ from partitions in the sequence \mathcal{A}_{jk} , $k = 1, 2, \dots, nP_j$). For the model constructed with x_j , determine (i) degrees of freedom $df_j = 3nP_j - 1$, (ii) F -statistic F_j for comparison against mean only model, and (iii) resultant p -value p_j (see Eq. (3.54)). Variable \tilde{x}_1 with smallest p -value is selected as most important variable at Step 1; the corresponding model and degrees of freedom are represented by $y \cong f_1(\tilde{x}_1)$ and \tilde{df}_1 , respectively. The process terminates with no variable selected if all p_j are greater than a specified cutoff (e.g., $\alpha = 0.02$).

Step 2. Estimate $y \cong f_{2j}(\tilde{x}_1, x_j)$ by performing RP_REG on $([\tilde{x}_{i1}, x_{ij}], y_i)$, $i = 1, 2, \dots, nS$, for $j = 1, 2, \dots, nX$ and $x_j \neq \tilde{x}_1$ as described in Sect. 3.3.4. This results in partition sets \mathcal{A}_{jk} , $k = 1, 2, \dots, nP_j$, for the vectors $[\tilde{x}_{i1}, x_{ij}]$, $i = 1, 2, \dots, nS$, and associated regressions $\hat{y}_{jk} = \hat{\beta}_{0jk} + \hat{\beta}_{1jk}\tilde{x}_1 + \hat{\beta}_{2jk}x_j$, $k = 1, 2, \dots, nP_j$, for each x_j (Note: Construction of the partition sets for $[\tilde{x}_{i1}, x_{ij}]$, $i = 1, 2, \dots, nS$, starts *ab initio* and does not involve the partition sets constructed for \tilde{x}_1 in Step 1). At each step in the partitioning process for x_j (i.e., for $nP_j = 2$, then $nP_j = 3$, and so on), the partition that results in the highest R^2 value is retained (see Eqs. (3.47) – (3.48) and associated discussion); the partitioning process associated with x_j is stopped when the partitioning of \mathcal{A}_{jk} , $k = 1, 2, \dots, nP_j$, into $\tilde{\mathcal{A}}_{jk}$, $k = 1, 2, \dots, nP_j + 1$, results in the model associated with the partitions $\tilde{\mathcal{A}}_{jk}$ having a higher adjusted PRESS value $PRSA$ (see Eqs. (3.22) and (3.25)) than the model associated with the partitions \mathcal{A}_{jk} (Note: Because of the sequential partitioning process, only two of the partitions in the sequence $\tilde{\mathcal{A}}_{jk}$, $k = 1, 2, \dots, nP_j + 1$, differ from partitions in the sequence \mathcal{A}_{jk} , $k = 1, 2, \dots, nP_j$). For the model constructed with x_j , determine (i) degrees of freedom $df_j = 5nP_j - 2$, (ii) F -statistic F_j for comparison against model $y \cong f_1(\tilde{x}_1)$ selected in Step 1, and (iii) resultant p -value p_j (see Eq. (3.54)). Variable \tilde{x}_2 with smallest p -value is selected as most important variable at Step 2; the corresponding model and degrees of freedom are represented by $y \cong f_2(\tilde{x}_1, \tilde{x}_2)$ and \tilde{df}_2 , respectively. The process terminates with no variable selected if all p_j are greater than a specified cutoff (e.g., $\alpha = 0.02$).

Step 3. Estimate $y \cong f_{3j}(\tilde{x}_1, \tilde{x}_2, x_j)$ by performing RP_REG on $([\tilde{x}_{i1}, \tilde{x}_{i2}, x_{ij}], y_i)$, $i = 1, 2, \dots, nS$, for $j = 1, 2, \dots, nX$, $x_j \neq \tilde{x}_1$ and $x_j \neq \tilde{x}_2$ as described in Sect. 3.3.4. Continue as in Step 2.

...

Step N. Terminate process when no variable satisfies specified cutoff.
