NAC-0029_R1

Machine Learning for Sensitivity Analysis of Probabilistic Environmental Models

5 June 2014

Prepared by
NEPTUNE AND COMPANY, INC.
1505 15th St, Suite B, Los Alamos, NM 87544

1. Title: Machine Learning for Sensitivity Analysis of Probabilistic Environmental Models		
2. Filename: Sensitivity Analysis v1.2.docx		
3. Description:		
	Name	Date
4. Originator	Paul Black	31 May 2014
5. Reviewer		31 May 2014
6. Remarks		

This page is intentionally blank, aside from this statement.

CONTENTS

FIGURES

TABLES

Executive Summary

The purpose of this document is to explain the development and application of the method used for sensitivity analysis (SA) of Performance Assessment (PA) models constructed in GoldSim. The overarching goal of the SA is to determine which explanatory variables (e.g. Kd in Sand for Tc) have the largest impact on specific endpoints of interest (e.g. Peak Ground Surface Flux of Radon-222). The SA procedure implemented for the Clive DU PA assesses the importance of every explanatory variable (input parameters) used in the GoldSim PA model. In practice, this means that for each endpoint of interest, every explanatory variable in the model has a quantitative measure of importance associated with it. For a given explanatory variable, the quantitative measure of importance depends on the endpoint of interest. All input parameters are included and are essentially varied simultaneously. This is very different from the one-at-a-time SA approaches that are used on deterministic models. This global SA approach allows all levels of interactions to be evaluated, so that the variation in the output can be measured at every level of possible interaction. The effects are collected together during SA processing, and then separated out to attribute the overall contribution of each input parameter to the model output.

As described below, some global SA procedures have the ability to characterize non-linear and non-monotonic relationships between explanatory variables (input parameters) and endpoints of interest (output parameters). This is critically important because of the need to characterize complex interactions among multiple explanatory variables in the PA models. These interactions are often non-linear and non-monotonic

The approach to simulating the PA model affects how the SA should be set up. Each PA simulation is set up to draw random numbers from the input distributions at the beginning of time, and then those random numbers are used throughout time in that simulation. This is one of the reasons why the input distributions are set up to describe the mean of the factor of interest. If, instead, random numbers were drawn at every time step, then the net effect over a long period of time (e.g., 10,000 years) is to create an overall averaging effect. The SA is essentially a regression model that uses the simulated inputs as observations of the input parameters, and the simulated outputs as observations of the output parameters. The form of regression used accommodates non-linear and non-monotonic effects, which are inherent in PA models.

The results of the SA provide a clear indication of the explanatory variables that most strongly influence a given endpoint, since the result of the SA, applied to a given PA endpoint, is a quantitative metric of importance for each explanatory variable. This information can be used in a number of ways. During model development it is a very useful tool for model evaluation, often leading to a better understanding of model constructs and modifications to the model structure as necessary. This leads to iterative model development. Also, if there is an unacceptable level of uncertainty associated with an endpoint of interest (for decision making purposes), the sensitive parameters can be targeted for effective uncertainty reduction; that is, further data or information should be collected to reduce the uncertainty on these sensitive input parameters. Another possibility is to use the results of the SA to simplify the PA model, although the simplification would depend on each specific endpoint.

The remainder of this document provides some background information on SA methods, leading to the choice of SA methods that are used for the Clive DU PA model. This starts with one-at-a-time SA methods for deterministic models, and moves through linear modeling approaches before finishing with global SA approaches.

1.0 Introduction

Decision making for the management of complex systems in the presence of uncertainty requires an explicit characterization of the current state of knowledge. In this context, a model is a valuable tool for understanding the interactions and influence of explanatory variables on the response of interest (e.g. media concentrations or future potential doses). The quantitative assessment of the importance of inputs is critical, and this is especially true when uncertainty in the response is deemed to be unacceptable for the decision at hand. Sensitivity analysis (SA) can be used to help identify those inputs with the greatest impact on uncertainty in the model response (Saltelli *et al*. 1999, Marrell *et al*. 2009, Nossent *et al*. 2011, Morris *et al*. 2014). Specifically, SA helps quantify the benefit of subsequent data collection through the identification of the explanatory variables for which uncertainty reduction through further information collection will yield the most effective decrease in uncertainty for the response of interest. Both analytical and simulation approaches can be implemented to develop SA models that characterize the state of knowledge for a system, and the approach selected depends on the application in question. Analytical representations (e.g. systems of differential equations) have the advantage of allowing more straightforward analysis of model effects; however, some systems possess sufficient complexity that simulation approaches need to be utilized.

Performance assessment (PA) is an integral part of post-closure for radioactive waste disposal facilities and it requires characterization of the fate and transport of waste through space and time. The main goal of a PA is to provide reasonable assurance that performance objectives for radioactive waste disposal will be met; hence it is necessary to depict the relevant dynamics of the facility of interest. This requires the characterization and simulation of multiple system components including: hydrologic, edaphic, radiologic, biotic, and structural features. The overarching model used for a PA is an aggregation of information from multiple sources including: field studies, literature review, and output from other models. PAs are a modeling application where analytical approaches alone are insufficient and the use of a high-dimensional simulation model is essential. SA of high dimensional probabilistic models can be computationally challenging; however, these challenges can be met through the application of machine learning methods applied to probabilistic simulation results. This approach is sometimes referred to as meta-modeling (Marrell *et al*. 2010, Coutts and Yokomizo 2014).

2.0 Sensitivity Analysis Approaches

This section provides a brief review of the common approaches to SA and reviews their pros and cons in the context of application to PA models. Generally speaking, SA deals with the estimation of influence measures for input variables that are components of a given model. In the application to PA models, it is of interest to determine which input variables are driving the uncertainty associated with an endpoint of interest (dose, flux, concentration, etc.). This can be accomplished with either a *qualitative* (Melbourne-Thomas *et al*. 2012) or *quantitative* (Liu *et al*. 2006, Storlie *et al*. 2009) approach applied across a spectrum ranging from *local* (McKay *et al*. 1979) to *global* (Sobol 2001, Friedman 2002) analyses.

Q*ualitative* SA provides a relative ranking of the importance (sensitivity) of input factors without incurring the computational cost of *quantitatively* estimating the percentage of the response (e.g. media concentrations or future potential doses) uncertainty accounted for by each input factor. It was considered more useful prior to the availability of the computational capability needed for quantitative SA approaches. Currently, for the purpose of PA modeling, qualitative SA is of little

utility beyond perhaps the initial preliminary model development stages. It is not considered further in this document.

A *local* SA varies one input factor while holding all other input factors constant and assesses the impact on the corresponding model response. This is often accomplished in an analytical context by examining partial derivatives evaluated at the solution of interest locally. This analysis is *local* in the sense that only a minimal portion of the full volume of the input factor space is explored (*i.e.*, the point at which all but one of the input factors are held constant). Although local sensitivity analysis is useful in some applications, the region of possible realizations for the model of interest is left largely unexplored. Global sensitivity analysis attempts to explore the possible realizations of the model more completely. The space of possible realizations for the model can be explored through the use of search curves or evaluation of multi-dimensional integrals using Monte Carlo methods.

An example of *quantitative local* SA approach is differential analysis based on the partial derivatives of the model with respect to each input factor. Given a model of the form $y = f(X)$, the *local* relative sensitivity measure, S_i, of each input factor, x_i, on model response y can be calculated as:

$$S_i = \left[E_x \left[\frac{\partial f(X)}{\partial x_i} \right]^2 \mathrm{var}_x[x_i] \right]^{1/2} \quad (1)$$

Typically, evaluation of this derivative at a specific solution is performed; hence the analysis is relevant in a small local neighborhood around the solution of interest.

Quantitative global SA attempts to explore the full hyper-volume defined by the collective ranges of possible values for the input factors. Sensitivity indices (SIs) for a single value are obtained by averaging over the variation of all other input factors to provide an estimate of sensitivity:

$$S_i = \frac{\mathrm{var}_{x_i}[E(y \mid x_i)]}{\mathrm{var}(y)} \quad (2)$$

The degree of success for this type of analysis is measured using the quantity, $\sum_{i=1}^{p} S_{xi}$, where p is the number of model parameters. If this sum is approximately 1, then the analysis is considered successful in terms of depicting (and hence allowing the ability to decompose) the observed variability in the response. Given the complexity of PA models, and the rigor that needs to be applied in order to provide reasonable assurance that PA goals being met, the global, quantitative SA is the most effective approach.

Quantitative global SA approaches can be partitioned into two groups: analytical; and, meta-model. Several approaches have been proposed to implement analytical SA for nonlinear, nonmonotonic models. Two of the analytical approaches that are considered here are the Fourier Amplitude Sensitivity Test (FAST) (Saltelli *et al.* 1999) and Sobol's design of experiment (SDOE) approach (Sobol 1993). Application of the meta-model approach (Borgonovo *et al.* 2012, Coutts and Yokomizo 2014) consists of the development of a statistical model that

quantifies the relationship between the response of interest and all of the explanatory variables that enter into the PA model.

The following is a brief review of these approaches. The goal of this review is to articulate the logic for the implementation of the preferred approach, which is the application of the meta-model using gradient boosting machines (GBM).

2.1 Analytical Approach: Sobol Design of Experiment and Fourier Amplitude Sensitivity Test (FAST)

Several analytical approaches have been proposed to handle nonlinear, nonmonotonic models. Two of these approaches include the Fourier Amplitude Sensitivity Test (FAST) (Saltelli *et al*. 1999) and Sobol's design of experiment (SDOE) approach (Sobol 1993). These methods provide an estimate of the proportion of the variation in the model response from an explanatory variable, by using an analysis of variance (ANOVA) type of decomposition of the variability in the model response. These two analytical methods use a different computational strategy for decomposing the partial variances corresponding to increased dimensionality (main-effects, two-way interactions, three-way interactions, etc.).

FAST and SDOE have been shown to be effectively equivalent with respect to SA application (Saltelli *et al*. 1999). In the context of SA, the ANOVA decomposition can be described in terms of total sensitivity indices for each input factor, S_{Ti}. The S_{Ti} for explanatory variable i is calculated as the sum across all main and interaction sensitivities that involve the i^{th} input explanatory variable:

$$S_{T_i} = S_i + \sum_{j \neq i}^{n} S_{ij} + \sum_{j,k \neq i}^{n} S_{ijk} + \cdots \qquad (3)$$

where S_i is the first-order or (main effect) sensitivity index and S_{ij} is the second-order (two- way interaction effect) sensitivity index and so on. For a single $S_{Ti,}$, the total number of interactive terms for sensitivity indices is $2^n - 1$, where n is the number of explanatory variables. Because SDOE requires multi-dimensional integration to estimate the sensitivity indices, this method can be prohibitive computationally for moderately complex models. These approaches become more computationally intensive as the dimensionality of the model (i.e., the number of model parameters) increases and can be prohibitive for models that include hundreds or thousands of stochastic explanatory variables.

2.2 Meta-models: Regression Based Methods

Regression based approaches are an option for the q*uantitative* global SA of PA models. These approaches include squared standardized regression coefficients (SSRC) and squared standardized rank regression coefficients (SSRRC) (Storlie *et al*. 2009, Cea *et al*. 2011). Some of the benefits of these methods are ease of implementation and relative familiarity of the basic output of regression models for most members of a target audience. However, one of the main drawbacks of these approaches is that the methods assume a monotonic and linear relationship between the input factors and the model response. The degree to which the relationships between explanatory and response variables follow these assumptions impacts the validity of the results of these simple linear regression-based analyses. That is, this method does not provide reliable SA for systems with a high degree of non-linearity or moderate lack of monotonicity.

A linear regression model has the following form

$$y = \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \varepsilon \quad (4)$$

where *y* represents the output from a PA model (i.e. the response of interest), and the x_i's corespond to each of the explanatory variables built into the PA model. The variance of the linear regression model in Equation 4 can be estimated as

$$\widehat{v(y)} = \sum_{i=1}^{p} \hat{\beta}_i^2 \, \widehat{var(x_i)} \quad (5)$$

where $\hat{\beta}_i^2$ is the square of the standard parameter estimate for the i^{th} variable in the regression model. Equation [5] requires an assumption that the input factors are independent, which is often not the case. For example, many of the explanatory variables in a PA model are correlated as a direct consequence of the nature of the system. If the model response and explanatory variables are standardized to a mean of 0 and a variance of 1 then the square of the regression coefficients (i.e. $\hat{\beta}_i^2$) provides an estimate of S_i (in the form of Equation 3).

Regressing the *ranks* of the model response on explanatory variables can help to mitigate (but not obviate) the impact of nonlinearities in the model on the lack of validity of output for the regression model. The coefficient of determination associated with the regression model in Equation [4], R^2, measures the percent of variability in the ranks of the model response that is explained by the linear combination of the explanatory variables multiplied by their respective parameters. The closer R^2 is to one, the less unexplained variability there is in the rank response and the better the regression model performs as a meta-model estimating the dynamics of the more complex PA model.

Alternatively, the quantity 1- R^2 represents the percent of rank response variation not accounted for by the SA method. As this percent increases, confidence in the analysis is reduced although the resulting relative ranking may still be of value. For models with low enough values of R^2, the validity of the relative rankings also comes into question. In this context, it is worth recalling that SSRC and SSRRC assume a monotonic linear relationship between the explanatory variables and the response of interest for the PA model output. A low R^2 might be reflective of a model structure that does not meet this assumption.

2.3 Meta-models: Machine Learning Approaches

Because of the computational cost, SA of high-dimensional probabilistic PA models requires efficient algorithms for practical application. Machine learning is a general approach that provides tools for the construction of meta-models that can be used for subsequent SA. These models allow for the partitioning of the variance in the model response in a manner that allocates a proportion to the explanatory variables. Two common machine learning approaches that could be used for SA are Multivariate Adaptive Regression Splines (MARS) (Friedman 1991) and the gradient boosting machine (GBM) (Friedman 2001, Friedman 2002). Several of the most important advantages of machine learning approaches are: the ability to fit non-monotonic and non-linear effects; the ability to fit parameter interaction effects; and, the ability to visualize these effects and their interaction across the range of the response and input parameters. MARS, boosting and other machine learning approaches typically produce similar results for noisy data.

In the case of realizations from a probabilistic process model, each realization is a deterministic evaluation of the model and all the stochastic predictor variables are available. As such there is no unexplainable variation in the process model response, which is a stark contrast to the case with observed data. Hence, machine learning algorithms should theoretically be able to construct models that have R^2 values very close to one. Given that, it might be tempting to conclude that machine learning SA methods are all effectively equivalent; however, with respect to the ease of implementation and interpretability of results, GBM has an advantage. The remainder of this section articulates this with a high level review of both MARS and GBM.

2.3.1 Multivariate Adaptive Regression Splines (MARS)

MARS is a recursive partitioning approach that directly addresses the ANOVA decomposition "curse of dimensionality", making estimation of sensitivity indices computationally achievable for large *n* (Friedman, 1991). MARS accomplishes this by optimally partitioning or, splitting the model response and explanatory variables into subsets, from which splines (i.e. piecewise smooth polynomial functions) are fit. The recursive nature of the algorithm results in increasingly local splits of the model response in which all significant interaction effects in sub-regions are found. MARS is able to find and fit significant nonlinear and threshold relationships between the model input and explanatory variables. An explanatory variable's influence is calculated using MARS as the sum of the partial residuals removing all main and interaction effects that variable enters.

$$S_{x_i} = \sum \left[f - a_o + \sum_{i \neq 1} f(x_i) + \sum_{i,j \neq 1} f(x_i, x_j) + \sum_{i,j,k \neq 1} f(x_i, x_j, x_k) + \cdots \right]^2 \qquad (6)$$

2.3.2 Gradient Boosting Machines (GBM)

Boosting of regression trees provides a technique that adds to the flexibility offered by recursive partitioning methods such as MARS. The GBM (Friedman 2001, Friedman 2002) approach utilizes boosting of binary recursive partitioning algorithms that deconstructs a response into the relative influence from a given set of explanatory variables (i.e. PA model input parameters). The deconstruction breaks the PA model into separate parts (braches of the regression tree), and each part is examined separately. This process is repeated with smaller and smaller parts, each analyzed for the relationship between the explanatory variables and the PA model output (i.e. the response of interest). The deconstructed parts are then collected together to provide estimates of the sensitivity of each exploratory variable for a specific response variable from the PA model.

Hence, GBM provides a method for constructing a statistical meta-model of the more complex PA simulation model. The GBM approach identifies the most influential explanatory variables in the context of the observed uncertainty in the PA model output or response. Critically, the GBM method also identifies the range over which the influence is strongest, which can be used to better understand the full effect of a sensitive explanatory variable on the output results.

Variance decomposition of the GBM fit is then used to estimate SIs. Under this decomposition approach, the goal is to identify the most influential explanatory variables that are identified within a model. The necessary degree of model complexity can be assessed using validation metrics, based on comparison of model predictions, with randomly selected subsets of the data. This approach uses the "deviance" of the model as a measure of goodness of fit. The concept of deviance is fundamental to classical statistical hypothesis tests (e.g., the common *t*-test can be

derived using a deviance-based framework) and guides the model selection process applied here. The deviance for a model given a set of data, y, is defined as

$$D(y) = -2(\log\left(p\left(y|\hat{\theta}_{sub}\right)\right) - \log\left(p\left(y|\hat{\theta}_{full}\right)\right) \quad (7)$$

Where $\hat{\theta}_{sub}$ is the vector of fitted values from the model of interest, and $\hat{\theta}_{full}$ is the vector of fitted values from the saturated model. Equation 7 is effectively the log likelihood ratio of the fitted versus the full model. In this sense, it measures the deviance of the fitted model from the full model.

Given that the PA models are deterministic simulations, where the only stochasticity comes from the distributions assigned to each of the explanatory variables, the variance of the response from the PA is solely attributable to the model uncertainty. Hence, these models can't be overfit in the same way that they would when applied to observational data. Because of this, the determination of sufficient GBM model complexity should be focused on the minimal amount necessary to explain approximately the maximal amount of the observed variance in the PA model response.

The GBM fitting approach is based on finding the values of each explanatory variable that result in the greatest difference in the mean for the corresponding subsets of the response. For example, if there were only a single explanatory variable, the GBM would identify the value of the explanatory variable that corresponds to a split of the response into two parts for which the means are more different than those corresponding to any other split of the response into two subsets. When multiple explanatory variables are present, multiple splits are made corresponding to each of the explanatory variables and the collection of splits is referred to as a "tree". Each tree results in an estimate (e.g., prediction) of the response. As multiple potential trees are evaluated, they are compared to the observed data using a loss function. The selection of the loss function is an important aspect of the GBM process, and depends on the distribution of the response variable. For data that are sufficiently skewed (e.g., non-normal), experience has shown the absolute error loss function typically produces more reliable results. This is often the case with realizations of responses generated from PA models since there are often many cases of small response (e.g., dose, flux, concentration) and only a few for which large responses are simulated by the PA model.

There is a trade-off that exists when considering which loss function to use. The squared-error loss function results in better fitting models, but can do so at the expense of introducing spurious variables into the model selection process when the response distribution is sufficiently skewed. The absolute error loss function produces model predictions with more variability, but is less likely to result in the selection of spurious variables in the model. This is due to the squared-error loss function methods increased sensitivity to results from the tails of distributions.

When considering the utility of the GBM approach and its increased computational burden relative to simpler linear regression based approaches, it is important to recall that linear regression techniques (e.g. SSRC) assume that the relationship between the response and the explanatory variable is a constant (i.e. the statistical model is linear in the parameter space). With the GBM approach, this relationship is not constrained by assumptions of linearity, and partial dependence plots are used to show the estimate of the relationship between the response (i.e. the output from the PA model) and the explanatory variables.

Partial dependence plots are used to describe and interpret the results of the SA. The partial dependence curve depicts the change in the value of an endpoint as a function of the values of the response variable (see blue curves in Figure 1). It is conceptually equivalent to the slope in a linear regression model, but shows the non-linear impact across the range of input values. In a linear regression model, this relationship is constrained to be constant. That is, the relationship between a change in the explanatory variable and the endpoint of interest does not change across the range of values for the explanatory variable. In the GBM approach, the relationship between the explanatory variable and the endpoint of interest is allowed to change across the range of values for the explanatory variable. For example, the x2 explanatory variable in Figure 1 displays a different relationship to the response between the values of 0 and 0.5 than it does between 0.5 and 1.0. Specifically, for the range of explanatory variable values between 0 and 0.5 there is a decrease in the response; however for the range of values between 0.5 and 1.0, there is an increase in the response. This is a powerful distinction between the GBM approach and other meta-modeling applications that do not allow this functional flexibility in the relationship between the explanatory variable and response to be evaluated.

2.4 Example: Comparison of SA methods

2.4.1 "Sobol g-function"

The Sobol *g*-function (Saltelli *et. al.* 1999) provides an analytic non-monotonic test function for evaluating the performance of various SA methods. This function is defined as:

$$f = \prod_{i=1}^{p} g_i(x_i) \tag{7}$$

where p is the total number of input factors and $g_i(x_i)$ is given by

$$g_i(x_i) = \frac{|4x_i - 2| + a_i}{1 + a_i}, \tag{8}$$

with

$$x_i = \frac{1}{2} + \frac{1}{p}\arcsin(\sin(w_i s + f_i)), \tag{9}$$

and s varying along $(-\pi,\pi)$, $\varphi_i \sim U[0,2\pi)$, and ω_i are specified frequencies.

The Sobol *g* function was simulated for $p = 8$ and frequencies $\{\omega_i\} = \{23, 55, 77, 97, 107, 113, 121, 125\}$ for a specific set of a_i's. Table 1 provides a comparison of sensitivity indices, *S*, calculated analytically (Saltelli *et. al.* 1999) versus those computed using GBM, MARS, FAST, differential analysis (DERIV), squared standardized regression coefficients (SSRC), and, squared standardized rank regression coefficients (SSRRC). DERIV is used to represent the calculus based approach based on derivatives evaluated at a point.

Note that the GBM, MARS and FAST methods all return sensitivity indices that are close to the actual sensitivities for the Sobol function (S). The Sobol function is highly non-linear; hence the standardized regression approaches (i.e. SSRC and SSRRC) do not work very well. As described

above, FAST is computationally challenging. The difference between MARS and GBM is small, but preference based on the results in this example is given overall to the GBM approach.

A goodness-of-fit statistic is also presented in the bottom row of Table 1. This is calculated as the standard chi-square goodness-of-fit statistic (i.e. the sum of the square of the observed (SA method) minus the expected (*S* value)) all divided by the expected value, in which case a small value implies a better fit. These goodness-of-fit statistics show that the GBM method outperforms the other methods, although the difference is small for GBM and FAST.

Table 1. Sensitivity Indices by Sensitivity Analysis Method for Sobol g-function application with p = 8.

	a	***S***	***GBM***	***MARS***	***FAST***	***DERIV***	***SSRC***	***SSRRC***
x_1	99	0.0001	0.0003	0.0000	0.0043	0.0037	0.6880	0.7805
x_2	0	0.4227	0.4146	0.4397	0.4287	0.3151	0.0137	0.0036
x_3	9	0.0058	0.0011	0.0084	0.0190	0.0401	0.0003	0.0000
x_4	0	0.4227	0.4200	0.4239	0.4269	0.3169	0.0163	0.0098
x_5	99	0.0001	0.0001	0.0000	0.0006	0.0037	0.0350	0.1152
x_6	4.5	0.0182	0.0335	0.0239	0.0141	0.0787	0.0012	0.0554
x_7	1	0.1304	0.1303	0.1041	0.1063	0.2382	0.0574	0.0344
x_8	99	0.0001	0.0000	0.0000	0.0002	0.0037	0.1881	0.0012
Goodness-of-Fit statistic			3.3	14.8	3.6	470	7,250	535

GBM is run on the realizations themselves, whereas FAST requires set up in terms of an embedded signal. This makes FAST relatively cumbersome to deal with. Also, GBM outperforms MARS, which is not as flexible and takes more time to implement computationally. GBM tends to provide the best fit, is flexible and is applied directly to the realizations from the PA model. Consequently, GBM is the preferred method, and the one that is used for the sensitivity analyses for the Clive DU PA.

2.4.2 Visualization

Once a GBM has been is constructed, every explanatory variable in the PA model has a corresponding sensitivity index (SI). Experience has shown that for PA models with hundreds of parameters, the majority of them will have SI values that are very near zero. That is, for a given response from the PA model, the majority of the uncertainty in the response values simulated by the PA model will be attributable to a handful of explanatory variables. The collection of important variables will change as different responses from the PA model are considered.

The SI is obtained through variance decomposition and can be interpreted as the percent of variability in the PA model output explained by a given explanatory variable. The sum of the

SI's across the entire set of explanatory variables in the PA model will approximately equal the R^2 of the linear regression of the realizations from the PA model on the GBM predictions.

For a GBM model, the partial dependence of the response on each explanatory variable is determined through the integration across the joint density of the explanatory variables to obtain a marginal distribution for each explanatory variable. That is, if there are *n* explanatory variables in a given PA model, the partial dependence for a single explanatory variable is obtained by integrating across the other *n*-1 explanatory variables in the PA model. The integration is performed using a "weighted tree traversal" measure that is analogous to more common integration procedures performed with Riemann or Lebesgue measures (Friedman 2001).

The vertical axis of a partial dependence plot has units corresponding to those of the response variable of interest for the meta-model built on the PA model output response. The partial dependence shows the change in the response variable as a function of the changes in the explanatory variable. If the underlying relationship between the response from the PA model and the explanatory variable of interest is linear in the parameters (as is the assumption for a linear regression model), then the partial dependence curve will be a line with slope equal to that of the corresponding regression model.

In order to assess the relationship between an individual explanatory variable and the response of interest, partial dependence plots are used (an example is provided in Figure 1). The first panel depicts a density estimate of the simulated response from the PA model as well as the R^2 and summary statistics for the response. The percentiles of the response distribution in this panel are shaded to provide a context for the partial dependence plots presented in the remaining panels. The colors indicate the percentile range of the response as follows:

1. The 0th - 25th percentile region is shaded orange-brown
2. The 25th - 50th percentile region is shaded dark yellow-green
3. The 50th - 75th percentile region is shaded light green
4. The 75th - 100th percentile region is shaded light blue

To reiterate, the y-axis of the partial dependence plots is in units of the response distribution (which is the x axis of the first panel in the upper left). Given that each parameter has a different range and strength of influence on the response, the y axes of the partial dependence panels have been constructed to depict only the range of the response over which a particular parameter is influential. This provides the most meaningful presentation of the variables and their relationships. In contrast, if the original scale of the response were maintained on each partial dependence panel, then the influence of the least influential parameter would not be visible in many cases. To mediate issues associated with this change in range of the vertical axes among the different partial dependence plots, the background of the partial dependence panels is colored to depict the percentiles of the response over which the parameter is influential. For example, if the background of the partial dependence plot under the partial dependence line is light blue, then that indicates the parameter's influence on the upper end of the response distribution (i.e., the 75th to 100th percentile of the response).

The partial dependence panels in each figure show the distributions of the explanatory variables (black line), and the partial dependence curve (blue line) shows changes in the response as a function of each explanatory variable.

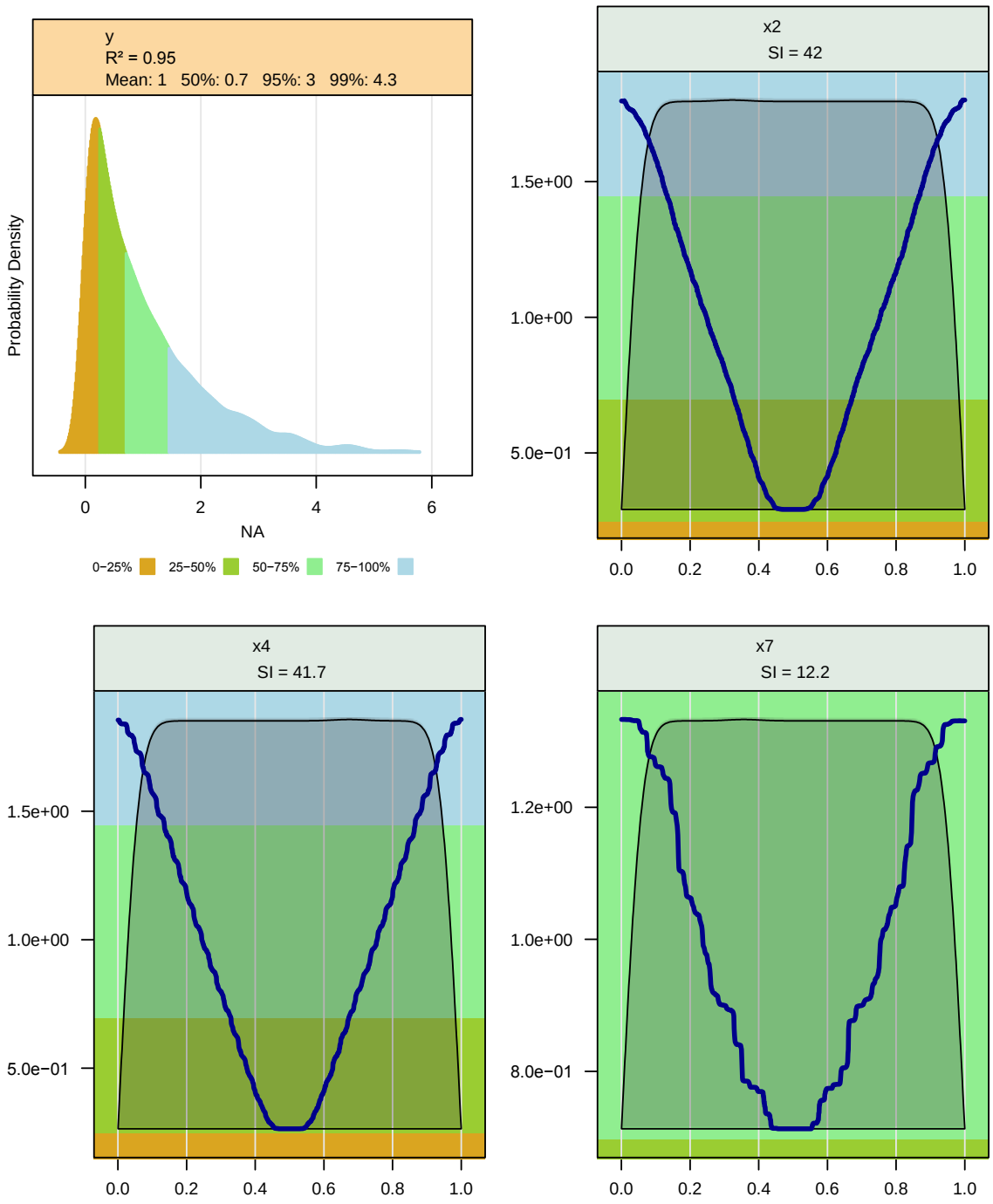

Figure 1. Sensitivity and Partial Dependence Plots for the GBM fit to the Sobol Function.

The plots show that the distributions for the three input parameters are uniform, and that the effects show sensitivity across the entire range of the inputs. The effects are first negative, and then positive, which is to be expected given Equation 15. Also note that the linear regression methods would not be able to track the non-linearity, and instead fits a straight, horizontal line for these parameters, which shows them to be non-sensitive. This is a prime example of why methods such as GBM are advantageous.

Table 2 shows a specific example for one of the Clive DU PA endpoints of interest – the peak groundwater concentrations within 500 years for technetium-99. As can be seen, all explanatory variables (input parameters) are included in the GBM-based global SA. The dependent variable (output variable or response variable) of interest is the peak groundwater concentration of ^{99}Tc in 500 years. The most sensitive input parameter is the ^{99}Tc Kd. This is consistent with the conceptual understanding of the model. That is, Kd dictates the rate of movement through the environmental system towards groundwater, and greater groundwater concentrations are likely to be associated with a lower Kd. Note that inventory of ^{99}Tc is not a sensitive parameter. Although it is the third most sensitive input parameter, its impact on the output is swamped by impact of the Kd of ^{99}Tc in sand. This means that the output is not affected very much by the inventory of ^{99}Tc because the inventory uncertainty is swamped by the uncertainty impact of the Kd of ^{99}Tc in sand. Or, if the inventory distribution is split in two, then the output ^{99}Tc groundwater concentration distribution is not very different for those two groups because it is dominated by the Kd of ^{99}Tc in sand.

Further examples are provided in the Sensitivity Analysis Results V1.2 White Paper.

Table 2. Peak Groundwater Well Concentrations within 500 years - Tc99

Explanatory Variable	Sensitivity Index
Kd Sand for Tc (mL/g)	76.63
Molecular Diffusivity in Water (cm2/s)	20.55
Activity Conc in SRS DU Waste: Tc99 (pCi/g)	1.47
Saturated Zone Water Table Gradient	0.92
Unit 2 Saturated Hyd Cond (cm/s)	0.17
log of van Genuchten's n for Unit 4	0.11
Unit 3 Bulk Density (g/cm3)	0.08
log of van Genuchten's α for Unit 4	0.03
Unsaturated Zone Thickness (m)	0.03
Unit 3 Bubbling Pressure Head (cm)	0.02
Unit 2 Porosity	0.00
Kd Sand for Th (mL/g)	0.00
Number of Gullies	0.00
Biomass % Cover Selector	0.00
Mammal Burrow Excavation Rate (m3/yr)	0.00
RipRap Porosity	0.00
Saltwater Solubility for Ra (mol/L)	0.00
Activity Conc in SRS DU Waste: Cs137 (pCi/g)	0.00
Kd Clay for Pu (mL/g)	0.00
Kd Clay for Cs (mL/g)	0.00
Plant/Soil Conc Ratio for Am	0.00
Kd Silt for Ra (mL/g)	0.00
Ant Colony Density - Plot 5 (1/ha)	0.00
Mammal Mound Density - Plot 4 (1/ha)	0.00
Kd Sand for U (mL/g)	0.00
Large Lake End (yr)	0.00

Mammal Mound Density - Plot 1 (1/ha)	0.00
Grass Root Shape Parameter b	0.00
Water Ingestion Rate for Cattle (kg/day)	0.00
Tortuosity Water Content Exponent	0.00
Activity Conc in SRS DU Waste: Rn222 (pCi/g)	0.00
Saltwater Solubility for Cs (mol/L)	0.00
Silt Sand Gravel BulkDensity (g/cm3)	0.00
Activity Conc in SRS DU Waste: Pu239 (pCi/g)	0.00
Siberia Gully Selector	0.00
Activity Conc in SRS DU Waste: U234 (pCi/g)	0.00
Resuspension Flux (kg/m2-yr)	0.00
Unit 4 Bulk Density (g/cm3)	0.00
Intermediate Lake Sed Thickness (m)	0.00
Kd Sand for Pu (mL/g)	0.00
Saltwater Solubility for Pb (mol/L)	0.00
Kd Silt for Pb (mL/g)	0.00
Ant Colony Density - Plot 2 (1/ha)	0.00
Beef Transfer Factor for Np (day/kg)	0.00
Surface Wind Speed (m/s)	0.00
Activity Conc in SRS DU Waste: U238 (pCi/g)	0.00
Activity Conc in SRS DU Waste: Ra226 (pCi/g)	0.00
Kd Silt for U (mL/g)	0.00
DCF Photon2 REF	0.00
Saltwater Solubility for Th (mol/L)	0.00
OHV Dust Adjustment	0.00
RipRap Bulk Density (g/cm3)	0.00
Fine Gravel Mix Porosity	0.00
Fine Gravel Mix BulkDensity (g/cm3)	0.00
Fine CobbleMix Porosity	0.00
Fine Cobble Mix BulkDensity (g/cm3)	0.00
Silt Sand Gravel Porosity	0.00
Natural Rn Barrier Clay Sat Hyd Cond (cm/s)	0.00
Soil Temperature (Â°C)	0.00
Unit 3 Porosity	0.00
Unit 3 Residual Water Content	0.00
Unit 3 Saturated Hyd Cond (cm/s)	0.00
Unit 3 Brooks-Corey Fractal Dimension	0.00
Unit 4 Porosity	0.00
Unit 4 Bubbling Pressure Head (cm)	0.00
Unit 4 Residual Water Content	0.00
Unit 4 Saturated Hyd Cond (cm/s)	0.00
Unit 4 Brooks-Corey Fractal Dimension	0.00
Unit 2 Bulk Density (g/cm3)	0.00

Saltwater Solubility for Ac (mol/L)	0.00
Saltwater Solubility for Am (mol/L)	0.00
Saltwater Solubility for Np (mol/L)	0.00
Saltwater Solubility for Pa (mol/L)	0.00
Saltwater Solubility for Pu (mol/L)	0.00
Saltwater Solubility for Sr (mol/L)	0.00
Saltwater Solubility for Tc (mol/L)	0.00
Saltwater Solubility for UO3 (mol/L)	0.00
Saltwater Solubility for I (mol/L)	0.00
Saltwater Solubility for Rn (mol/L)	0.00
Saltwater Solubility for U3O8 (mol/L)	0.00
Kd Sand for Ac (mL/g)	0.00
Kd Sand for Am (mL/g)	0.00
Kd Sand for Cs (mL/g)	0.00
Kd Sand for Np (mL/g)	0.00
Kd Sand for Pa (mL/g)	0.00
Kd Sand for Pb (mL/g)	0.00
Kd Sand for Ra (mL/g)	0.00
Kd Sand for Sr (mL/g)	0.00
Kd Sand for I (mL/g)	0.00
Kd Silt for Ac (mL/g)	0.00
Kd Silt for Am (mL/g)	0.00
Kd Silt for Cs (mL/g)	0.00
Kd Silt for Np (mL/g)	0.00
Kd Silt for Pa (mL/g)	0.00
Kd Silt for Pu (mL/g)	0.00
Kd Silt for Sr (mL/g)	0.00
Kd Silt for Th (mL/g)	0.00
Kd Clay for Ac (mL/g)	0.00
Kd Clay for Am (mL/g)	0.00
Kd Clay for Np (mL/g)	0.00
Kd Clay for Pa (mL/g)	0.00
Kd Clay for Pb (mL/g)	0.00
Kd Clay for Ra (mL/g)	0.00
Kd Clay for Sr (mL/g)	0.00
Kd Clay for Th (mL/g)	0.00
Kd Clay for U (mL/g)	0.00
Liner Clay Saturated Hyd Cond (cm/s)	0.00
Radon Escape/Production Ratio	0.00
Resuspended Particle Fraction	0.00
Surface Atmosphere Thickness (m)	0.00
Surface Atmosphere Diffusion Length (m)	0.00
Ant Nest Volume (m3)	0.00

Ant Colony Lifespan (yr)	0.00
Ant Nest Shape Parameter b	0.00
Ant Colony Density - Plot 1 (1/ha)	0.00
Ant Colony Density - Plot 3 (1/ha)	0.00
Ant Colony Density - Plot 4 (1/ha)	0.00
Mammal Burrow Shape Parameter b	0.00
Mammal Mound Density - Plot 2 (1/ha)	0.00
Mammal Mound Density - Plot 3 (1/ha)	0.00
Mammal Mound Density - Plot 5 (1/ha)	0.00
Plant/Soil Conc Ratio for Ac	0.00
Plant/Soil Conc Ratio for Cs	0.00
Plant/Soil Conc Ratio for I	0.00
Plant/Soil Conc Ratio for Np	0.00
Plant/Soil Conc Ratio for Pa	0.00
Plant/Soil Conc Ratio for Pb	0.00
Plant/Soil Conc Ratio for Pu	0.00
Plant/Soil Conc Ratio for Ra	0.00
Plant/Soil Conc Ratio for Sr	0.00
Plant/Soil Conc Ratio for Tc	0.00
Plant/Soil Conc Ratio for Th	0.00
Plant/Soil Conc Ratio for U	0.00
Grass Root/Shoot Ratio	0.00
Shrub Root Shape Parameter b	0.00
Shrub Root/Shoot Ratio	0.00
Tree Root Shape Parameter b	0.00
Tree Root/Shoot Ratio	0.00
Greasewood Root Shape Parameter b	0.00
Greasewood Root/Shoot Ratio	0.00
Forb Root/Shoot Ratio	0.00
Forb Root Shape Parameter b	0.00
Vegetation Association Selector	0.00
Biomass Production Rate (kg/ha/yr)	0.00
Tortuosity Porosity Exponent	0.00
Angle of Repose for Gullies (Â°)	0.00
Gully b Shape Parameter	0.00
Activity Conc in SRS DU Waste: Sr90 (pCi/g)	0.00
Activity Conc in SRS DU Waste: I129 (pCi/g)	0.00
Activity Conc in SRS DU Waste: Pb210 (pCi/g)	0.00
Activity Conc in SRS DU Waste: Ra228 (pCi/g)	0.00
Activity Conc in SRS DU Waste: Ac227 (pCi/g)	0.00
Activity Conc in SRS DU Waste: Th228 (pCi/g)	0.00
Activity Conc in SRS DU Waste: Th229 (pCi/g)	0.00
Activity Conc in SRS DU Waste: Th230 (pCi/g)	0.00

Activity Conc in SRS DU Waste: Th232 (pCi/g)	0.00
Activity Conc in SRS DU Waste: Pa231 (pCi/g)	0.00
Activity Conc in SRS DU Waste: U232 (pCi/g)	0.00
Activity Conc in SRS DU Waste: U233 (pCi/g)	0.00
Activity Conc in SRS DU Waste: U235 (pCi/g)	0.00
Activity Conc in SRS DU Waste: U236 (pCi/g)	0.00
Activity Conc in SRS DU Waste: Np237 (pCi/g)	0.00
Activity Conc in SRS DU Waste: Pu238 (pCi/g)	0.00
Activity Conc in SRS DU Waste: Pu240 (pCi/g)	0.00
Activity Conc in SRS DU Waste: Pu241 (pCi/g)	0.00
Activity Conc in SRS DU Waste: Pu242 (pCi/g)	0.00
Activity Conc in SRS DU Waste: Am241 (pCi/g)	0.00
GDP DU Inventory Storage Dead Space (m2)	0.00
Contaminated Fraction of GDP DU	0.00
Angle Of Repose Gully Fan (deg)	0.00
Distance to Gully Initiation (m)	0.00
Saturated Zone Thickness (m)	0.00
Large Lake Start (yr)	0.00
Large Lake Sedimentation Rate (m/yr)	0.00
Site Dispersal Area (km2)	0.00
Intermediate Lake Duration (yr)	0.00
Intermediate Lake Depth Above Clive (m)	0.00
Large Lake Depth Above Clive (m)	0.00
Forage Ingestion Rate for Cattle (kg/day)	0.00
Soil Ingestion Rate for Cattle (kg/day)	0.00
Soil Ingestion Rate for Antelope (kg/day)	0.00
Water Ingestion Rate for Antelope (kg/day)	0.00
Body Weight Factor for Antelope	0.00
Beef Transfer Factor for Ac (day/kg)	0.00
Beef Transfer Factor for Am (day/kg)	0.00
Beef Transfer Factor for Cs (day/kg)	0.00
Beef Transfer Factor for I (day/kg)	0.00
Beef Transfer Factor for Pa (day/kg)	0.00
Beef Transfer Factor for Pb (day/kg)	0.00
Beef Transfer Factor for Pu (day/kg)	0.00
Beef Transfer Factor for Ra (day/kg)	0.00
Beef Transfer Factor for Sr (day/kg)	0.00
Beef Transfer Factor for Tc (day/kg)	0.00
Beef Transfer Factor for Th (day/kg)	0.00
Beef Transfer Factor for U (day/kg)	0.00
Receptor Area (ha)	0.00
Antelope Range Area (acre)	0.00
DCF Alpha REF	0.00

DCF Beta REF	0.00
DCF Photon1 REF	0.00
Plant Fresh Weight Conversion	0.00
Soil Ingestion Tracer Element	0.00
Meat Preparation Loss	0.00
Meat Post-Cooking Loss	0.00

3.0 References

Borgonovo E., Castaings W., and Tarantols S. (2012) "Model emulation and moment-independent sensitivity analysis: An application to environmental modelling", *Environmental Modelling and Software* **34**:105-115.

Cea, L., Bermudez M., and Puertas J. (2011) "Uncertainty and sensitivity analysis of a depth-averaged water quality model for evaluation of *Escherichia Coli* concentration in shallow estuaries" *Environmental Modelling and Software* **26**:1526-1539.

Coutts S.R. and Yokomizo H. (2014) "Meta-models as a straightforward approach to the sensitivity analysis of complex models", *Population Ecology* **56**:7-19.

Friedman J.H. (1991) "Multivariate Adaptive Regression Splines", *Annals of Statistics* **19(1)**:1-67.

Friedman J.H. (2001) "Greedy function approximation: A gradient boosting machine", *Annals of Statistics* **29(5)**:1189-1232.

Friedman J.H. (2002) "Stochastic gradient boosting", *Computational Statistics and Data Analysis* **38**:367-378.

Liu H., Chen W., and Sudjianto A. (2006) "Relative Entropy Based Method for Probabilistic Sensitivity Analysis in Engineering Design" *Journal of Mechanical Design* **128**:326-336.

Marrel A.M., Iooss B., Laurent B., and Roustant O. (2009), "Calcluations of Sobol indices for the Gaussian process metamodel" *Reliability Engineering and Safety System* **94**:742-751.

Marrel A.M., Iooss B., Jullien M., Laurent B., and Volkova E. (2011), "Global sensitivity analysis for models with spatially dependent outputs" *Environmetrics* **22**:383-397.

McKay M. D., Beckman R.J., and Conover W.J. (1979) "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code" *Technometrics* **21(2)**: 239-245.

Melbourne-Thomas J., Wotherspoon S., Raymond B., and Constable A. (2012) "Comprehensive evaluation of model uncertainty in qualitative network analyses" *Ecological Monographs* 82(4):505-519.

Morris D.J., Speirs D.C., Cameron A.I., Heeath M.R. (2014) "Global sensitivity analysis of an end-to-end marine ecosystem model of the North Sea: Factors affecting the biomass of fish and benthos", *Ecological Modelling* **273**:251-263.

Nossent J., Elsen P., and Bauwens W. (2011) "Sobol's sensitivity analysis of a complex environmental model", Environmental Modelling and Software **26**:1515-1525.

Saltelli A., Tarantola S., and Chan K.P.-S. (1999) "A quantitative model-independent method for global sensitivity analysis of model output", *Technometrics* **41**:39-55.

Sobol I.M. (1993) "Sensitivity analysis for nonlinear mathematical models", *Mathematical Modeling & Computational Experiment* **1**:407-414.

Sobol, I.M. (1993), "Sensitivity Analysis for Nonlinear Mathematical Models," *Mathematical Modeling & Computational Experiment*, 1, 407-414.

Sobol I.M. (2001) "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates", *Mathematics and Computers in Simulation* **55**:271-280.

Storlie C.B., Swiler L.P., Helton J.C., and Sallaberry C.J. (2009) "Implementation and evaluation f nonparametric regression procedures for sensitivity analysis of computationally demanding models." *Reliability Engineering and Safety System* **94**:1735-1763.