MEAN-BASED SENSITIVITY OR UNCERTAINTY IMPORTANCE MEASURES FOR IDENTIFYING INFLUENTIAL PARAMETERS

Sitakanta Mohanty[1], Y-T. (Justin) Wu[2]

1. CNWRA, Southwest Research Institute, San Antonio, Texas, USA
2. Applied Research Associates, Inc., Raleigh, North Carolina, USA

ABSTRACT

Two sensitivity (or uncertainty importance) measures particularly relevant to the disposal of HLW are presented. These measures are referred to as performance-mean-based sensitivity measures, $\partial\mu_Y / \partial\mu_{X_i}$ and $\partial\mu_Y / \partial\sigma_{X_i}$, where μ_Y is the mean of the model output Y and σ_{X_i} is the standard deviation of the input variable X_i. These two sensitivity measures are demonstrated using the U.S. Nuclear Regulatory Commission's total-system performance assessment model, for evaluating the proposed repository at Yucca Mountain. Based on $\partial\mu_Y / \partial\mu_{X_i}$, fifteen out of 330 variables are identified as significantly contributing to sensitivities at 95% acceptance limit. Similarly, based on the calculated $\partial\mu_Y / \partial\sigma_{X_i}$, twenty variables are identified as significantly contributing to sensitivities. Because of the large variability in the performance, approximately 700 samples are needed for the ranking of the variables to be stabilized.

KEYWORDS

Sensitivity Analysis, Uncertainty Analysis, Risk Assessment, System Modeling, Nuclear Fuel Cycle, and Waste Management

INTRODUCTION

Physics-based probabilistic analysis of engineered and natural systems is emerging as an important tool for studying reliability in addition to field and laboratory tests. However, new challenges exist because highly complicated physics-based models are computationally intensive and involve a large number of parameters. The performance assessment of a high-level radioactive waste (HLW) disposal is an example. The performance assessment model has a large number of input parameters that are described by probability distribution functions representing uncertainty and variability. Sensitivity analysis of the performance assessment model is conducted to explain the variability in the output due to uncertainties in the model (not considered in the paper) and input parameters and to determine the most influential input parameters that control the behavior of the output. Knowledge of the most influential input parameters is important because (among other reasons) it can provide an insight on where more efforts should be devoted to reduce the uncertainties in the output and to significantly improve the understanding of the system.

A variety of sensitivity measures have been used in the literature to identify influential parameters emphasizing different aspects of the input-output relationships. In a recently published article by Mohanty and Wu [1], two sampling-based sensitivity measures in the context of the CDF-sensitivity analysis function were presented. However, the HLW problem requires sensitivity measures that are consistent with the regulatory criteria, such as the peak expected dose for compliance [2]. Two performance-mean-based sensitivity measures, $\partial\mu_Y / \partial\mu_{X_i}$ and $\partial\mu_Y / \partial\sigma_{X_i}$, have been proposed in the past in [3] for importance analysis for HLW applications in which components of the repository are artificially neutralized to identify important components. However, applicability of these measures has not been established in the context of sensitivity analysis.

This paper summarizes the development and application of these two mean-based sensitivity measures. Details of the development of these measures in conjunction with the cumulative distribution function (CDF)-based sensitivity analysis method and their comparison with the previously developed [4] and implemented [1] sensitivity measures is a subject of a future paper. In the following sections, we present a very brief description of the processes involved in the performance assessment model, a brief description of the mean-based sensitivity measures, and the results from the application of these measures to the NRC performance assessment model.

THE PERFORMANCE ASSESSMENT COMPUTER MODEL

Performance assessment models often use a probabilistic approach to propagate uncertainties (sometimes variability) in model parameters, conceptual models, and future system states (i.e., scenario classes). A probabilistic model, as implemented in the NRC TPA code [5], simulates (at the process level) thermal, hydrological, mechanical, and chemical processes of the repository system. This paper uses only the portion of the TPA code that models the most likely scenario. This scenario involves the degradation of waste package (WP) in which high-level waste is disposed in the engineered barrier system (EBS), the release of radionuclides when the water infiltrating the ground surface contacts exposed spent nuclear fuel, and transports the radionuclides through the partially water-saturated geologic medium beneath the repository and subsequently in the saturated zone to a reasonably maximally exposed individual assumed to be located at 20 km down-gradient of the repository [5]. The TPA code estimates dose from released radionuclides during specified time periods (e.g. regulatory compliance period). Input parameters are sampled from assigned probability distributions using Latin Hypercube Sampling (LHS). The code contains 961 input parameters out of which 330 are sampled from specified distribution functions. Several sampled input parameters are specified to have correlation with other parameters.

SENSITIVITY MEASURES

Based on a reliability sensitivity concept [4], the response CDF is defined as the integral of the joint probability-density-function of the parameters, with a domain of integration that corresponds to the domain of the identified samples. The response CDF sensitivities are then calculated from the derivatives of the probability integral. The derivatives are statistically estimated from the samples and used to identify and rank the importance of the random variables.

The CDF of a performance $Y = Y(X)$ can be represented as:

$$p = F_Y(y_0) = P(Y < y_0) = \int_\Omega \dots \int f_X(x)dx \quad (1)$$

where Ω is the region of X for $Y(X) < y_0$. From Eq. 1, the sensitivity of p with respect to a distribution parameter θ (e.g., mean or standard deviation) can be formulated as:

$$\frac{\partial p / p}{\partial \theta / \theta} = \int_\Omega \dots \int \frac{\theta \partial f_X}{f_X \partial \theta}\left(\frac{f_X}{p}\right)dx \quad (2)$$

in which (f_X / p) is the sampling density function that corresponds to the sampling region Ω. By applying Eq. 2 for a number of different percentiles, the sensitivities for the entire CDF of Y can be estimated from random samples. Two CDF sensitivities, the standard-deviation sensitivity, $S_{\sigma_i} = (\partial p / p)/(\partial \sigma_i / \sigma_i)$, and the mean sensitivity, $S_{\mu_i} = (\partial p / p)/(\partial \mu_i / \sigma_i)$, were developed in [4] and implemented in [1]. Parameters μ_i and σ_i are the mean and the standard deviation, respectively, of the random variable X_i.

New Mean Response-Based Sampling Sensitivity Measures

Other sensitivity measures proposed for HLW applications include two performance mean-based measures $\partial \mu_Y / \partial \mu_{X_i}$ and $\partial \mu_Y / \partial \sigma_{X_i}$. The sampling-based methods for estimating these two sensitivities have been derived and a summary is given herein. More detailed derivations will be published in a future paper.

The variable transformation is used to transform X_i to Z_i. This transformation can be expressed as

$$\frac{Z_i - \mu_{Z_i}}{\sigma_{Z_i}} = \Phi^{-1}(F_{X_i}(x_i)) = u_i \quad (3)$$

where Z_i is a normal variable with mean value of $\mu_{Z_i} = 0$ and standard deviation of $\sigma_{Z_i} = 1$. Sensitivities with respect to the original variables can be expressed as:

$$\frac{\partial \mu_Y}{\partial \mu_{X_i}} = \frac{\partial \mu_Y}{\partial \mu_{Z_i}} \cdot \frac{\partial \mu_{Z_i}}{\partial \mu_{X_i}} \quad (4)$$

$$\frac{\partial \mu_Y}{\partial \sigma_{X_i}} = \frac{\partial \mu_Y}{\partial \sigma_{Z_i}} \cdot \frac{\partial \sigma_{Z_i}}{\partial \sigma_{X_i}} \quad (5)$$

In Eqs. 3-4, $\partial \mu_{Z_i} / \partial \mu_{X_i}$ and $\partial \sigma_{Z_i} / \partial \sigma_{X_i}$ are calculated numerically or analytically based on Eq. 3. The sensitivities $\partial \mu_Y / \partial \mu_{Z_i}$ and $\partial \mu_Y / \partial \sigma_{Z_i}$ are calculated from the random samples as described below.

$\partial \mu_Y / \partial \mu_{Z_i}$ *Sensitivity from Random Samples*

After the transformation using Eq. 3, the mean value of Y is:

$$\mu_Y = \int Y \phi_u(\boldsymbol{u}, \mu_Z, \sigma_Z)\, d\boldsymbol{u} \quad (6)$$

in which ϕ_u is the joint standard normal pdf. The mean-based sensitivity is (several intermediate steps are not presented):

$$S_{Y_\mu} = \frac{\partial \mu_Y}{\partial \mu_{Z_i}} = \int_{All\ \boldsymbol{u}} Y(\boldsymbol{u}) \frac{\partial \phi(, \mu_Z, \sigma_Z)}{\partial \mu_{Z_i}} d\boldsymbol{u} = E[u_i Y(\boldsymbol{u})] \quad (7)$$

To distinguish if the sensitivity is statistically significant or not, we can test the hypothesis that $S_{Y_\mu} = 0$ and develop the acceptance limits. The test statistics is

$$Z_o = \frac{\overline{S}_{Y_\mu} - S_{Y_\mu}(=0)}{\sigma_{\overline{S}_{Y_\mu}}} \tag{8}$$

in which the sampling estimate is

$$\overline{S}_{Y_\mu} = \frac{1}{k}\sum_{j=1}^{k}[u_i Y]_j \tag{9}$$

Using normal distribution approximation, justified for sufficiently large k based on the central limit theorem, the following probability statement can be made:

$$P\left[-Z_{\alpha/2} \le \frac{\overline{S}_{Y_\mu} - S_{Y_\mu}}{\sqrt{E\left[Y^2\right]/k}} \le Z_{\alpha/2}\right] \le 1-\alpha \tag{10}$$

where $E[Y^2]$ can be estimated using the Monte Carlo or LHS samples. α is the significant probability level or the risk of making a wrong conclusion about the null hypothesis that u is unrelated to the performance Y and has zero sensitivity.

$\partial\mu_Y / \partial\sigma_{Z_i}$ *Sensitivity from Random Samples*

The mean-based sensitivity is:

$$S_{Y_\sigma} = \frac{\partial\mu_Y}{\partial\sigma_{Z_i}} = \int Y(\boldsymbol{u})\frac{\partial\phi(\boldsymbol{u},\mu_Z,\sigma_Z)}{\partial\sigma_{Z_i}}d\boldsymbol{u} = E\left[(u_i^2-1)Y(\boldsymbol{u})\right] \tag{11}$$

To test the hypothesis that $S_{Y_\sigma} = 0$, the test statistics is

$$Z_o = \frac{\overline{S}_{Y_\sigma} - S_{Y_\sigma}(=0)}{\sigma_{\overline{S}_{Y_\sigma}}} \tag{12}$$

in which the sampling estimate is

$$\overline{S}_{Y_\sigma} = \frac{1}{k}\sum_{j=1}^{k}\left[(u_i^2-1)Y\right]_j \tag{13}$$

Using the normal distribution approximation, the following probability statement can be made:

$$P\left[-Z_{\alpha/2} \le \frac{\overline{S}_{Y_\mu} - S_{Y_\mu}}{\sqrt{2\cdot E\left[Y^2\right]/k}} \le Z_{\alpha/2}\right] \le 1-\alpha \tag{14}$$

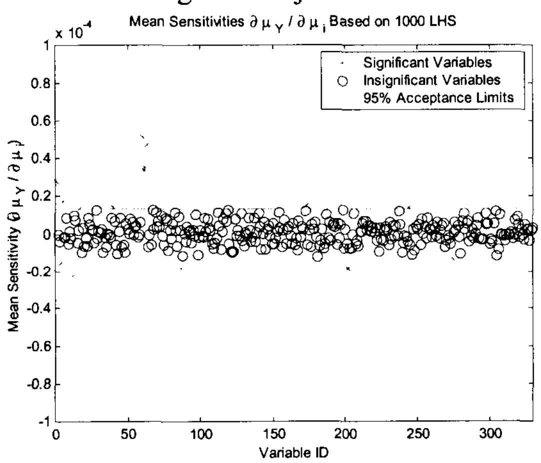
where $E[Y^2]$ can be estimated using the samples.

Acceptance Limits and Adaptive Sampling

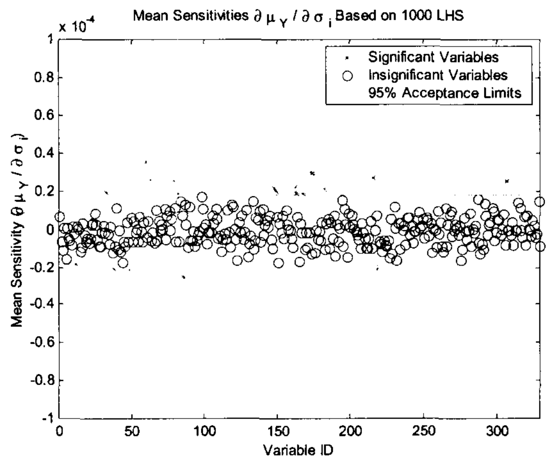
If the calculated sensitivities are outside of the acceptance limits defined by Eqs. 10 or 14, we will accept the alternative hypotheses that the sensitivities are greater than zero at the corresponding confidence level. If the calculated point lies well outside of the limits, then the variable is likely to be important. In such cases, the magnitudes of the sensitivities may be used to rank the important variables. The number of samples can be adaptively increased to reduce the sampling error and to identify the important variables and their ranking with confidence.

RESULTS

Figure 1 shows the calculated sensitivities from 1000 LHS samples and the nominal case 10,000-yr compliance period response (peak dose) calculations using the TPA code. Based on $\partial\mu_Y / \partial\mu_{Z_i}$, 15 variables (corresponding to the data that are outside the acceptance limits) are identified as having significant sensitivities at α = 5%. Similarly, based on $\partial\mu_Y / \partial\sigma_{Z_i}$, 20 variables are identified as significant at α = 5%. The identified important variables are listed in table 1. The results show that the two sensitivity measures produce substantially different set of influential variables. But, when these two measures are applied to the previous version of the TPA code, the difference between the two sets of influential variables is small. Therefore, we believe that the difference between these two measures when applied to the latest version of the TPA code is a result of the new process models and the associated parameter ranges. A formal validation study is currently underway to ensure that the differences are logical and justified.



(a)



(b)

Figure 1. Influential variables identified by (a) S_{Y_μ} and (b) S_{Y_σ} sensitivities (see table 1 for top 10)

Table 1. Top ten random variables identified by S_{Y_μ} and S_{Y_σ} sensitivity

S_{Y_μ} or ($\partial\mu_Y / \partial\mu_{Z_i}$) sensitivity		S_{Y_σ} or ($\partial\mu_Y / \partial\sigma_{Z_i}$) sensitivity	
Rank	Variable Name	Rank	Variable Name
1	WastePackageFlowMultiplicationFactor	1	WastePackageFlowMultiplicationFactor
2	Preexponential_SFDissolutionModel2	2	MatrixKD_UFZ_Ra[m3/kg]
3	DefectiveFractionOfWPs/cell	3	MatrixKD_CHnvPb[m3/kg]
4	SubAreaWetFraction	4	FracturePorosity_TSw_
5	ArealAvgMeanAnnualInfiltrationAtStart[mm/yr]	5	Preexponential_SFDissolutionModel2
6	DripShieldFailureTime[yr]	6	KD_Soil_Se[cm3/g]
7	SFWettedFraction_SEISMO1_7	7	SFWettedFraction_FAULTO
8	MatrixPermeability_TSw_[m2]	8	SFWettedFraction_SEISMO1_6
9	FractionOfCondensateTowardRepository[1/yr]	9	MatrixKD_CHnzTh[m3/kg]
10	FractionOfCondensateRemoved[1/yr]	10	MatrixKD_CHnzU[m3/kg]

The mean sensitivity is expected to stabilize as the number of samples is increased. Figure 2 shows that the ranking convergence seems to become stabilized as the number of samples exceeds about 700.

More samples will be generated to confirm the convergence. Several parameters that are known to have very little significance show up in table 1 (Rank 7 for S_{Y_μ} sensitivity), but this variable drops out as the number of samples is increased from 1000 to 2000. Investigation continues to address this issue.

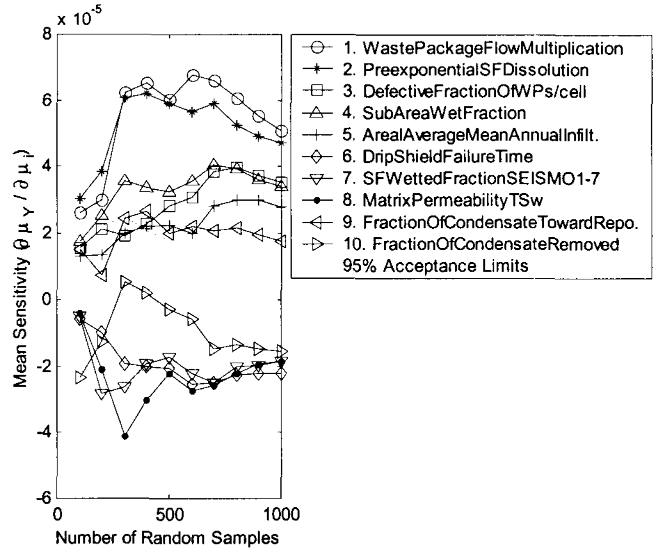


Figure 2. Mean sensitivity of performance to top 10 variables as a function of sample size

CONCLUSIONS

The development and successful implementation of two performance-mean-based sensitivity (or uncertainty importance) measures, $\partial\mu_Y / \partial\mu_{X_i}$ and $\partial\mu_Y / \partial\sigma_{X_i}$, that are particularly relevant to the disposal of HLW regulatory criteria are summarized. Based on $\partial\mu_Y / \partial\mu_{X_i}$ and $\partial\mu_Y / \partial\sigma_{X_i}$ sensitivities, fifteen and twenty out of 330 variables are identified as having significant sensitivities at 95% acceptance limit. Further studies are underway to determine the reason for significant differences in the list of influential variables identified through these two mean-based measures. It appears that 700 samples are sufficient for obtaining stable results at 95% confidence limit for the S_{Y_μ} sensitivity.

ACKNOWLEDGMENTS

The paper was prepared to document work performed by the Center for Nuclear Waste Regulatory Analyses (CNWRA) for the Nuclear Regulatory Commission (NRC) under Contract No. NRC-02-97-009. The activities reported here were performed on behalf of the Office of Nuclear Material Safety and Safeguards (NMSS). The paper is an independent product of the CNWRA and does not necessarily reflect the views or regulatory position of the NRC.

REFERENCES

[1] Mohanty, S. and Y-T. (Justin) Wu. 2001. CDF sensitivity analysis technique for ranking influential parameters in the performance assessment of the proposed high-level waste repository at Yucca Mountain, Nevada, USA, *Reliability Engineering and System Safety*, Vol 73/2, pp 167-176.

[2] U.S. Nuclear Regulatory Commission, 10 CFR Part 19, Disposal of High-Level Radioactive Wastes in a Proposed Geological Repository at Yucca Mountain, Nevada, Part 63, Washington, DC, U.S. Nuclear Regulatory Commission, 2001.

[3] Eisenberg NA, Sagar B. Importance Measures for Nuclear Waste Repositories. Reliability Eng. Safety. 70, 217-239: 2000.

[4] Wu Y-T. Computational methods for efficient structural reliability and reliability sensitivity analysis. AIAA J. 32(8), 1717–1723: 1994.
[5] Mohanty S, McCartin TJ (coordinators). NRC Sensitivity and Uncertainty Analyses for a Proposed HLW Repository at Yucca Mountain, Nevada Using TPA 3.1—Volume I: Conceptual Models and Data. NUREG–1668. Washington, DC: US Nuclear Regulatory Commission: 2001.