CNWRA 92-002

SENSITIVITY AND UNCERTAINTY ANALYSES APPLIED TO ONE-DIMENSIONAL TRANSPORT IN A LAYERED, FRACTURED ROCK PART 2: EVALUATION OF LIMIT STATE APPROACH

Iterative Performance Assessment- Phase II

Prepared for

Nuclear Regulatory Commission
Contract NRC-02-88-005

Prepared by

Center for Nuclear Waste Regulatory Analyses
San Antonio, Texas

March 1992

462.2 --- T199810140006
Sensitivity and Uncertainty
Analyses Applied to
One-Dimensional Transport in

CNWRA 92-002

SENSITIVITY AND UNCERTAINTY ANALYSES APPLIED TO ONE-DIMENSIONAL TRANSPORT IN A LAYERED FRACTURED ROCK

Part 2: Evaluation of the Limit State Approach

Iterative Performance Assessment - Phase II

Prepared for

Nuclear Regulatory Commission
Contract NRC-02-88-005

Prepared by

Y.-T. Wu
A. B. Gureghian
B. Sagar

Center for Nuclear Waste Regulatory Analyses
San Antonio, Texas

and

R. B. Codell

Nuclear Regulatory Commission
Office of Nuclear Materials Safety & Safeguards

March 1992

TABLE OF CONTENTS

Page

LIST OF FIGURES iv
LIST OF TABLES vi
ACKNOWLEDGMENTS vii
EXECUTIVE SUMMARY viii

1 INTRODUCTION 1-1

2 DEFINITIONS OF UNCERTAINTY AND SENSITIVITY ANALYSES 2-1

3 REVIEW OF CURRENT PROBABILISTIC METHODS FOR PERFORMANCE ASSESSMENT 3-1

3.1 MONTE CARLO METHOD 3-1
3.2 RESPONSE SURFACE APPROACH 3-2
3.3 DIFFERENTIAL ANALYSIS APPROACH 3-2

4 ALTERNATIVE PROBABILISTIC APPROACH BASED ON LIMIT STATE FORMULATION 4-1

4.1 INTRODUCTION 4-1
4.2 LIMIT STATE FORMULATION 4-1
4.3 APPROXIMATION POINT FOR DEVELOPING APPROXIMATE LIMIT STATE SURFACE 4-3
4.4 FAST PROBABILITY INTEGRATION 4-6
4.5 ITERATION METHODS FOR MOST PROBABLE POINT SEARCH 4-6
4.6 PROBABILISTIC SENSITIVITY FACTORS 4-7
4.7 CORRELATED RANDOM VARIABLES 4-10
4.8 MPP SEARCH AND CDF ANALYSIS BY THE ADVANCED MEAN VALUE (AMV) METHOD 4-12
4.9 PROBABILITY UPDATING BY ADAPTIVE IMPORTANCE SAMPLING METHOD 4-13
4.10 SUMMARY OF THE LIMIT STATE APPROACH 4-16

TABLE OF CONTENTS (Continued)

Page

5 APPLICATION EXAMPLE 5-1

5.1 DESCRIPTION OF THE RADIONUCLIDE TRANSPORT MODEL 5-1
5.2 APPLICATION OF THE LIMIT STATE APPROACH 5-1
5.3 COMPARISON OF LIMIT STATE APPROACH WITH OTHER APPROACHES 5-7

6 SUMMARY AND DISCUSSIONS 6-1

7 REFERENCES 7-1

APPENDICES

A CDF Analysis by the Advanced Mean Value (AMV) Procedure - Illustrative Example A-1

B Sensitivity Updating Algorithm for Problems with Large Number of Input Random Variables B-1

LIST OF FIGURES

No.		Page
4-1	Limit state surfaces for v_i = -5, 0, and 1	4-2
4-2	Transformation and most probable point	4-4
4-3	Limit state surfaces and most probable points in the transformed u-space for velocity v_i = -5, 0, and 1	4-5
4-4	Most probable point locus (MPPL)	4-8
4-5	Limit state surfaces and most probable point locus	4-9
4-6	The definition of sensitivity factors	4-10
4-7	Sensitivity factors for velocity v_i = -5, 0, and 1	4-11
4-8	MPP search by the AMV method	4-14
4-9a	MPPL tracking points	4-15
4-9b	Calculated CDF	4-15
4-10	The concept of the curvature-based adaptive importance sampling method	4-17
4-11	Calculation of P $\{V > 1\}$ by the Adaptive Importance Sampling Method	4-18
5-1	Vertical cross-section of a layered rock system intersected by a vertical fracture subject to surface ponding	5-2
5-2	Description of migration pathways in a system of homogeneous layers of fractured rock	5-3
5-3	Cumulative mass of Np-237 per unit in the fracture versus time at different positions x = 100, 200, and 500 meters (exponentially decaying source and band release mode)	5-5
5-4(a)	Sensitivity of cumulative mass to pore diffusivity versus time for Np-237 (exponentially decaying source)	5-6

LIST OF FIGURES (Continued)

No.		Page
5-4(b)	Sensitivity of cumulative mass to distribution coefficient in rock versus time for Np-237 (exponentially decaying source)	5-6
5-5	Comparison of limit state approach with Monte Carlo	5-7
5-6	CCDF results using Latin Hypercube sampling	5-8
5-7	Comparisons of the AMV method with the simulation methods	5-9
5-8	Comparisons of the regression model with the exact model	5-13
5-9	Comparisons of the differential analaysis model with the exact model	5-14
A-1	Most probable point locus (MPPL) of Z_{MV}	A-3
A-2	CDF's of Z_{MV} and Z_{AMV}	A-5

LIST OF TABLES

No.		Page
5-1	Input parameters for the radionuclide transport problems	5-4
5-2	Comparison of sensitivities from regression to differential analysis (50 LHS samples)	5-10
5-3	Comparison of sensitivities from regression to differential analysis (300 LHS samples)	5-12
5-4	Sensitivity coefficients from regression and differential analysis	5-15
5-5	Probabilistic sensitivity factors	5-17

ACKNOWLEDGMENTS

Several people assisted in the preparation of this report. The authors would like to express their appreciation for their efforts which greatly helped the document to reach its final form. In particular, we would like to thank Drs. W. C. Patrick and P. K. Nair for their reviews.

EXECUTIVE SUMMARY

Applicability of the Limit State method, which was developed initially for structural reliability analyses, to performance assessment of geologic repositories is investigated. This study was undertaken as an auxiliary analysis for the Iterative Performance Assessment (IPA), Phase II project.

The Limit State method belongs to a class of methods that deal with uncertainty propagation and sensitivity analysis of system response that depends upon a multitude of parameters, some of which are random. Other such methods applied to repository performance assessment are the Monte Carlo, Differential Analysis, Response Surface, and Regression Analysis. The Limit State method has found considerable application in reliability estimation of structural components, but is relatively new to the repository performance assessment field.

The Limit State approach is based on partitioning the parameter space into two parts: one in which the performance measure is smaller than a chosen value (called the limit state), and the other in which it is larger. Through a Taylor expansion at a suitable point, the partitioning surface (called the limit state surface) is approximated as either a linear or quadratic function. The success and efficiency of the limit state method depends upon choosing an optimum point for the Taylor expansion. The point in the parameter space that has the highest probability of producing the value chosen as the limit state is optimal for expansion. When the parameter space is transformed into a standard Gaussian space, the optimal expansion point, known as the Most Probable Point (MPP), has the property that its location on the Limit State surface is closest to the origin. Additionally, the projections onto the parameter axes of the vector from the origin to the MPP are the sensitivity coefficients. Once the MPP is determined and the Limit State surface approximated, formulas (see Equations 4-7 and 4-8) are available for determining the probability of the performance measure being less than the limit state. By choosing a succession of limit states, the entire cumulative distribution of the performance measure can be determined. Methods for determining the MPP and also for improving the estimate of the probability are discussed in this report.

The Limit State method is obviously more complex than the more commonly used Monte Carlo method. To aid understanding, all steps of the method are explained by applying them to a simple example of obtaining the cumulative probability distribution function of Darcy velocity, V, given by $V = -KI$, where K and I are the hydraulic conductivity and hydraulic gradient, respectively. Although simple, this example turned out to be not so easy for the application of the Limit State method because of the possibility of change of sign of I (and hence of V). All steps of this example are boxed in the text for easy identification.

Finally, the method is applied to a one-dimensional transport problem and compared to the Monte Carlo method. Sensitivity coefficients were computed and compared with those obtained from differential analysis.

The evaluation of the Limit State method documented in this report indicates that it has merit. It is recommended that it be evaluated next by applying it to a more complex problem, such as the total system performance assessment which may have several modules linked together. It appears that some developmental work regarding efficient determination of the MPP and estimation of the derivatives of the performance measure required for Taylor expansion will also be needed.

1 INTRODUCTION

In assessing high-level waste (HLW) repository performance, mathematical models will be used to estimate system behavior in response to future perturbations, including the release of radionuclides into the geosphere. Performance assessment models are based on a combination of scientific principles and empirical data. It can be expected that, because of inherent variabilities in the model inputs due to the difficulties in obtaining relevant field data and in validating the models, uncertainties will exist in the model input parameters as well as in the description of processes included in the models. Therefore, the predicted response or performance will also be uncertain.

One way to address the uncertainties is to use probabilistic sensitivity analysis approaches which assess the variabilities in the response due to input or model uncertainties. There are several well-known uncertainty and sensitivity analysis methods that can be used. The most common of these analysis approaches generally depend upon repeated evaluation of the response for different combinations of input parameters. However, for complex numerical models generally applicable to HLW performance assessment, these methods may demand excessive computational effort or may not be sufficiently accurate. The Center for Nuclear Waste Regulatory Analyses (CNWRA) and the Nuclear Regulatory Commission (NRC) are exploring an alternative uncertainty and sensitivity analysis approach that may save computational effort without losing significant accuracy. This investigation is a part of the Iterative Performance Assessment (IPA) effort to develop HLW system performance assessment capabilities within NRC and CNWRA. A series of two reports regarding uncertainty and sensitivity analysis methodologies has been developed. In Part 1 [Gureghian, et al. (1991)], a mathematical model and a computer code for one-dimensional radionuclide transport in a layered fracture rock were developed. This report employs this model to evaluate the alternative sensitivity and uncertainty approach.

In the context of this report, uncertainty analysis means probabilistic analysis in which the uncertain input parameters are treated as random variables with assigned probability distributions. The goal of the uncertainty analysis is to compute the cumulative probability [or cumulative distribution function, (CDF)] of one or more responses or performance measures. Sensitivity analysis, on the other hand, refers to the estimation of change in the performance measure if any one of the input variables is varied. Probabilistic sensitivities are defined as variations in the performance CDF because of a change in an input variable.

For problems related to repository performance assessment, applicable probabilistic methods include the widely-used Monte Carlo method, the response surface method, and the differential analysis method. These methods are well documented in the literature (e.g., see Helton et al. 1991). The goal of this study is to investigate an alternative approach which has a relatively short history of development. The basic idea of this alternative approach was originally developed to solve problems encountered in the field of mechanical and structural reliability design and analysis especially when high reliability (or low probability-of-failure) was of particular interest. More recently, the ideas have been extended to develop full CDF of performance measures. However, the alternative approach addressed in this report has not been

sufficiently investigated for its applicability to repository system performance assessment problems that are characterized by large uncertainty, large number of random variables, and complicated nonlinear performance functions.

Traditionally, sensitivity analysis is based on the perturbation of a system parameter at a "reference" point, typically the mean-value. In this report, the meaning of sensitivity analysis will be extended to include probabilistic sensitivity, e.g., the sensitivity of a probability-based measure with respect to a system parameter, evaluated at a performance value of interest that may not correspond to the mean-value of the parameter. The alternative approach provides a framework to determine an optimum reference point in the parameter space and for performing probabilistic sensitivity analysis based on such an optimum point.

The alternative approach is evaluated by considering: (1) the efficiency and accuracy of the approach when the models are complicated and require extensive computations; (2) the feasibility of the method when the numerical models cannot be modified easily and must be treated as "black-boxes"; and (3) the efficiency and accuracy of the method when problems involve a large number of random variables.

Section 2 of the report defines uncertainty and sensitivity analyses and discusses different methods of computing sensitivities. Section 3 provides a summary of the more established methods. Section 4 describes the alternative approach. It provides a review of the fundamental concept as well as a scheme that was developed during the development of this report to support the alternative approach. In Section 5 several methods are applied to the transport model example for illustrating and comparing different methods. The advantages and limitations of the alternative approach are summarized in Section 6. Cited references are given in Section 7. In addition to the general description of the alternative method, some detailed technical discussions are given in the appendices to supplement Section 4.

2 DEFINITIONS OF UNCERTAINTY AND SENSITIVITY ANALYSES

In this report, uncertainty analysis is defined as being equivalent to probabilistic analysis in which the uncertain input variables are treated as random variables with assigned probability distributions. The goal of the uncertainty analysis is to compute the CDF or the complementary CDF (CCDF = 1 - CDF) of a performance measure such as cumulative radionuclide release.

The term "sensitivity" has several meanings. For distinguishing the different meanings, this report defines sensitivity as $dZ/d\alpha$ where dZ is the change in a measure Z related to system response or performance, R, and $d\alpha$ is the change in a system parameter α. In general, Z could be R itself, or a characteristic of R such as the CDF of R, $F_R(r)$. With α deterministic, $dR/d\alpha$ can, in general, be calculated by obtaining R at a "reference" value $\bar{\alpha}$, and then at a perturbed value $\bar{\alpha} \pm d\alpha$. In this case, $dR/d\alpha$ will be called a "deterministic" sensitivity.

In probabilistic performance assessment when the response R is a random variable (or vector), the usefulness of $dR/d\alpha$ is limited since the expression does not include input uncertainty information. The dimensionless "sensitivity coefficient," $[dR/R]/[d\alpha/\alpha]$ discussed in Gureghian, et al (1991), which represents the percentage change in the response relative to the percentage change in a parameter is more useful for identifying "important" parameters. However, the sensitivity coefficient still does not include parameters that are related to probability distributions; hence, it is also a deterministic sensitivity.

A commonly used "probabilistic" sensitivity measure that relates to input and response probability distributions, in terms of the standard deviations, is $d\sigma_R/d\sigma_\alpha$. This measure is useful because it provides information on the degree of response uncertainty (or variability) reduction as a function of reduction of the input uncertainty. Similarly, a probabilistic measure associated with a change in the mean value is useful in examining the impact of the uncertainty in the mean value of an input random variable. Probabilistic sensitivity measures can be used to identify key parameters that contribute to the uncertainty in the performance measure. Relative importance or ranking of input random parameters may be used to assist in making decisions regarding uncertainty reduction.

More generally, in probabilistic performance assessment, when there is a change in the system parameter α, either deterministic or random, the CDF of R will change (e.g., its mean and standard deviation will change). Thus, it is reasonable to define "probabilistic" sensitivity as the derivative of the cumulative probability, $F_R(r)$, with respect to a parameter, α. The probability sensitivity, $dF_R(r)/d\alpha$, can be evaluated either at a specified probability, P, or at a specified response, r. Here, the parameter α is related to the input (deterministic or random) variables. For example, α could be a system parameter or a mean value or a standard deviation. Other probabilistic sensitivity measures could also be defined. In a more general sense, a probabilistic sensitivity analysis can be performed with respect to a change in the distribution type (e.g., Normal to Lognormal distribution). This type of sensitivity analysis may be applied for situations where there are several competing probability distribution models. It will be

shown in Section 4 that the alternative approach provides a reasonable way to compute probabilistic sensitivity measures.

The most direct method for calculating a deterministic or a probabilistic sensitivity is by calculating the performance measure at two levels: reference and perturbed. When each individual solution requires significant computer time, repetitive analyses can put excessive computational burden on the performance assessment. This approach of numerical differentiation requires a minimum of (n + 1) solutions for n input parameters. It does not require much knowledge about the modeling details and is essentially a "black-box" approach.

In a "non-black box" approach, sensitivities could be obtained by analytical means such as the transport model discussed in Part 1 (Gureghian, et al. 1991). Analytical sensitivity provides accurate and possibly faster calculations than the numerical differentiation approach. For some problems, sensitivity could be computed much more efficiently using the adjoint method. The advantages and the limitations of these methods are discussed in a report by Thomas (1982). Another way of computing the sensitivities efficiently is the use of perturbation method as discussed by Dias and Nagtegaal (1985). In general, the adjoint method and the perturbation method might require significant modifications to the original computer codes. Detailed investigation of these methods is outside of the scope of this study.

3 REVIEW OF CURRENT PROBABILISTIC METHODS FOR PERFORMANCE ASSESSMENT

Commonly used probabilistic methods for performance assessment include the Monte Carlo method, the response surface method, and the differential analysis method. These methods are well explored (Helton, et al 1991, and Karamchandani 1987), and are briefly reviewed below.

3.1 MONTE CARLO METHOD

This method generates K realizations of n input parameters, $X_1,\ldots,X_n$, from specified probabilistic distribution information. These K realizations are then fed into the model to compute K responses or performances. The K response values are synthesized to derive probabilistic information including the CDF. Several sampling techniques are available.

In *random sampling*, each sample of the parameter vector is obtained from the entire joint probability distribution of the input random parameters. This sampling is the most straightforward, and is often called the "standard" Monte Carlo method.

In *stratified sampling*, the domain of $X_1,\ldots,X_n$ is divided into regions of (typically) equal probability, and then from each region a random sample is generated. This approach ensures that samples cover the entire domain. In *Latin Hypercube sampling* (LHS), the range of each input variable is divided into intervals of equal probability, and one value based on its marginal distribution is then randomly selected from each interval. For each X_i the K sample values thus obtained are rearranged according to an independent random permutation. The resulting permutations form an n by K matrix, with the nth row consisting of the permuted values of X_i. The K columns of this matrix constitute the Latin Hypercube sample. For correlated input variables, the permutation is based on the specified correlation matrix. As is true for stratified sampling, it can be shown that, under appropriate conditions, estimates based on LHS have smaller variance than estimates based on random sampling.

Monte Carlo method has several advantages. As summarized by Helton et al., (1991), Monte Carlo techniques do not require a large amount of sophistication that goes beyond the analysis problem of interest. Also, they do not require modifications to the original model or additional numerical procedures, and can be used to propagate uncertainties through a sequence of separate models.

The major drawback is that a large number of samples might be required to achieve a result with sufficiently narrow confidence bounds. In particular, when the interests are at the tails of the distribution, the sample sizes will need to be large in order to generate sufficient "hits" in the "corner(s)" of the parameter space that corresponds to extreme performance values.

Probabilistic sensitivity analysis can be performed by combining the Monte Carlo method with the regression analysis. The accuracy of this approach depends on the distribution of the data points used in the regression analysis. In the "corner" region where there are fewer

samples, the answers generated by the regression-based sensitivity analysis might not be accurate. Alternatively, a probabilistic sensitivity analysis can be performed by generating a perturbed solution using a new Monte Carlo analysis in which the input parameters are perturbed (e.g., a standard deviation is redefined). This process, while able to provide accurate results, is clearly computationally more expensive than the regression-based method.

To reduce the computational burden in the above Monte Carlo approaches, an *importance sampling* technique, which forces sampling in the "important" region (e.g., a "corner" region), may be used. The main difficulty is that the important region is not known *a priori*. A good guess will be needed to use this technique.

3.2 RESPONSE SURFACE APPROACH

The response surface approach consists of replacing the original response function Z by a simple analytical function of a limited number N of the initial n input variables, with $N \leq n$. Most often that function is a linear combination of the X_i, $i=1,\ldots, N$:

$$Z \approx a_0 + \sum_{i=1}^{N} a_i X_i \quad . \tag{3-1}$$

In a more complex case, the fitted function may include higher-order terms, e.g.,

$$Z \approx a_0 + \sum_{i=1}^{N} a_i X_i + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \leq i}}^{N} b_{ji} X_j X_i \quad . \tag{3-2}$$

The coefficients a_0, a_i, and b_{ji} are fitted by a regression analysis based on a number of solutions at designed or randomly selected input points. After the response surface model is developed, probabilistic analysis can be performed efficiently because the developed response surface is simple. Thus, for complicated models, the efficiency in using the response surface approach depends mainly on the efficiency in generating the coefficients of the response surface.

In order to obtain accurate information at the tails of the distribution, higher-order response surfaces will be required when the response functions are highly nonlinear. In practice, unless n is small, higher than second-order response surfaces are rarely developed because they require much more data points and because nonlinear regression models may produce large modelling errors in the regions where few or no curve-fitting data are available.

3.3 DIFFERENTIAL ANALYSIS APPROACH

The differential analysis approach consists of approximating the actual function by Taylor's series expansion about a reference point, typically the mean (μ) or median value,

$$Z \approx Z(\mu) + \sum_{i=1}^{n} \frac{\partial Z(\mu_i)}{\partial X_i} (X_i - \mu_i) \quad . \tag{3-3}$$

While not intended for full CDF analysis, the above first-order approximation can be used to provide the estimates of the response mean and standard deviation and to produce the sensitivity coefficients.

The Taylor's series model is more accurate in the neighborhood of the expansion (mean) point while the response surface model is intended to be used for a wider region by optimizing the modelling errors. Similar to the response surface approach, a higher-order Taylor's series expansion (Sagar and Clifton, 1984) may be better, but would require extra computations to obtain higher-order term coefficients.

4 ALTERNATIVE PROBABILISTIC APPROACH BASED ON LIMIT STATE FORMULATION

4.1 INTRODUCTION

In the alternative formulation, a particular value identified as the "limit state" of the performance measure is selected. Having selected a limit state, the probability of the performance measure being less than (or exceeding) the limit is computed. The CDF of the performance measure can be established by selecting a number of limit states, and performing probability analysis at each of these selected limit states.

The limit state approach is based on approximating the performance function at a suitable point in parameter space by a simple expression (linear or quadratic) so that the probabilities can be calculated relatively easily. The efficiency and accuracy of the results depend upon the selection of the point on the limit state surface at which the performance function is simplified. The key step in applying the limit state approach is in determining this approximation point.

In the following, the procedures required to compute the approximation point and the probability estimates efficiently will be described. An adaptive importance sampling approach for improving the approximate probability solutions is also described. A simple example will be used in this section to illustrate the approach. This example, described in Figure 4-1, pertains to the estimation of CDF of Darcy velocity. For easy identification, application of various steps of the limit state approach to this example are enclosed in boxes. The radionuclide transport model developed in Part 1 of this report (Gureghian, et al. 1991) is used in Section 5 to demonstrate the applicability of the limit state approach.

4.2 LIMIT STATE FORMULATION

Define the performance function as: $Z = Z(X_1, X_2, \ldots X_n)$, where X_i are the input random variables. In the limit state formulation, Z is approximated at a selected value (called the limit state), for example, $Z = z$. Numerous combinations of $(X_1, X_2, \ldots X_n)$ that lead to the limit value z constitute a "limit state surface" which can be regarded as the "response surface" but is *constrained* by the condition that $Z = z$. The limit state surface partitions the parameter space into two parts. On one side of the limit state surface, $Z \leq z$, and on the other side $Z > z$. Once this partition is determined, the probability of Z being less than (or exceeding) z is easily computed. A full CDF can be established by selecting a number of z_i, and performing probability analysis at each of these selected z_i.

Equivalently, for each limit state, define,

$$g(\mathbf{X}) = Z(\mathbf{X}) - z = 0 \,, \tag{4-1}$$

which separates the variable space into two sets: "g less than or equal to 0" and "g greater than 0" (this corresponds to "failure" and "safe" in a structural reliability problem). Given the joint

Example: $V = -KI$

Consider a simple flow problem described by Darcy's law, $V = -KI$, where K is the hydraulic conductivity, I is the hydraulic gradient, and V is the Darcy velocity. Assume that both K and I are random variables and that the performance function is the flow velocity V(K,I). The goal of the probabilistic analysis is to compute the CDF of V.

To formulate a limit state, let $V = v_i$. Figure 4-1 shows three limit state surfaces for $v_i = -5$, 0, and 1. Each limit state surface determines two regions: $V \leq v_i$, and $V > u_i$. The goal of the analysis is to compute the probabilities of those two regions and the approach is to simplify the limit state surface using a linear or quadratic surface.

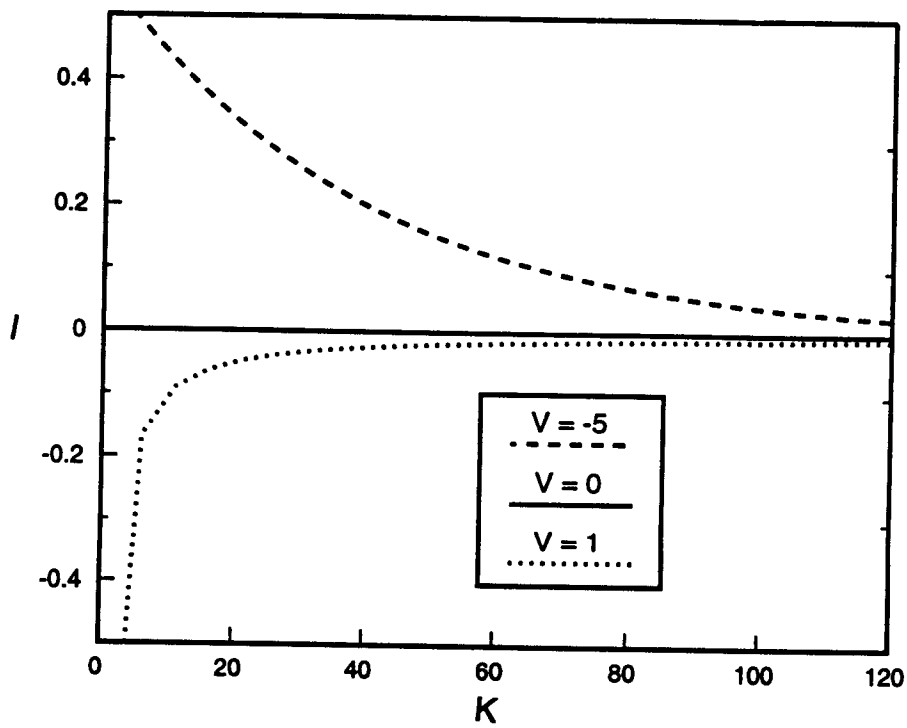

Figure 4-1. Limit state surfaces for v_i = -5, 0, and 1

probability density function, $f_X(x)$, the CDF, $F_Z(z)$, of the performance function, in theory, is given by:

$$F_Z(z) = P[g \le 0] = \int_{g \le 0} \cdots \int f_X(x)\,dx \tag{4-2}$$

In practice, this integral is very difficult to evaluate analytically. For practical purposes, efficient numerical solutions are desired. One approach is to approximate an exact limit state surface by a more simple, but approximate (say linear), limit state surface, such that the above integral can be computed easily.

4.3 APPROXIMATION POINT FOR DEVELOPING APPROXIMATE LIMIT STATE SURFACE

In order to develop an acceptable approximate limit state surface, a good approximation point in the parameter space needs to be identified. Assuming that a first- or second-order Taylor's series will be used to approximate the exact function, ideally a best point can be defined as that point on the limit state surface that minimizes the probability in the region between the exact and the approximate linear (or quadratic) surface. However, such a point is, in general, difficult to find. An easier and widely used procedure requires a transformation of the original problem to a mathematically more tractable problem, as described below.

First, the generally correlated non-normal random variables **X** are transformed into independent, standardized normal vector **u**. One such transformation is the Rosenblatt transformation [Rosenblatt (1952), and Hohenbichler and Rackwitz (1981)]:

$$\begin{aligned} u_1 &= \Phi^{-1}[F_1(x_1)] \\ u_2 &= \Phi^{-1}[F_2(x_2 \mid x_1)] \\ &\cdot \\ &\cdot \\ &\cdot \\ u_n &= \Phi^{-1}[F_n(x_n \mid x_1, x_2, \ldots, x_{n-1})] \end{aligned} \tag{4-3}$$

where $\Phi^{-1}(\cdot)$ represents the inverse standardized normal CDF, and $F_n(x_n \mid x_1, x_2, \ldots x_{n-1})$ denotes the conditional CDF. In practice, if all that is known are the marginal distributions and the correlation matrix of **X**, a transformation can be made to generate a joint normal distribution that gives the same correlation matrix. This topic is discussed in Section 4.7.

In the cases where all the X_i are mutually independent, Eqn. (4-3) reduces to

$$u_i = \Phi^{-1}[F_i(x_i)] \tag{4-4}$$

and the inverse transformation is:

$$x_i = F_{x_i}^{-1}[\Phi(u_i)] \, . \tag{4-5}$$

By using the above inverse transformation, the original g(**X**)-function is transformed to g(**u**)-function. In the u-space, the point closest to the origin called the minimum distance point, as illustrated in Figure 4-2, can be determined. In the u-space, the joint probability density is rotationally symmetric about the origin, and is an exponentially decreasing function of the distance from the origin. Thus, the minimum distance point has the highest probability of producing the value z among all the points on the limit state surface. Therefore, this point is the most-probable-point (MPP), denoted by **u*** in the following. See Figure 4-3 for an example of MPP.

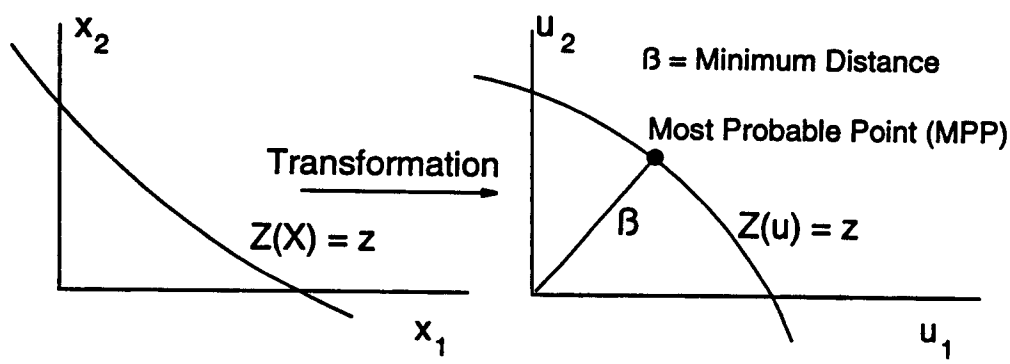

Figure 4-2. Transformation and most probable point

Example: $V = -KI$

Let K be a lognormal random variable with a mean of 13.4 and a standard deviation of 14.4 (i.e., coefficient of variation (COV) is 107 percent), and I be a normal random variable with a mean of 0.05 and a standard deviation of 0.033 (i.e., COV = 66 percent). The median value of K is 9.141. Based on Eqn. (4-5), it can be shown that

$$\ln K = \ln 9.141 + u_1\sqrt{\ln (1 + COV^2)}$$
$$= \ln 9.141 + 0.876\, u_1, \tag{4-6}$$

and

$$I = 0.033u_2 + 0.05 \tag{4-7}$$

Therefore, the transformed limit state is

$$\ln V = \ln v_i = (\ln 9.141 + 0.876\, u_1) + \ln (-0.033u_2 - 0.05) \tag{4-8}$$

The three transformed limit states corresponding to $v_i = -5, 0, 1$ are plotted in Figure 4-3. The **u*** of the three corresponding MPP's are (2.16, 0.984), (0., -1.515), and (1.622, -2.31), respectively.

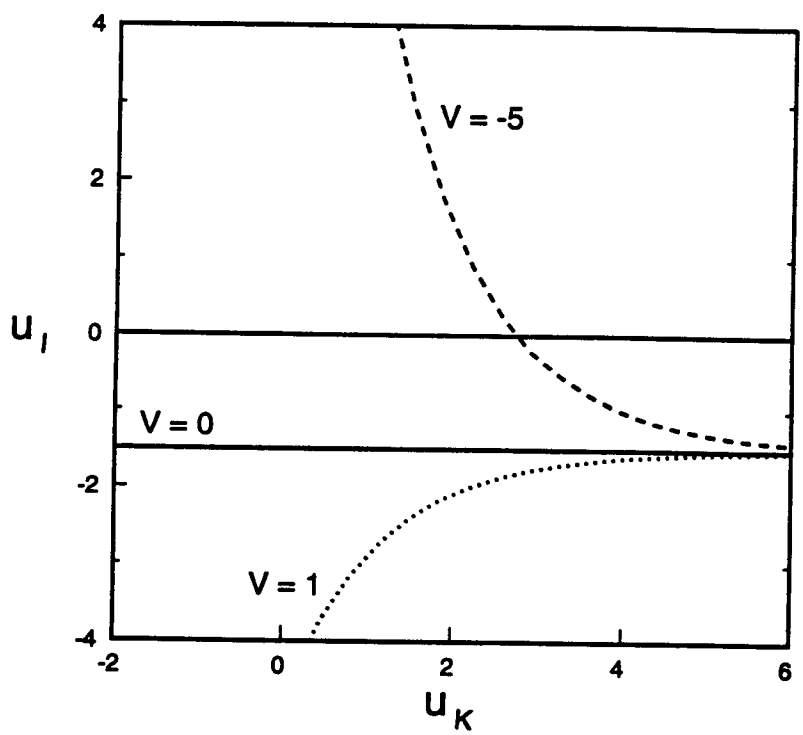

Figure 4-3. Limit state surfaces and most probable points in the transformed u-space for velocity v_i = -5, 0, and 1

4.4 FAST PROBABILITY INTEGRATION

Assuming that the MPP is found, the g(**u**) or g(**X**) function is then approximated by a linear or quadratic polynomial function and the associated probability can be computed relatively easily.

When the g(**u**) formulation is used, several analytical solutions are available for linear and quadratic approximations. For example, with β as the distance of the MPP from the origin, the First-Order Reliability Method (FORM) estimate [Ang and Tang (1984), and Madsen et al. (1986)], which requires only the minimum distance, is:

$$P(g \le 0) \approx \int_{g \le 0} \cdots \int f_U(\mathbf{u})d\mathbf{u} = \Phi(-\beta) \tag{4-9}$$

where $\Phi(\cdot)$ is the CDF of a standard normal distribution. The asymptotic, Second-Order Reliability Method (SORM) estimate [Breitung (1984), and Breitung (1989)] is:

$$P(g \le 0) \approx \int_{g \le 0} \cdots \int f_U(\mathbf{u})d\mathbf{u} = \Phi(-\beta) \prod_{j=1}^{n-1} (1 + \beta \kappa_j)^{-1/2}, \text{ (Exact as } \beta \to \infty) \tag{4-10}$$

which requires β as well as κ_j, j = 1, ...n - 1, the main curvatures (Fiessler et al., 1979) of the limit-state surface at the MPP **u***. A rather complete treatment of probability analysis based on quadratic functions can be found in Tvedt (1990). It should be noted that in the above two equations, the minimum distance is always positive, meaning the calculated probability (usually < 0.5) is always associated with the region which does not contain the origin. Thus, the computed results might be either CDF or CCDF.

When the g(**X**) formulation is used (Wu and Wirsching, 1984), a numerical procedure based on the convolution theorem and Fast Fourier Transformation has also been developed for efficient first-order and second-order probability analysis (Wu and Torng, 1990).

Example: V = -KI

Referring to Figure 4-3, the minimum distances corresponding to v_i = -5, 0, and 1 are 2.372, 1.515, and 2.826. The corresponding probability estimates are: P {V < -5} = Φ(-2.372) = 0.0088, P {V < 0} = Φ(1.515) = 0.935, and P {V < 1} = Φ(2.826) = 0.9976.

4.5 ITERATION METHODS FOR MOST PROBABLE POINT SEARCH

The MPP, **u***, can be found by using an optimization procedure based on the following formulation:

Minimize: Distance $\beta = \sqrt{u_1^2 + u_2^2 + \ldots + u_n^2}$

Subject to: $g(\mathbf{u}^*) = 0$

The MPP can also be found by methods that involve iterations in which a sequence of successive approximation are generated by updating the *g*-function and its gradients (Liu and Der Kiureghian 1986). For example, a widely used iteration method was proposed in Rackwitz and Fiessler (1978). This method has several other versions [Ang and Tang (1984), Thoft-Christensen and Baker (1982), and Madsen et al. (1986)] with subtle differences. The advanced mean-value (AMV) method, which will be discussed in Section 4.8, is another iteration algorithm.

There is no single algorithm that can be identified as universally superior. In fact, in using an optimization program a poor initial guess point may lead to a local, rather than a global extremum. At the same time, in using an iterative procedure, slow or unstable convergence may occur for some problems. Two approaches are recommended to help alleviate these problems.

For problems in which many local minimums are present, a checking procedure such as the adaptive importance sampling method (explained in Section 4.9) can be used to validate (or invalidate) the approximate solution. Problems in which convergence is slow typically occur when the dominant random variables switch from one input parameter region to another. In such problems a linear approximation function is good only in a local region and extrapolation results in poor iteration performance. To resolve this problem, the following strategy based on the most-probable-point-locus (MPPL) concept (Wu et al. 1988) might be useful.

The MPPL is defined by connecting all the MPP's for all the performance levels, as illustrated in Figure 4-4. A unique property of MPPL is that it always passes through the origin, which is because the minimum distance is zero for the performance evaluated at the origin. This point is usually used as a starting point in an MPP search procedure. However, if one is interested in a limit state that is far away from the origin, then the origin may not be a good starting point. To ensure a robust iteration procedure, one can start from the origin and keep track of the MPPL by using a small increment of performance measure until the performance reaches the desired limit state. The success of this approach depends upon the choice of a suitable increment in discretizing the performance measure.

4.6 PROBABILISTIC SENSITIVITY FACTORS

Under the limit state approach framework, first-order sensitivity factors which provide approximate relative importance of the random variables are produced as a by-product. This is a major advantage of the limit state approach. The *probabilistic* sensitivity factors (referred to as sensitivity factors from now on) are defined as follows [Ang and Tang (1984), Thoft-Christensen and Baker (1982), and Madsen et al. (1986)].

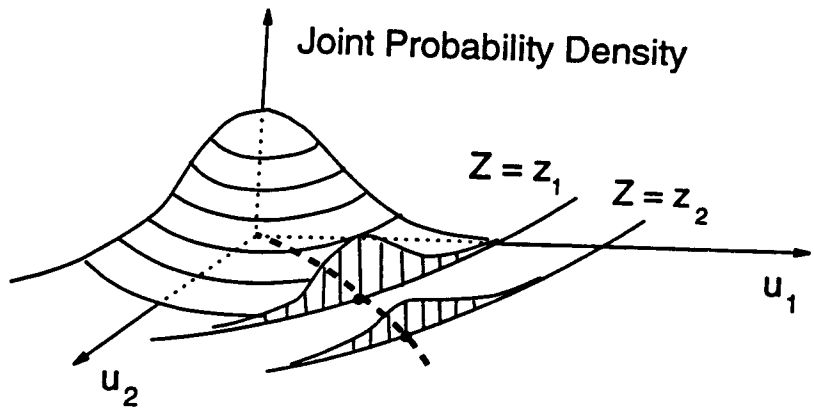

Figure 4-4. Most probable point locus (MPPL)

At the most probable point $\mathbf{u}^* = (u_1^*, u_2^*, ..u_n^*)$, the first-order probability estimate is simply $\Phi(-\beta)$ (Eqn. 4-9) where

$$\beta^2 = u_1^{*2} + u_2^{*2} + \ldots u_n^{*2} . \tag{4-11}$$

The sensitivity factor, s, is defined as

$$s_i = \frac{u_i^*}{\beta} \tag{4-12}$$

which is the direction cosine of the OP vector (from the origin to the minimum distance point) as shown in Figure 4-6. Thus

$$s_1^2 + s_2^2 + \ldots s_n^2 = 1 \tag{4-13}$$

which implies that each s_i^2 is a measure of the *contribution* to the probability (since the probability is related to β); the larger the s the larger is the contribution. In general, the

Example: V = -KI

The most-probable-point-locus (MPPL) is shown in Figure 4-5. It can be seen that the slopes for different V-values range from negative to positive.

Assuming that we are interested in the V = 1 limit state, the computational efficiency in determining the MPP depends on the optimization algorithm and the tolerances used in the analyses. The required number of V-function calculations for this example are in the order of 50 to 150 using different algorithms and tolerances.

Figure 4-5 suggests that a linear approximation for V at the origin is valid only in a small region close to the origin. By starting at the origin, several iteration methods failed to converge in locating the MPP for V = 1. However, by using a small enough incremental value of V, and keeping track of the MPPL, the iteration methods converged to the desired MPP.

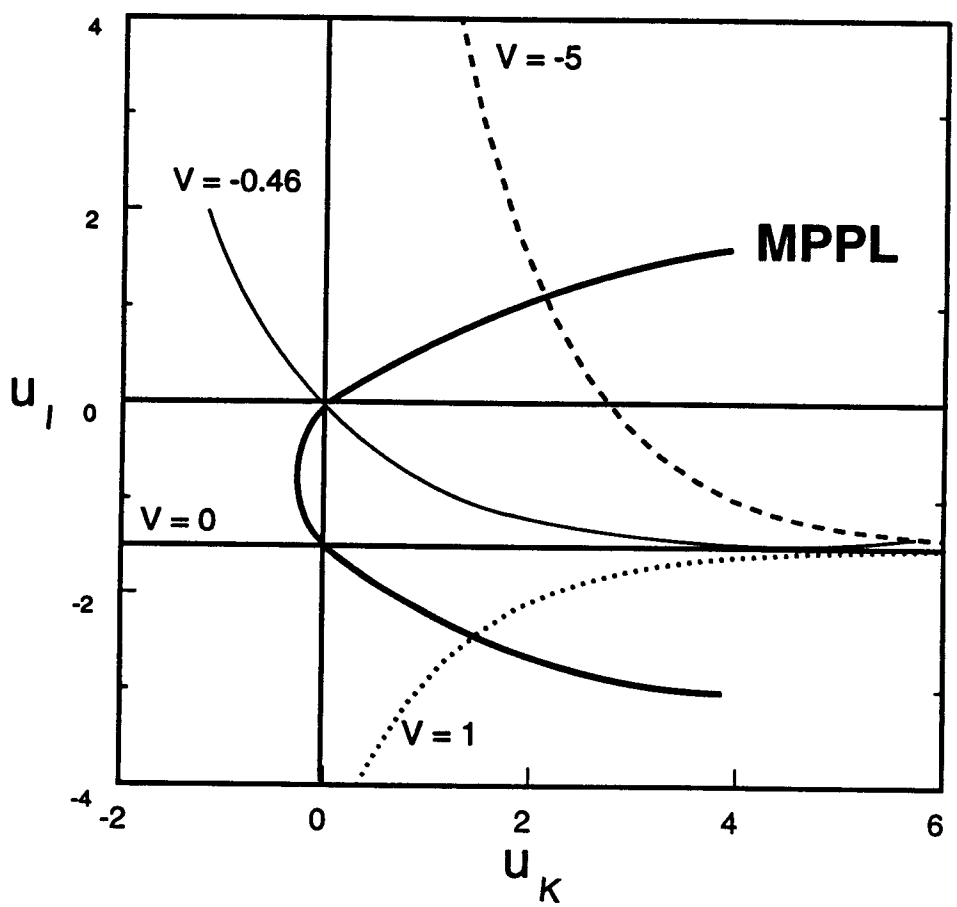

Figure 4-5. Limit state surfaces and most probable point locus

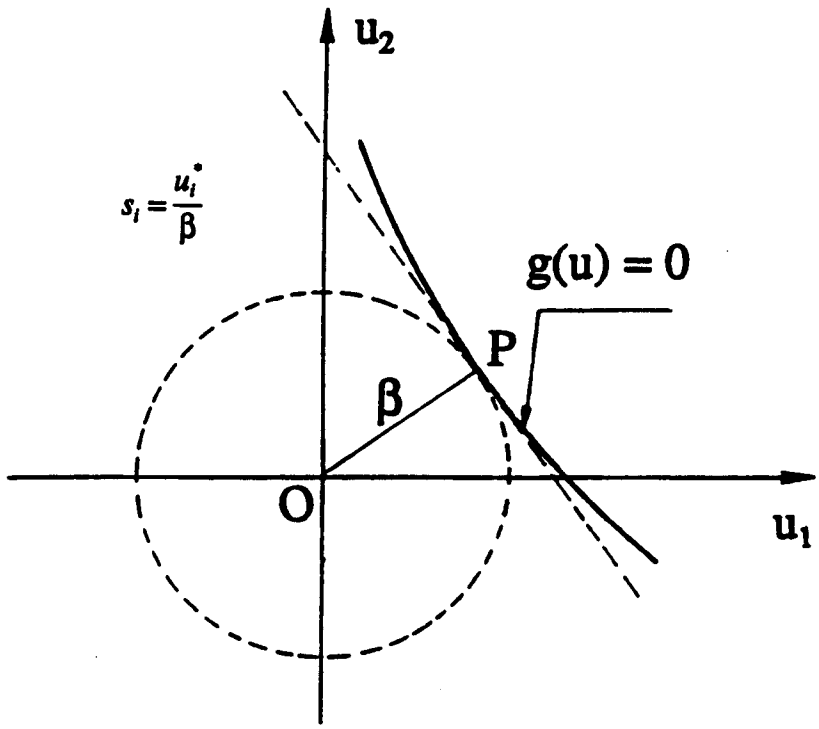

Figure 4-6. The definition of sensitivity factors

sensitivity factors depend on the g-function as well as the probability distribution. In a CDF analysis, the sensitivity factors will usually be different for different Z levels.

Other probabilistic sensitivities such as $d\beta/d\sigma_i$, and those described in Section 2 can be computed easily once a linear or a quadratic limit state surface is approximated.

4.7 CORRELATED RANDOM VARIABLES

For many applications, the information is seldom sufficient to establish the joint probability distributions. A more realistic model assumes the knowledge of marginal distributions for **X** and the associated correlation coefficients [Grogoriu (1983), Liu and Der Kiureghian (1986), and Wu et al. (1989)]. The transformation for this model first requires the application of Eqn. (4-5). The next step is to obtain the correlation coefficient matrix for the transformed normal variables **u**. Finally, given the correlation coefficient matrix, a linear transformation can be applied to find the uncorrelated normal random variables **v**. The basic assumption for this model is that random variables v_i are jointly normal.

Consider two random variables X_i and X_j with correlation coefficient R. The correlation coefficient of the transformed normal variables u_i and u_j denoted as r is needed in the analysis

Example: V = -KI

Based on Figure 4-3, the sensitivity factors can be computed using the MPP coordinate and Eqn. (4.9). The results are plotted in Figure 4-7 for three values of V: -5, 0, and 1. The figure shows that for V = -5, variable K is more important than I, for V = 0, I is the dominant variable, and for V = 1, I is more important than K.

Note that for V = 0, the solution is I = 0 or K = 0. But since K can take only positive values, the only possibility for V = 0 to happen is when I = 0. This explains why the sensitivity factor is zero for K.

This example illustrates that sensitivity factors depend strongly on the performance levels. In general, the sensitivity factors also depend on the probability distributions.

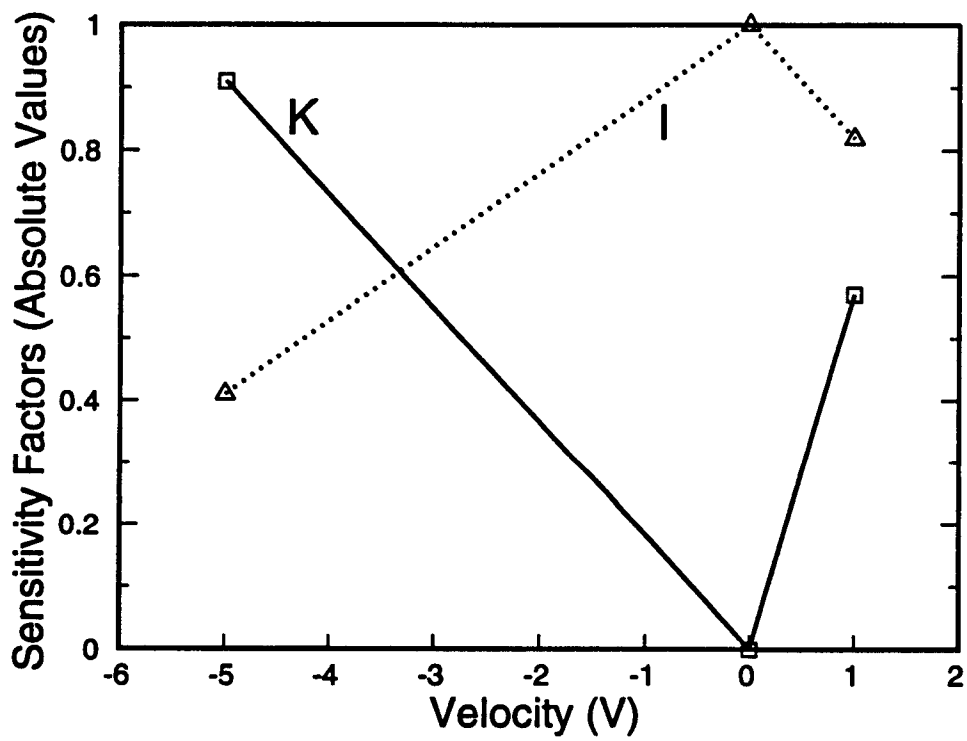

Figure 4-7. Sensitivity factors for velocity v_i = -5, 0, and 1

and can be found by solving the following equation:

$$R = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{x_i - \mu_i}{\sigma_i} \right) \left(\frac{x_j - \mu_j}{\sigma_j} \right) \phi_{ij} du_i \, du_j \quad , \tag{4-14}$$

where

$$\phi_{ij} = \frac{1}{2\pi\sqrt{1-r^2}} \exp\left(-\frac{u_i^2 - 2ru_iu_j + u_j^2}{2(1-r^2)} \right) \quad . \tag{4-15}$$

Because of the implicit nature of Eqn. (4-14), the calculation of r generally requires iterative solution.

4.8 MPP SEARCH AND CDF ANALYSIS BY THE ADVANCED MEAN VALUE (AMV) METHOD

The advanced mean value (AMV) [Wu et al. (1989), Wu and Burnside (1988), and Wu et al. (1990)] is designed to improve the estimate of CDF obtained from the mean value (MV) method. In the MV method, the performance function Z is approximated by the first-order terms of the Taylor's series as follows:

$$\begin{aligned} Z_{MV} &= Z(\boldsymbol{\mu}) + \sum_{i=1}^{n} \left(\frac{\partial Z}{\partial X_i} \right) \cdot (X_i - \mu_i) \; ; \; \mu_i = E\{X_i\} \\ &= a_o + \sum_{i=1}^{n} a_i X_i \end{aligned} \tag{4-16}$$

where $\boldsymbol{\mu}$ is the mean value vector and the derivatives are evaluated at the mean values of X_i. Selecting a limit state value z, the probability of $\{Z_{MV} < z\}$ can be computed using the methods described in Section 4.6. The AMV method improves this estimate as explained in the following.

In the AMV method, the approximation of Eqn. (4-16) is improved by writing:

$$Z_{AMV} = Z_{MV} + H(Z_{MV}) = \left(a_0 + \sum_{i=1}^{n} a_i X_i \right) + H(Z_{MV}) \tag{4-17}$$

where the $H(Z_{MV})$-function is defined as the difference between the values of Z_{MV} and Z_{EXACT} where it is assumed that an algorithm, possibly in the form of a computer program, is available. The computational steps of the AMV method are:

1. Select a limit state $Z_{MV} = z$.

2. Use Eqn. (4-16) to compute the CDF $P\{ Z_{MV} < z\}$.

3. In the transformed u-space compute the MPP **u***.

4. Transform **u*** to **x***.

5. Compute $Z_{EXACT}(\mathbf{x}^*)$ to update z for the CDF computed in Step 2.

Step 5 updates z at the MPP computed in Step 4, thus the MPP, **x***, defines a limit state of $Z_{EXACT} = Z(\mathbf{x}^*)$. This update requires one Z_{EXACT}-function calculation. By selecting different Z_{MV} = z-values and repeating the above steps, the CDF curve estimate can be constructed. An illustrative example for the above AMV analysis is given in Appendix A.

The accuracy of the AMV solution depends on the quality of the MPPL approximation. The accuracy will tend to be good if the MPPL developed from Z_{MV} is close to the exact MPPL and poor if the Z_{EXACT}-function is very nonlinear such that the MPPL of Z_{MV} is good only in a small region near the mean-values point. In this case, the AMV solution can only be applied effectively to that small region.

The AMV method applies the limit state concept and it can be improved by searching and using the exact MPPL, which requires more function calculations. In the context of MPP search (Section 4.5), the AMV method can be considered as providing a first guess of the MPP. Further improvement of MPP and CDF estimate may be made by successive linearization at the previous MPP and applying other MPP search procedures. These analyses may be combined with the "MPPL tracking" procedure proposed in Section 4.5 to minimize the possibility of convergence instability problems.

In summary, the AMV method bridges the differential analysis method and the limit state concept and provides a first-order determination of CDF, as well as other information regarding the behavior of the performance function. For example, it may be used to detect whether the function has more than one MPP (Wu et al., 1990). As a general procedure, the AMV solution should be updated at critical CDF points using improved MPP and adaptive sampling method described next. The application of AMV to the example problem is shown in Figures 4-8 and 4-9.

4.9 PROBABILITY UPDATING BY ADAPTIVE IMPORTANCE SAMPLING METHOD

The approximating function developed around the MPP provides only approximate probability without estimates of error bounds. It is useful to confirm the solution with minimal extra computations. One approach is to use an importance sampling method [Melchers (1989), Dolinski (1990), and Hohenbichler and Rackwitz (1988)] that samples the parameter space only in the region close to the MPP. This procedure requires fewer samples than are required by the standard Monte Carlo approach.

Example: V = -KI

By expanding the V-function at the mean value, we obtain:

$$V = 0.671 - 0.05K - 13.42I \tag{4-18}$$

From which a most probable point locus (MPPL) can be developed. Figure 4-8 shows the mean value-based MPPL as well as the exact MPPL based on the original equation. It appears that, because of the drastic change in the sensitivity factors for different V-values, the MV solution is a poor approximation, which results in a poor MPPL approximation. However, this can be detected easily by checking the sensitivity factors at the calculated MPP. Once this is detected, the MPPL-tracking procedure discussed above can be activated to search for the exact MPPL. This process was performed for V = 1, and the CDF results are shown in Figure 4-9.

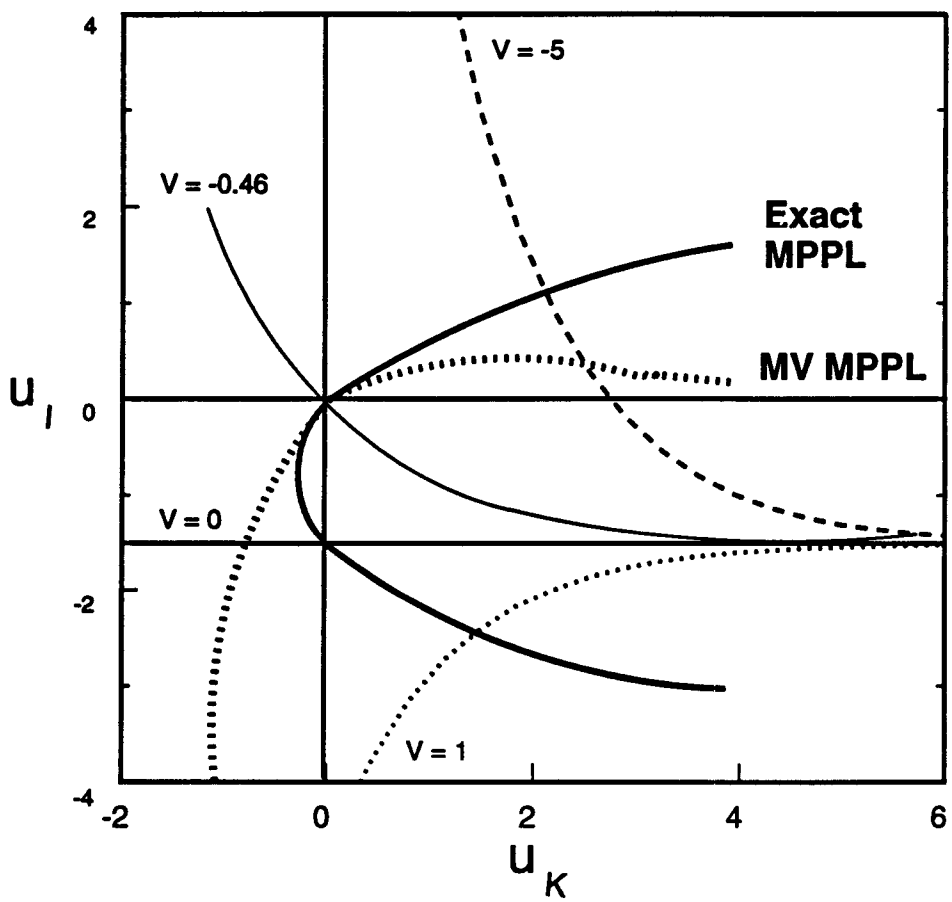

Figure 4-8. MPP search by the AMV method

Example: V = -KI

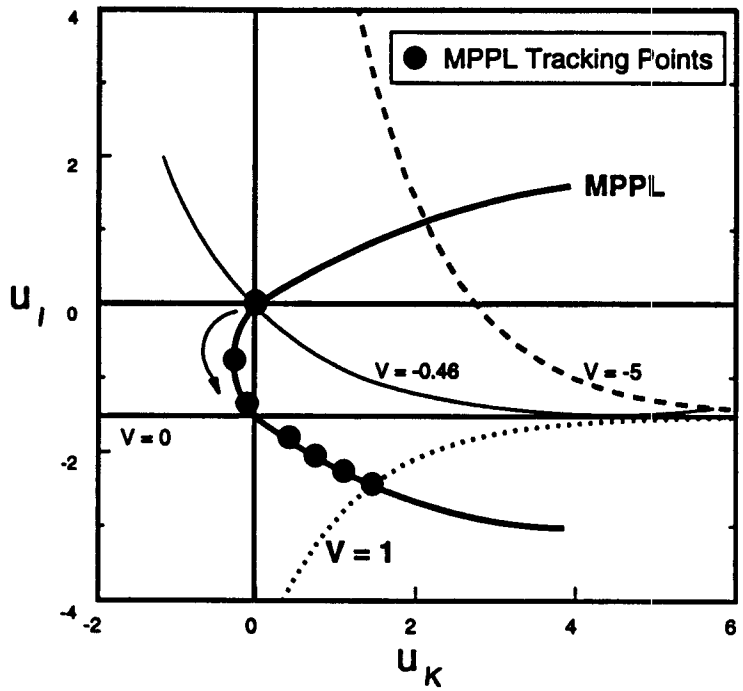

Figure 4-9(a). MPPL tracking points

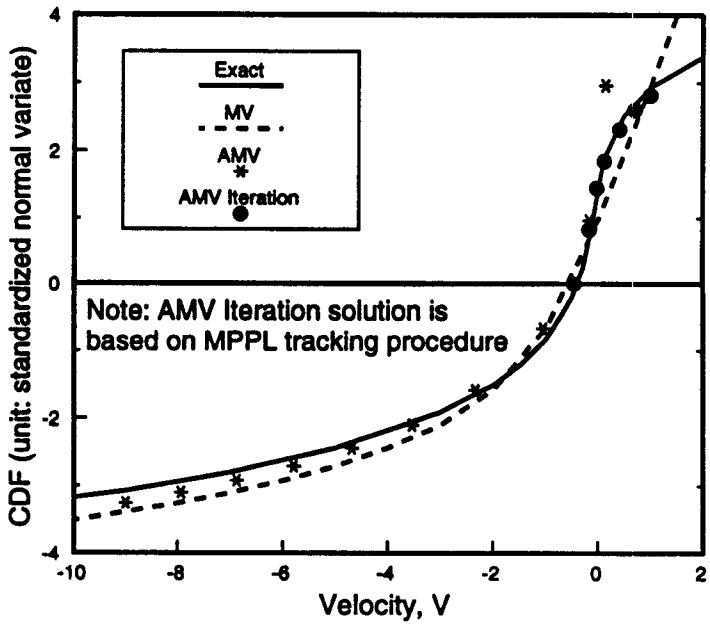

Figure 4-9(b). Calculated CDF

Assume that the interest is in calculating $P\{Z < z\}$. The region $\Omega = [Z \leq z]$ is called the target region. Denotes S as the (importance) sampling region. Let $P\{S\} = p_s$, the total number of samples in S be K, and the number of samples in the target region Ω be K_Ω. The probability of $[Z \leq z]$ is:

$$p \approx p_s \cdot \frac{K_\Omega}{K} \quad . \tag{4-19}$$

The sampling efficiency achieves its maximum if the selected importance sampling region equals the target region. In this extreme case, every point sampled will be in the target region, therefore $K_\Omega = K$ and $p = p_s$. The other extreme case corresponds to a standard Monte Carlo approach where $p_s = 1$, and

$$p \approx \frac{K_\Omega}{K} \quad . \tag{4-20}$$

An adaptive importance sampling method (Wu et al. 1991) lies between these two extremes. Since the region of $[Z \leq z]$ is not known *a priori*, one starts with a guess and subsequently adjusts the region based on the observed results. For example, one can start with a sampling region that is based on the first or second-order limit state surface developed at the MPP. The sampling region is adjusted by deforming the limit state surface. The deformation can be designed to gradually increase the sampling region until it fully covers the target region. An example based on changing the curvature of the surface is illustrated in Figure 4-10. When the sampling region fully covers the target region, the probability solution will converge, indicating that no more deformation is required. Since the efficiency is dominated by the ratio of p/p_s, the procedure is essentially independent of the number of random variables.

The above procedure is intended for problems with $p < 0.5$. For optimal efficiency, when the CDF of interest is greater than 0.5 the sampling should be performed in the $[Z > z]$ region and then use the result to compute the CDF. Application of importance sampling is shown in Figure 4-11.

4.10 SUMMARY OF THE LIMIT STATE APPROACH

In summary, the limit state approach consists of the following steps:

1. Select a set of discrete values of performance measure, z_i.

2. For each z_i search for the most probable point (MPP) (Section 4.5). Transformation from the correlated non-normal random variables to uncorrelated normal random variables is required (Sections 4.3 and 4.7).

3. Approximate the performance function at the MPP by a linear or quadratic function (Section 4.4).

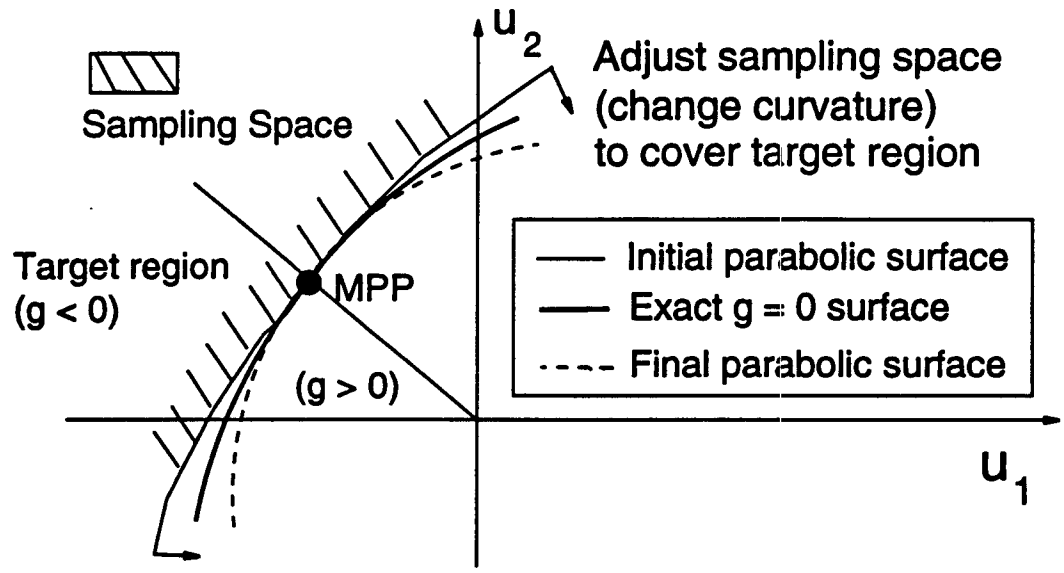

Figure 4-10. The concept of the curvature-based adaptive importance sampling method

4. Calculate the CDF at z_i based on the approximate function (Section 4.4).

5. At critical tail regions, confirm the probability solution by the importance sampling method (Section 4.9).

6. Calculate the probabilistic sensitivities based on the approximate function (Section 4.6).

7. Plot the CDF of Z.

Some further details of these steps are included in the next section and Appendix A.

Example: V = -KI

The adaptive sampling method described above was applied to $V \geq 1$. The probability in the sampling region is 0.00239, the total number of sampling points is 61, of which 40 is in the target region. The probability of $V \geq 1$ is:

$$p_s \cdot K_0/K = 0.00239 \cdot 40/61 = 0.00157. \quad \textbf{(4-21)}$$

The V = 1 limit state and the sampling points are shown in Figure 4-11.

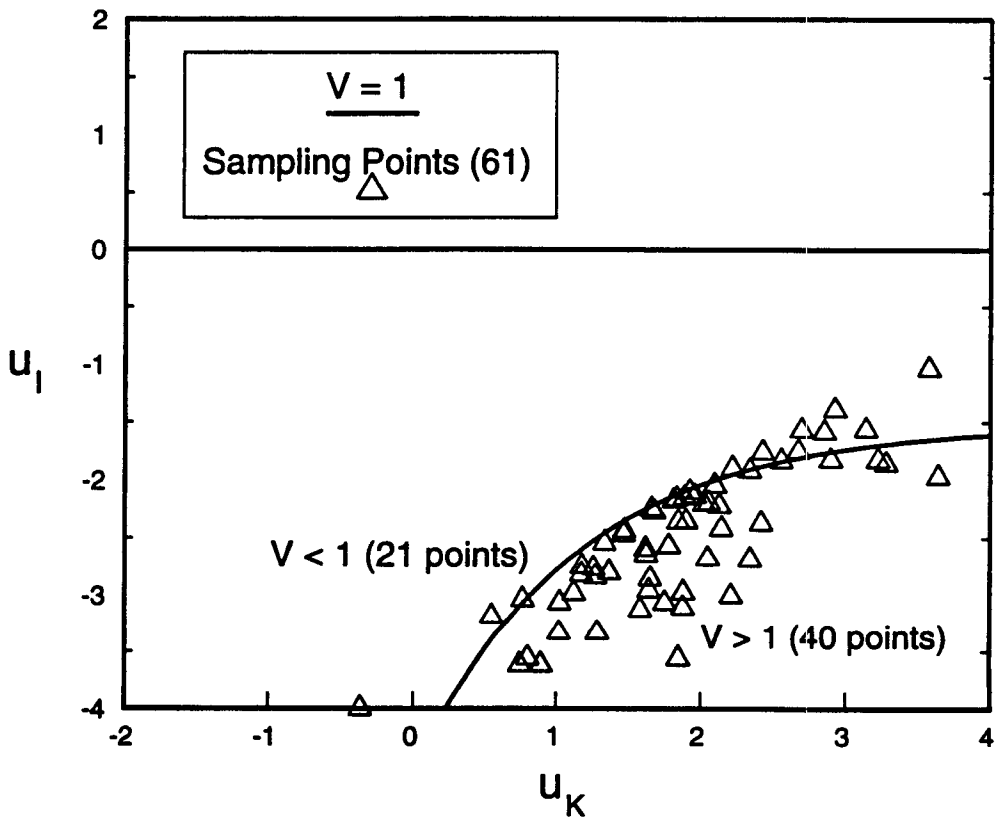

Figure 4-11. Calculation of P { V ≥ 1 } by the Adaptive Importance Sampling Method

5 APPLICATION EXAMPLE

5.1 DESCRIPTION OF THE RADIONUCLIDE TRANSPORT MODEL

This example presents an uncertainty and sensitivity study of a hypothetical High Level Waste (HLW) underground repository intersected by a vertical fracture or fault and under saturated conditions, as illustrated in Figure 5-1. A one-dimensional analytical model based on the Laplace transform method yielding the concentration in the fracture and rock layers and the cumulative mass release of radionuclides at any point within the fracture was developed in Part 1 (Gureghian et al. 1991). In this instance, the cumulative mass is adopted as the performance measure. The deterministic sensitivity of the performance measure to some of the system parameters affecting the transport process was derived analytically in Part 1 and the same is used in the uncertainty and sensitivity analysis in the following.

The demonstration example reported here refers to the one-dimensional transport of a radionuclide, Np-237, in a heterogeneous saturated fractured rock system composed of five layers (the last extending to infinity, see Figure 5-2), with piece-wise constant parameters. The steady flow rate of water per unit width of fracture is 0.1 m^2/yr. The type of solute release mode at the source investigated here corresponds to a band release, where the leaching time T = 5×10^3 years. Note that the flow domains in both fracture and rock layers are assigned non-zero initial concentrations. The input data pertaining to this test case is presented in Table 5-1.

Figure 5-3 depicts the time-dependent evolution of the cumulative mass (per unit width of the fracture) profile at three different observation points in the fracture. These are located at distances of 100, 200 and 500 meters downstream from the source, lying within the second, third and fifth layer respectively. A comparison of our analytical solution results with those yielded by Stefhest's solution indicates excellent agreement (Gureghian et al. 1991). Note that all three profiles tend to become asymptotic to three specific values of the cumulative mass namely, 4.902×10^2, 4.7×10^2 and 4.309×10^2 (Units of Activity/m) at times greater than 10^4 years.

Figures 5-4a and 5-4b shows a comparison between the analytical and numerical spatial and temporal variations of sensitivity of the cumulative mass to two typical rock matrix parameters: pore diffusivity and distribution coefficient respectively (See Part 1 report for details). These results suggest that for the given range of parameters, pore diffusivity seems to have an overall greater impact on the cumulative mass than the distribution coefficient. In both cases, maximum sensitivity seems to be achieved at time close to 2×10^4 years.

5.2 APPLICATION OF THE LIMIT STATE APPROACH

The above example was used to test the performance of the probabilistic methods discussed in previous Sections. The performance measure is taken as the cumulative mass release at an observation point in the fracture located at a distance 500m downstream from the

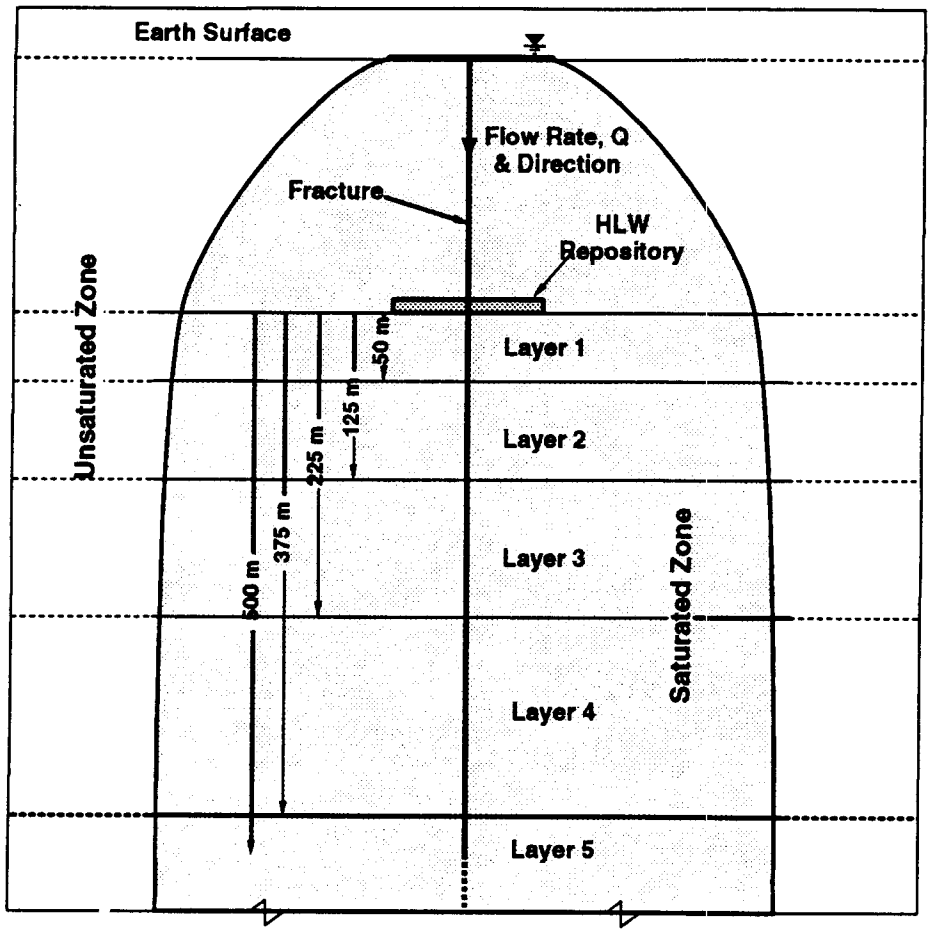

Figure 5-1. Vertical cross-section of a layered rock system intersected by a vertical fracture subject to surface ponding. Calculations apply to the saturated zone only.

source (i.e., in the fifth rock layer) and for a simulation time corresponding to 10^4 years. The mean values of the 25 random variables (i.e., b, ϕ, D_p, K_f, and K_r in each layer) are listed in Table 5-1. All the random variables are assumed to be independent lognormally distributed with coefficients of variation of 0.5. As pointed out in Section 4, the methodology can apply to correlated and non-normal distributions.

In Figure 5-5, several results using various approaches are plotted. The probability (CCDF) is shown on a semi-logarithmic scale to emphasize the low probability region. Initially, a Monte Carlo-based reference solution was developed using 5000 simulations. Using the procedure described in Section 4.8, the MV model using analytical sensitivity calculations was

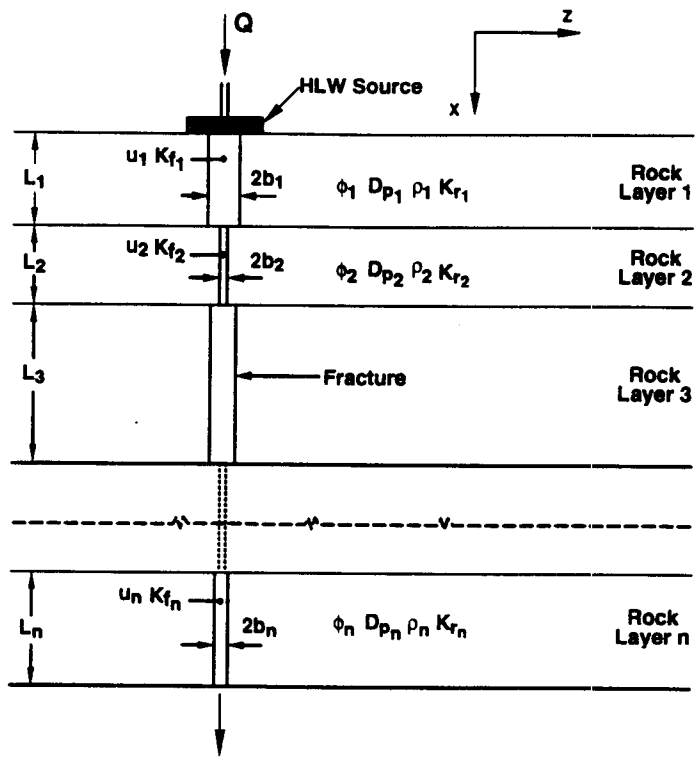

Figure 5-2. Description of migration pathways in a system of homogeneous layers of fractured rock. In the worked example, number of layers, n = 5, and migration path lies entirely in the saturated zone.

first developed (the MV solution is not shown on the plot), then the AMV solutions were obtained for seven selected probability values (shown as discrete points in Figure 5-5). Had the MV sensitivities been calculated through numerical differentiation, a minimum of 33 (26 for the first-order coefficients and 7 for the selected CCDF levels) function evaluations would have been required. The AMV solution produced a CCDF whose shape resembles the one obtained through the Monte Carlo method. To demonstrate how to improve the AMV solution, a limit state of Z = 223 units of activity/m was selected. Note that this value of Z has a low probability of exceedance or, in other words, it lies in the tail region of the CCDF. Subsequently, the converged MPP was obtained using three additional iterations by successive linearization (see Section 4.5). Finally, a quadratic function without product terms was developed to approximate the exact limit state function. If the sensitivities were calculated by numerical differentiation, a minimum of 129 (26 x 4 = 104 calculations for finding the MPP and another 25 for calculating the second-order coefficients) function evaluations would have been required for this approximation. The above numbers indicate that the number of function evaluations and the computational time depend on the number of random variables and the way the sensitivities are computed. It can be seen that the converged AMV solution approaches the Monte Carlo results.

TABLE 5-1. INPUT PARAMETERS FOR THE RADIONUCLIDE TRANSPORT PROBLEMS. SPECIES: Np-237, $T_{1/2}$ = 2.3 x 10^6yr, LEACHING TIME T = 5 x 10^3yr, A^0 = 1.0*, Q = 0.1(m^2/yr).

Layer	L(m)	b(m)	u(m/yr)	ϕ
1	50.0	5.0E-03	10.0	0.01
2	75.0	4.0E-03	12.5	0.008
3	100.0	3.0E-03	16.666	0.006
4	150.0	2.0E-03	25.0	0.004
5	∞	1.5E-03	33.333	0.002

Layer	ρ(g/cm^3)	D_p(m^2/yr)	K_f(m)	K_r(cm^3/g)
1	2.0	0.01	5.0E-03	0.5
2	2.3	0.02	8.0E-03	0.6978
3	2.6	0.06	2.7E-02	1.158
4	2.65	0.05	1.0E-02	1.059
5	2.7	0.03	3.0E-03	0.741

Layer	a_1*	a_2*	α(m^{-1})	b_1*
1	1.50E-04	-0.50E-04	0.02	1.00E-05
2	2.00E-04	-0.25E-05	0.02	1.75E-05
3	1.75E-04	-0.20E-05	0.02	1.25E-05
4	2.00E-04	-0.15E-05	0.02	1.05E-05
5	1.50E-04	-0.20E-05	0.02	1.05E-05

* (UA/L^3 = arbitrary units of activity/L^3)

Notes:

L	=	rock layer thickness
b	=	half thickness of fracture
u	=	average fluid velocity in the fracture
ϕ	=	rock porosity
ρ	=	bulk rock density
D_p	=	pore diffusivity
K_f	=	surface distribution coefficient in fracture
K_r	=	distribution coefficients in rock matrix
a_1	=	coefficient in the initial fracture concentration equation
a_2	=	coefficient in the initial fracture concentration equation
α	=	coefficient in the initial fracture concentration equation
b_1	=	initial concentration in the rock matrix

See Gureghian et al., (1991) for details.

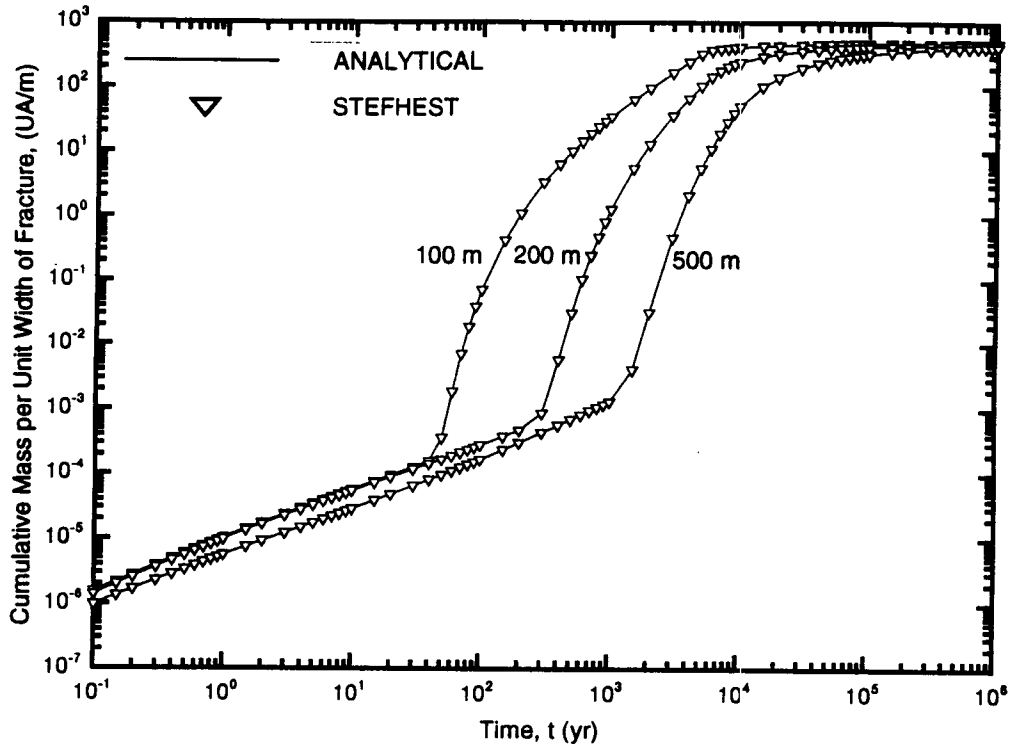

Figure 5-3. Cumulative mass of Np-237 per unit in the fracture versus time at different positions x = 100, 200, and 500 meters (exponentially decaying source and band release mode)

Based on the exact MPP, the adaptive sampling scheme is applied to the limit state, Z = 223 Units of Activity/m. For simplicity, the adaptive limit state surface has a common curvature at the MPP. Consequently, 532 simulations were needed (after the MPP is found) to achieve the "exact" (Monte Carlo) solution. The required nurnber of simulations could be significantly reduced if separate curvatures are used so that the final sampling region is closer to, but still covers, the target region. Since the efficiency is dominated by the ratio of the probability in the target region to the probability in the sampling region (see section 4.9), the required number of samples is essentially independent of the probability level or the number of random variables. Thus, the method offers significant advantage for calculating small CDF or small CCDF.

Because the final MPP-based approximate CCDF is reasonably good, as confirmed by the adaptive sampling method, it suggests that the MPP-based polynomial function can be used to replace the exact function for further analyses such as probabilistic sensitivity analyses.

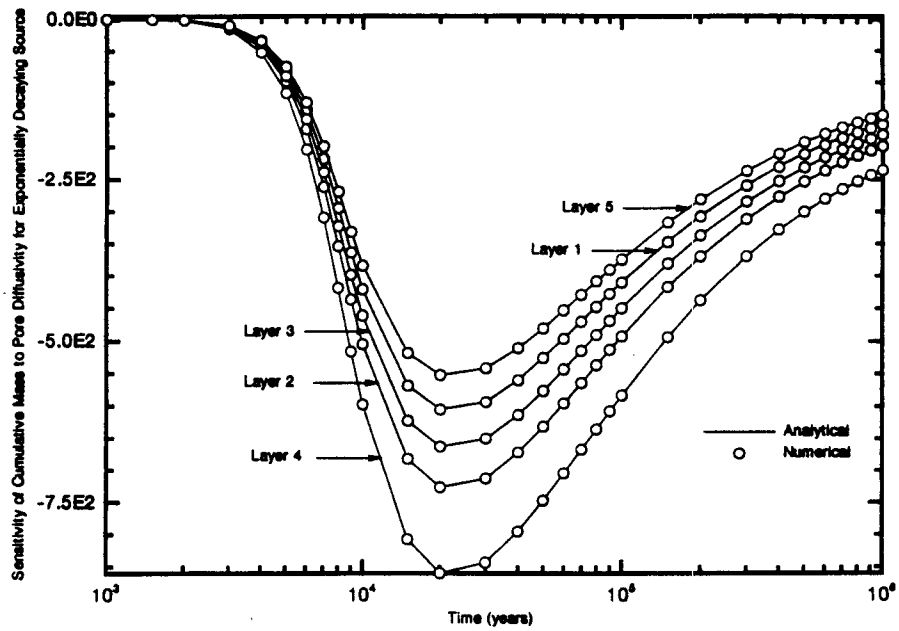

Figure 5-4(a). Sensitivity of cumulative mass to pore diffusivity versus time for NP-237 (exponentially decaying source). Curves marked layer 1 to 4 pertain to cumulative release at layer interfaces, and the curve marked layer 5 pertains to cumulative release at 500 m.

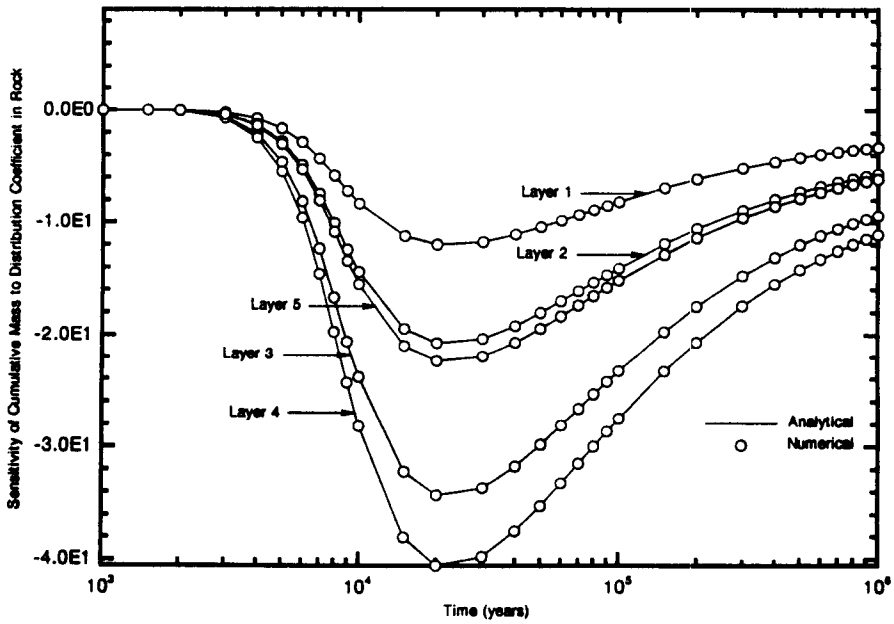

Figure 5-4(b). Sensitivity of cumulative mass to distribution coefficient in rock versus time for Np-237 (exponentially decaying source). Curves marked layer 1 to 4 pertain to cumulative release at layer interfaces, and the curve marked layer 5 pertains to cumulative release at 500 m.

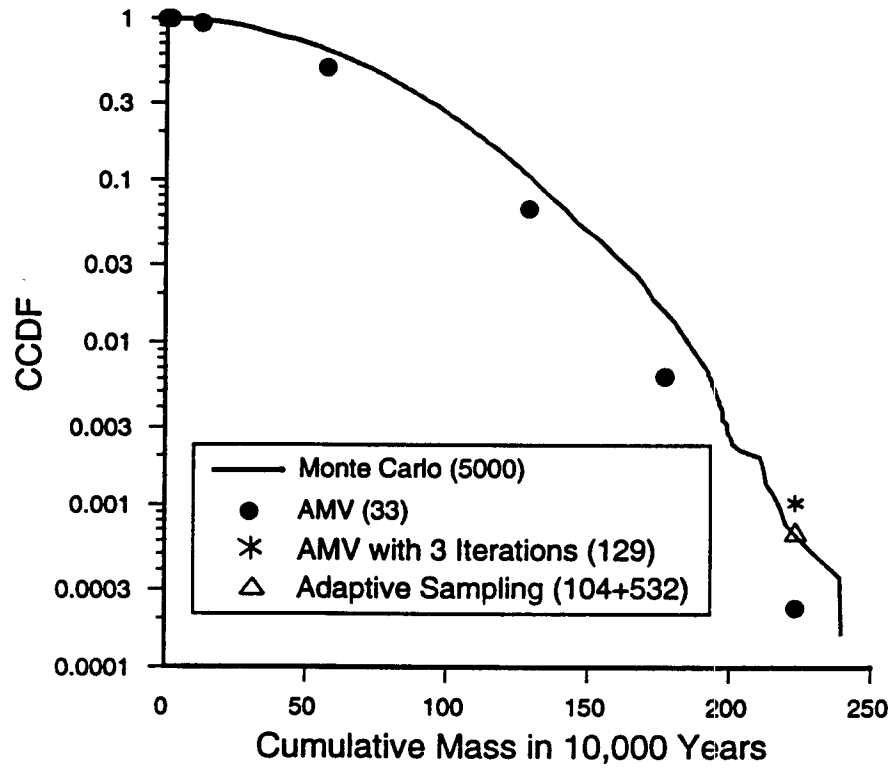

Figure 5-5. Comparison of Limit State Approach with Monte Carlo

5.3 COMPARISON OF LIMIT STATE APPROACH WITH OTHER APPROACHES

Two computer programs LHS (Iman and Shortencarier, 1984) and STEPWISE (Iman et al., 1980) were used to compare the LHS, the regression analysis, and the limit state method. The fractile limits of 0.001 and 0.999 were built into the LHS program.

In Figures 5-6 and 5-7, the CCDF results are plotted on a semi-logarithmic paper. In Figure 5-6, the LHS results for 50, 100 and 300 samples are shown to illustrate the improvement in the CCDF for increasing numbers of samples. Figure 5-7 compares the 50-sample LHS run, the 5000-sample Monte Carlo run and the AMV analysis results. This figure shows that the AMV method works well for the present example and was able to generate with only 33 samples a CCDF comparable to 5000 Monte Carlo runs, or more than 300 LHS runs.

The LHS results were also used to determine sensitivities; i.e., the derivatives of the cumulative release with respect to the input variables. The sensitivities of the output result (cumulative release for 10,000 years) were determined by the stepwise linear regression program STEPWISE. In this method, the input and output for all samples are fitted to a linear equation:

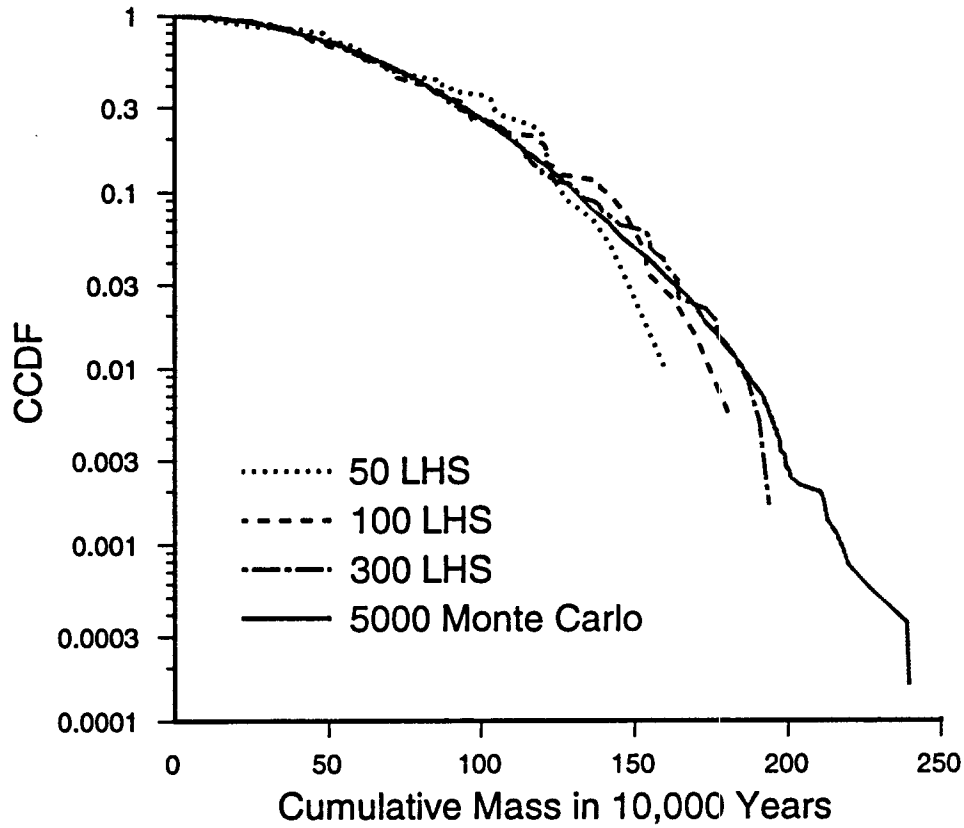

Figure 5-6. CCDF results using Latin Hypercube sampling

$$Z = a_0 + \sum_{i=1}^{N} a_i X_i \ , \qquad (5\text{-}1)$$

where Z is the cumulative release for 10,000 years, a_i is the proportionality coefficient for input variable X_i, and a_0 is the constant term. The choice of the functional form for Eqn. (5-1) could have been different than linear; e.g., it could include transformations of the input variables or combinations of several input variables to form new input variables. For the present example, only the linear form of the regression equation in terms of the input variables was chosen. The linear format of Eqn. (5-1) allows approximate comparison between the stepwise linear regression and differential analysis, since both results reflect sensitivities near the mean-values

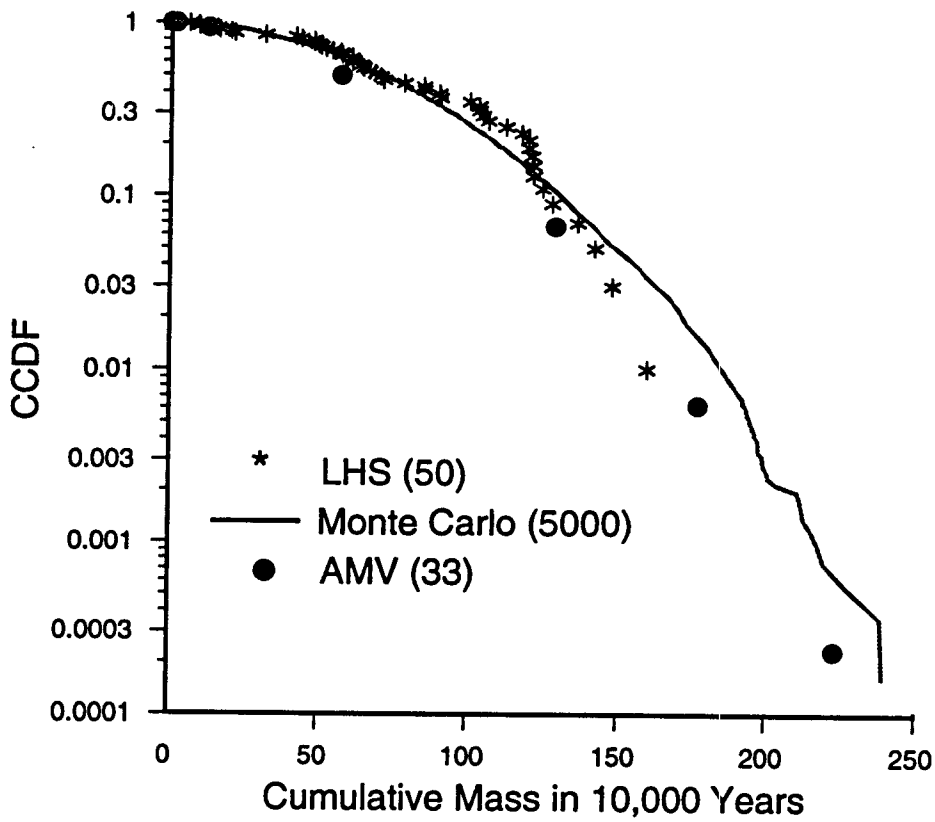

Figure 5-7. Comparisons of the AMV method with the simulation methods

point of the space of input variables; i.e., the a_i terms are like sensitivities dZ/dX_i.

Sensitivities generated by the differential analysis were compared to those generated from the regression analysis. Table 5-2 compares the sensitivities generated by the differential analysis and the regression coefficients resulting from 50 LHS samples. The first column denotes the name of the 25 input variables for the five layers. The next five columns in Table 5-2 give the regression coefficients in ascending order of the regression equation; i.e., N = 5 gives the coefficients for the regression equation which contains the five most significant input variables. The column for N = 25 contains the coefficients of the regression equation where

TABLE 5-2. COMPARISION OF SENSITIVITIES FROM REGRESSION TO DIFFERENTIAL ANALYSIS (50 LHS SAMPLES)

N	5	7	11	13	25	Diff. Anal.
a_0	216.3	258.0	320.5	335.9	348.2	
b_1					821.1	-9.7
b_2					-205.0	-15.4
b_3					-409.4	-21.1
b_4					1504	-32.4
b_5					-3094	-26.8
ϕ_1			-1197	-1334	-1390	-415.7
ϕ_2					-652.3	-1247
ϕ_3		-3763	-3996	-4023	-3856	-4550
ϕ_4	-6871	-7127	-7835	-7516	-7381	-7384
ϕ_5			-7848	-7407	-8352	-5704
D_{p1}					-31.8	-415.7
D_{p2}			-808.4	-805.0	-850.0	-499.2
D_{p3}	-442.9	-452.2	-414.8	-413.3	-421.1	-455.4
D_{p4}	-518.1	-494.4	-471.4	-454.9	-449.8	-590.8
D_{p5}		-637.6	-641.8	-585.1	-552.9	-379.2
K_{f1}					-1240	-9.7
K_{f2}					268.1	-15.4
K_{f3}					70.0	-21.1
K_{f4}					-15.9	-32.4
K_{f5}					-1516	-26.7
K_{r1}				-23.9	-23.4	-6.6
K_{r2}			-19.6	-18.6	-17.7	-12.6
K_{r3}	-24.3	-28.4	-28.0	-27.5	-28.0	-22.0
K_{r4}	-29.36	-27.9	-29.0	-28.7	-28.4	-26.3
K_{r5}				-11.3	-10.9	-13.8
r^2	0.659	0.770	0.890	0.920	0.940	---

all 25 input variables were included. The last column gives the sensitivity coefficients for the differential analysis. The bottom row in Table 5-2 gives the correlation coefficient r^2 value for each of the regression equations. Table 5-3 presents the same comparison, but for the 300 LHS samples.

Tables 5-2 and 5-3 demonstrate good qualitative agreement on the magnitude and sign of the sensitivities for the regression and differential analyses for the five most important parameters. On the less important parameters, there is significant disagreement.

By inspection of the governing equations, all sensitivity coefficients should be negative; that is, an increase in any of the 25 input variables should lead to a decrease in the cumulative release. Regression of the LHS results which forced significance on all 25 input variables illustrates that in some cases the magnitude and sign of the sensitivity coefficients disagreed markedly, which is a clear sign of "overfitting" the data to the regression equation. Program STEPWISE performs regression with stepwise addition of important variables, and stops addition of variables when r^2 reaches a predetermined value, typically 0.95. This practice limits overfitting the regression equation to the data. In the case of 50 samples and r^2 cutoff of 0.95, STEPWISE limited the order of the equation to the 13 most significant variables. In the case of 300 samples, equations up to order 16 were admitted. Note however that the 16th order equation for 300 samples case also had positive signs for some of the sensitivity coefficients.

The differential analysis or the regression equations from the LHS analysis can be considered as the approximate model to the true cumulative release model. Based on the generated LHS samples, approximate cumulative release results, Z', can be computed and compared with the exact results, Z. If Z' is close to Z, then the "approximate" model is good. In general, the degree of linear relationship, determined by performing a statistical correlation analysis is an indication of how good the approximate model is.

Figures 5-8 and 5-9 are the Z versus Z' plots. Figure 5-8 is based on the full 25 variable polynomials developed for the 50 and 300 sample cases using the regression results. The agreement is quantitatively good except for a few "outliers" for both regression equations. The standard error, S, was chosen to evaluate the approximate model. It is defined as:

$$S = \frac{1}{K} \sqrt{\sum_{i=1}^{K} (Z_i' - Z_i)^2} , \qquad (5\text{-}2)$$

where K = number of samples. The standard errors for the outputs was 1.083 and 0.827 for the 50 and 300-sample regression equations respectively.

Figure 5-9 shows Z versus Z', where Z' is based on the differential analysis. The standard error S was 1.667. Thus the linear approximate model using the differential analysis is a relatively poor approximate model. Not surprisingly, agreement between Z and Z' is closest at the cumulative release value predicted by the mean values of the input variables, since this

TABLE 5-3. COMPARISION OF SENSITIVITIES FROM REGRESSION TO DIFFERENTIAL ANALYSIS (300 LHS SAMPLES)

N	5	7	11	13	25	Diff. Anal.
a_0	215.4	252.7	305.3	326.0	328.7	
b_1					183.6	-9.7
b_2					-728.1	-15.4
b_3					226.3	-21.1
b_4					-751.5	-32.4
b_5					441.8	-26.8
ϕ_1				-517.5	-518.3	-415.7
ϕ_2			-1507	-1521	-1477	-1247
ϕ_3	-4313	-4273	-4286	-4323	-4347	-4550
ϕ_4	-7360	-7253	-7454	-7399	-7431	-7384
ϕ_5		-6944	-7044	-7261	-7251	-5704
D_{p1}				-409.2	-404.8	-415.7
D_{p2}			-548.6	-542.3	-537.6	-499.2
D_{p3}	-454.4	-458.2	-459.4	-458.5	-460.1	-455.4
D_{p4}	-596.3	-589.1	-599.4	-504.5	-605.3	-590.8
D_{p5}			-479.4	-505.1	-506.0	-379.2
K_{f1}				1060	1043	-9.7
K_{f2}					193.9	-15.4
K_{f3}					6.96	-21.1
K_{f4}					92.7	-32.4
K_{f5}					-4089	-26.7
K_{r1}				-9.04	-9.13	-6.6
K_{r2}				-15.95	-16.01	-12.6
K_{r3}		-21.6	-21.6	-21.9	-22.0	-22.0
K_{r4}	-25.7	-25.2	-25.4	-25.7	-25.8	-26.3
K_{r5}			-17.8	-16.6	-16.7	-13.8
r^2	0.617	0.739	0.839	0.871	0.940	---

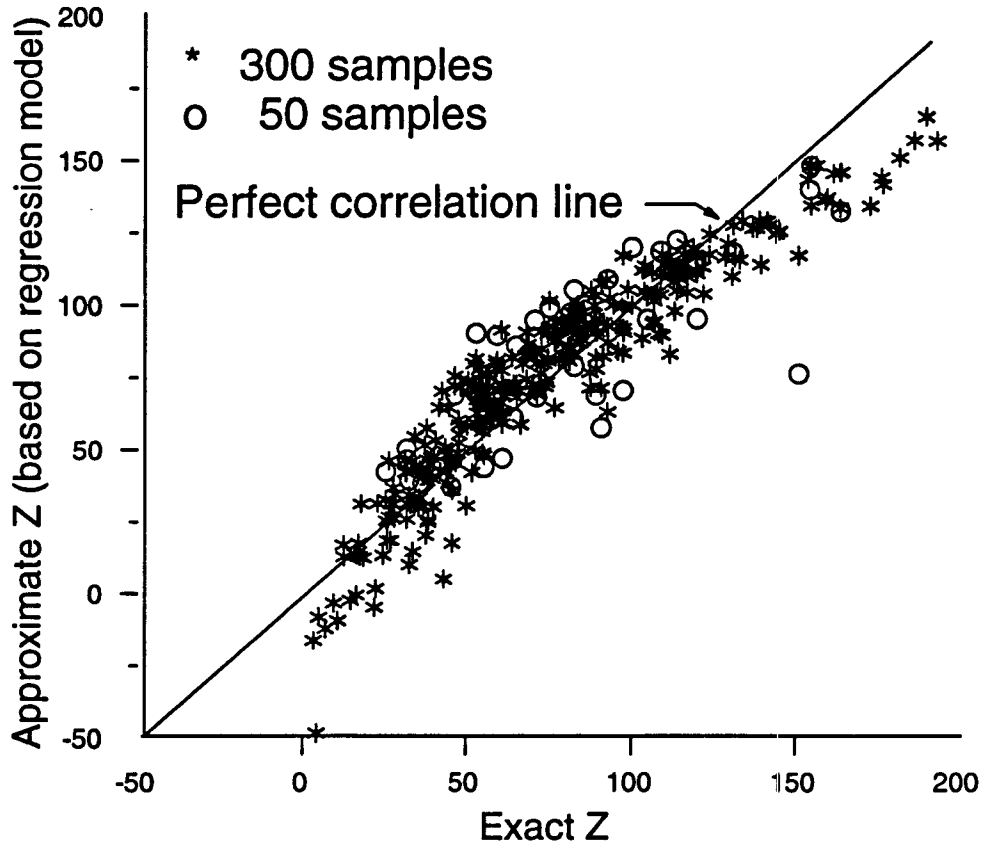

Figure 5-8. Comparisons of the regression model with the exact model

is the reference point in parameter space at which the sensitivities were calculated.

Cumulative release based on the regression coefficients, while deviating at the extremes of the cumulative release range, predicts the results better than those based on the differential analysis. This result is understandable, since the differential analysis model relies only on the behavior of the model at one point in parameter space, while the regression model is optimized to perform over a wide range of input parameters.

The sensitivities, dZ/dX_i, express the change in the cumulative release with respect to changes in the input variables. Since the dimensional units are different, the sensitivities cannot be used directly to determine the relative importance of the input variables.

In order to rank the importance of the input variables, the sensitivities must be normalized to make them dimensionless. As discussed in Section 2, the differential analysis

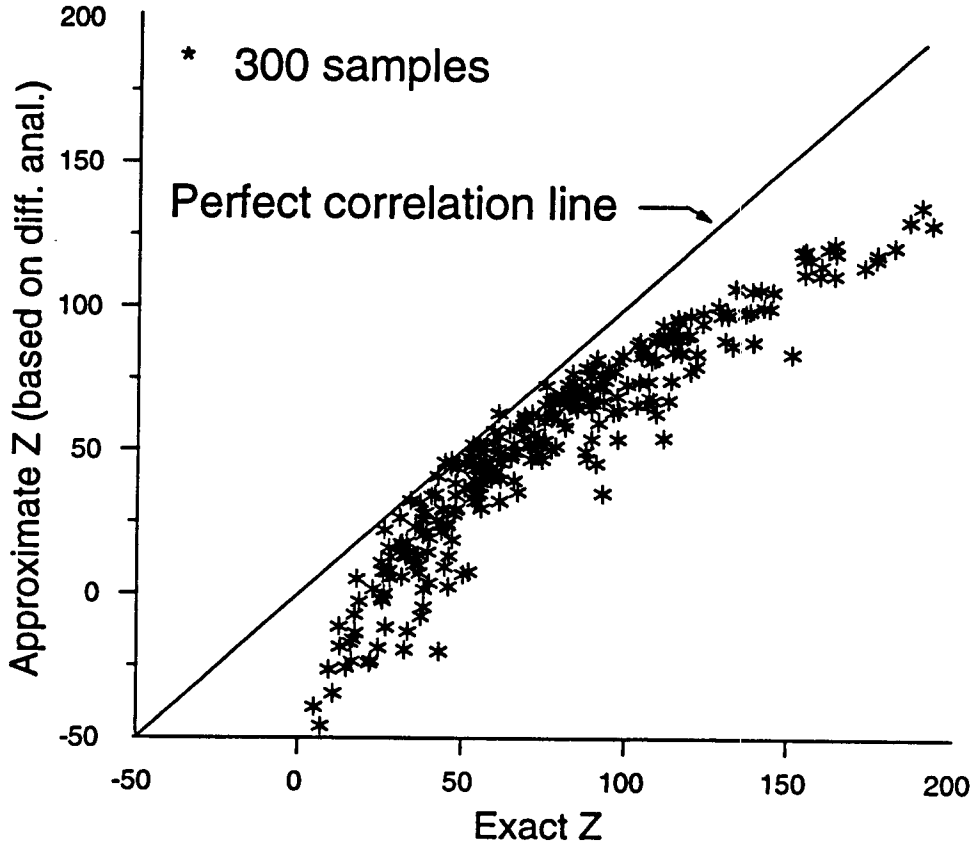

Figure 5-9. Comparisons of the differential analysis model with the exact model

normalizes the first derivatives with respect to the mean values of the input and dependent variables. The stepwise regression method normalizes the regression coefficients with respect to the standard deviation of the input and dependent variables. Although these two normalizations are different, they may be used to rank the "overall" relative importance of the input variables. The normalized sensitivity coefficients are compared in Table 5-4 for the differential analysis and the 5th and 7th order polynomial regressions on 50 and 300 samples.

The normalized differential sensitivity coefficients compared reasonably well to the regression analyses. The absolute values of the normalized sensitivities for the differential analysis were similar in magnitude. In nearly all cases, the normalized sensitivities from the regression analyses were also similar in magnitude for the same set of input variables which were most important for the differential analysis. There were different rankings within each group, but absolute differences between the normalized sensitivity coefficients within each group were small.

Table 5-4. SENSITIVITY COEFFICIENTS FROM REGRESSION AND DIFFERENTIAL ANALYSIS (RELATIVE RANK IN SQUARE BRACKETS)

N	5	7	5	7	Diff. Anal.
a_0					
b_1					
b_2					
b_3					
b_4					
b_5					
Φ_1					
Φ_2					
Φ_3		-.268[6]	-.320[5]	-.317[5]	
Φ_4	-.316[3]	-.328[3]	-.364[2]	-.359[1]	-1.805[3]
Φ_5				-.172[7]	
D_{p1}					
D_{p2}					
D_{p3}	-.309[4]	-.316[4]	-.335[4]	-.337[3]	-1.675[4]
D_{p4}	-.299[5]	-.285[5]	-.364[1]	-.357[1]	-1.810[1]
D_{p5}		-.221[7]			
K_{f1}					
K_{f2}					
K_{f3}					
K_{f4}					
K_{f5}					
K_{r1}					
K_{r2}					
K_{r3}	-.365[2]	-.382[2]		-.305[6]	-1.672[5]
K_{r4}	-.443[1]	-.421[1]	-.337[3]	-.335[4]	-1.807[2]
K_{r5}					

The probabilistic sensitivity factors (see Section 4.6) based on the limit state approach are shown in Table 5-5 for Z = 51 and Z = 223. The results show how sensitivity factors change for different Z values. At the mean value point, where the corresponding Z is 51 UA/m, the six most significant random variables are ϕ_4, D_{p4}, K_{r4}, ϕ_3, D_{p3}, K_{r3}, and the importance of these six variables is about equal. At Z = 223 UA/m, the above six variables remain the most dominant ones indicating no significant change in the transport process. In this instance, the sensitivity factors for ϕ_4, D_{p4}, K_{r4}, are higher than ϕ_3, D_{p3}, K_{r3}.

It can be concluded that the differential analysis is limited to a "local" (mean-based) analysis, the regression analysis is more appropriate for a "global" analysis, and the limit state approach provides both global and local analysis capability.

TABLE 5-5. PROBABILISTIC SENSITIVITY FACTORS (RELATIVE RANK IN SQUARE BRACKETS)

	Z = 51	Z = 223
b_1	0.0	0.0
b_2	0.0	0.0
b_3	0.0	0.0
b_4	0.0	0.0
b_5	0.0	0.0
ϕ_1	0.0	0.046
ϕ_2	0.18	0.116
ϕ_3	0.35[4]	0.366[4]
ϕ_4	0.36[1]	0.405[1]
ϕ_5	0.198	0.134
D_{p1}	0.09	0.046
D_{p2}	0.18	0.116
D_{p3}	0.34[4]	0.368[4]
D_{p4}	0.36[1]	0.405[1]
D_{p5}	0.198	0.134
K_{f1}	0.0	0.0
K_{f2}	0.0	0.0
K_{f3}	0.0	0.0
K_{f4}	0.0	0.0
K_{f5}	0.0	0.0
K_{r1}	0.09	0.046
K_{r2}	0.18	0.116
K_{r3}	0.34[4]	0.367[4]
K_{r4}	0.36[1]	0.405[1]
K_{r5}	0.198	0.134

6 SUMMARY AND DISCUSSIONS

This report investigates several uncertainty and sensitivity analysis methods for problems related to HLW repository performance assessment. The alternative approach proposed in this report can be viewed as a combination of differential analysis, response surface, and importance sampling. The limit state and the most probable point concepts provide the required linkage to switch from the differential analysis to the response surface, and finally to the adaptive importance sampling.

In general, computational efficiency is proportional to the desired accuracy and the choice of an approach will depend on the nature of the problem. However, this report demonstrates that the limit state-based approach is efficient by focusing on a single response or probability level. This permits the analyst to concentrate on the critical performance region. In addition, the approach naturally leads to probabilistic sensitivity analysis. In particular, the efficiency of the alternative approach is independent of the probability level, therefore it is most suitable for evaluating the tails (i.e., CDF close to 0 or 1) of the distribution. However, the limit state is relatively difficult to implement.

The computational efficiency of the limit state approach in general depends on the number of random variables, n, the way the sensitivities are computed, and the required number of sensitivity calculations. Assuming that a numerical differentiation scheme is used, the number of limit states of interest is p, and the number of sensitivity calculations using an iteration procedure is k, then the required number of response-function calculations will be in the order of $n \cdot p \cdot k$. To minimize this number, efficient methods should be fully investigated to minimize the required k.

When n is very large[1], the alternative approach may no longer be efficient unless the sensitivities can be determined by other more efficient approaches including the perturbation method and the adjoint method. It is also possible, as proposed in Appendix B, to minimize sensitivity updating and increase efficiency by focusing more on the important random variables during the iteration analysis. The stochastic finite element expansion for random media (Spanos and Ghanem 1989) is another possibility for simplifying the problems associated with a large number of random variables due to finite element discretization.

[1] Since the efficiency of the standard Monte Carlo method depends on the probability level and the desired accuracy, it is difficult to define quantitatively what is "large." Assuming that 10,000 samples is required for a Monte Carlo analysis, $p = 2$, and $k = 5$, then the limit state will lose its efficiency edge if n is greater than 1,000. Notice the additional advantage offered by the limit state approach, that of producing probabilistic sensitivity results essentially as a by-product.

7 REFERENCES

Ang, A. H.-S., and W. H. Tang. 1984. *Probability Concepts in Engineering Planning and Design. Volume II: Decision, Risk, and Reliability*. New York: John Wiley & Sons, Inc.

Breitung, K. 1984. Asymptotic Approximations for Multinormal Integrals. *Journal of Engineering Mechanics* 110 (3): 357-366. American Society of Civil Engineers (ASCE).

Breitung, K. 1989. Asymptotic Approximations for Probability Integrals. *Probabilistic Engineering Mechanics* 4 (4): 187-190.

Decker, K. M. 1991. The Monte Carlo method in science and engineering: Theory and application. *Computer Methods in Applied Mechanics and Engineering* 89: 463-483. Elsevier Science Publishers.

Dias, J. B. and J. C. Nagtegaal. 1985. Efficient Algorithms for Use in Probabilistic Finite Element Analysis. *Advanced in Aerospace Structural Analysis* AD-09: 37-50. American Society of Mechanical Engineers (ASME).

Dolinski, K. 1990. Improvement of Second-Order Reliability Estimates by Importance Sampling. *Journal of Engineering Mechanics* 116 (7): 1668-1670. ASCE.

Fiessler, B., H. J. Neumann, and R. Rackwitz. 1979. Quadratic Limit States in Structural Reliability. *Journal of Engineering Mechanics* 105: 661-676. ASCE.

Grogoriu, M. 1983. Approximate Analysis of Complex Reliability Problems. *Structural Safety* 1: 277-288.

Gureghian, A. B., Y.-T. Wu, B. Sagar, and R. B. Codell. 1991. Sensitivity and Uncertainty Analyses Applied to One-Dimensional Transport in a Layered Fractured Rock. *Part 1: Analytical Solutions and Local Sensitivities*. CNWRA 91-010. San Antonio, Texas: Center for Nuclear Waste Regulatory Analyses (CNWRA).

Harbitz, A. 1986. An Efficient Sampling Method for Probability of Failure Calculation. *Structural Safety* 3: 109-115.

Harper, W. V. 1983. *Sensitivity/Uncertainty Analysis Techniques for Nonstochastic Computer Codes*. BMI/ONWI-444. Battelle Memorial Institute (BMI), Office of Nuclear Waste Isolation (ONWI).

Harper, W. V. and S. K. Gupta. 1983. *Sensitivity/Uncertainty Analysis of a Borehole Scenario Comparing Latin Hypercube Sampling and Deterministic Sensitivity Approaches*. BMI/ONWI-516. BMI, ONWI.

Hasofer, A. M. and N. C. Lind. 1974. Exact and Invariant Second - Moment Code Format. *Journal of Engineering Mechanics* 100 (EMI): 111-121. ASCE.

Helton, J. C., J. W. Garner, R. D. McCurley, and D. K. Rudeen. 1991. Sensitivity Analysis Techniques and Results or Performance Assessment at the Waste Isolation Pilot Plant. SAND90-7103. Albuquerque, New Mexico: Sandia National Laboratories (SNL).

Hohenbichler, M. and R. Rackwitz. 1981. Non-normal Dependent Vectors in Structural Safety. *Journal of Engineering Mechanics* 100 (EM6): 1227-1238. ASCE.

Hohenbichler, M. and R. Rackwitz. 1988. Improvement of Second-Order Reliability Estimates by Importance Sampling. *Journal of Engineering Mechanics* 114 (12): 2195-2199. ASCE.

Iman, R. L., J. M. Davenport, E. L. Frost, and M. J. Shortencarier. 1980. *Stepwise Regression with PRESS and Rank Regression (Program User's Guide)*. SAND79-1472. Albuquerque, New Mexico: SNL.

Iman, R. L. and M. J. Shortencarier. 1984. A FORTRAN 77 Program and user's guide for the generation of Latin Hypercube and random samples for use with computer models. NUREG/CR-3624. Washington, D. C.: U.S. Nuclear Regulatory Commission (NRC).

Karamchandani, A. 1987. *Structural System Reliability Analysis Methods*. 83. Stanford, California: The John A. Blume Earthquake Engineering Center, Dept. of Civil Engineering.

Liebetrau, A. M. and P. G. Doctor. 1987. *The Generation of Dependent Input Variable to a Performance Assessment Simulation Code*. Proceedings of a Nuclear Energy Agency Workshop on Uncertainty Analysis for Performance Assessments of Radioactive Waste Disposal Systems: 85-115.

Liu, P.-L. and A. Der Kiureghian. 1986. Multivariate Distribution Models with Prescribed Marginals and Covariances. *Probabilistic Engineering Mechanics* 1 (2): 105-115.

Liu, P.-L. and A. Der Kiureghian. 1986. Optimization Algorithms for Structural Reliability Analysis. UCB/SESM-86/09. Berkeley, California: Department of Civil Engineering, University of California, Berkeley (UCB).

Madsen, H. O., S. Krenk, and N. C. Lind. 1986. *Methods of Structural Safety*. Englewood Cliffs, New Jersey: Prentice Hall, Inc.

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* 21 (2).

Melchers, R. E. 1989. Importance Sampling in Structural Systems. *Structural Safety* 6: 3-10.

Rackwitz, R. and B. Fiessler. 1978. Structural Reliability Under Combined Load Sequences. *Journal of Computers and Structures* 9: 489-494.

Rosenblatt, M. 1952. Remarks on a Multivariate Transformation. *The Annals of Mathematical Statistics* 23 (3): 470-472.

Rubinstein, R. Y. 1981. *Simulation and the Monte Carlo Method*. New York: John Wiley & Sons.

Sagar, B. and P. M. Clifton. 1984. *Stochastic Groundwater Flow Modeling Using the Second-Order Method*. RHO-BW-SA-346P. Rockwell International, Rockwell Hanford Operations (RHO).

Sitar, N., J. D. Crawfield, and A. Der Kiureghian. 1987. First-Order Reliability Approach to Stochastic Analysis of Subsurface Flow and Containment Transport. *Water Resources Research* 23 (5): 794-804.

Spanos, P. D. and R. Ghanem. 1989. Stochastic Finite Element Expansion for Random Media. *Journal of Engineering Mechanics* 115 (4): 1035-1053. ASCE.

Stein, M. 1987. Large Sample Properties of Simulations Using Latin Hypercube Sampling. *Technometrics* 29 (2): 143-151.

Thoft-Christensen, P. and M. J. Baker. 1982. *Structural Reliability Theory and Its Applications*. New York: Springer-Verlag.

Thomas, R. T. 1982. *Uncertainty Analysis*. BMI/ONWI-380. BMI, ONWI.

Tvedt, L. 1990. Distribution of Quadratic Forms in Normal Space - Application to Structural Reliability. *Journal of Engineering Mechanics* 116 (6): 1183-1197. ASCE.

Wu, Y.-T. and P. H. Wirsching. 1984. Advanced Reliability Methods for Fatigue Analysis. *Journal of Engineering Mechanics* 110 (4). ASCE.

Wu, Y.-T. and O. H. Burnside. 1988. Validation of the NESSUS Probabilistic Analysis Computer Program. *Proc. of the 29th AIAA/ASME/ASCE/AHS Structures, Structural Dynamics and Materials Conference* Part 3: 1267-1274.

Wu, Y.-T., O. H. Burnside, and T. A. Cruse. 1988. Probabilistic Methods for Structurable Response Analysis. W. K. Liu and T. Belytschko, eds. *Computational Mechanics of Reliability Analysis: 181-195*. Lausanne, Switzerland: Elmepress International.

Wu, Y.-T., H. R. Millwater, and T. A. Cruse. 1990. An Advanced Probabilistic Structural Analysis Method for Implicit Performance Functions. *American Institute of Aerospace and Astronautics (AIAA) Journal* 28 (9): 1663-1669.

Wu, Y.-T. and T. Y. Torng. 1990. *A Fast Convolution Procedure for Probabilistic Engineering Analysis*. Proceedings of the First International Symposium on Uncertainty Modeling and Analysis. Institute of Electrical and Electronics Engineers (IEEE) Computer Society: 72-77.

Wu, Y.-T., A. G. Journal, L. R. Abramson, and P. K. Nair. 1991. Uncertainty Evaluation Methods for Waste Package Performance Assessment. *NUREG/CR-5639*. Washington, D.C.: NRC.

APPENDIX A

CDF ANALYSIS BY THE ADVANCED MEAN VALUE (AMV) PROCEDURE - ILLUSTRATIVE EXAMPLE

CDF Analysis by the Advanced Mean Value (AMV) Procedure - Illustrative Example

Introduction

A performance function is used to illustrate the step by step procedure of the advanced mean value (AMV) method which was discussed in Section 4.8 and applied to the transport example problem in Section 5. For illustration purposes, the performance function selected is a simple second-order equation with only two normal random variables to allow a closed-form AMV solution. It should be noted that the AMV method can be used for more complicated problems involving implicitly defined functions and non-normal random variables.

Performance Function

The performance function is:

$$Z = X_1^2 + X_2^2 \tag{A-1}$$

in which the two input two random variables have normal distributions. The mean values are 10 for both variables, and standard deviations are 1 and 2, respectively.

MV Analysis

Denote mean values as $\mu_i (i = 1,2)$ and standard deviations as $\sigma_i (i = 1,2)$. Taking a Taylor's series expansion at the mean values and including only up to the first-order terms, the MV approximation (see Section 4.8) is:

$$Z_{MV} = (\mu_1^2 + \mu_2^2) + 2\mu_1 (X_1 - \mu_1) + 2\mu_2 (X_2 - \mu_2) = 20X_1 + 20X_2 - 200 \tag{A-2}$$

In order to define the most-probable-point (MPP), the original random variables must first be transformed to standardized normal variables, u_i. Because X_i are normal variables, the transformations are:

$$u_1 = (X_1 - \mu_1)/\sigma_1 = X_1 - 10 \tag{A-3}$$

$$u_2 = (X_2 - \mu_2)/\sigma_2 = (X_2 - 10)/2 \tag{A-4}$$

Using Eqns. (A-3) and (A-4), Eqn. (A-2) can be expressed as:

$$Z_{MV} = 20u_1 + 40u_2 + 200 \tag{A-5}$$

At the mean-values point, $u_1 = u_2 = 0$, therefore $Z_{MV} = 200$. Now consider a limit $Z_{MV} = 200$. The limit state surface will be represented by:

$$200 = 20u_1 + 40u_2 + 200$$

or

$$u_1 = -2u_2, \tag{A-6}$$

which is the equation of a straight line (Figure A-1) passing through the origin. Other Z_{MV}-based limit state surface will be parallel to the line defined by Eqn. (A-6). For any limit state the MPP is a point with the smallest distance. Based on Eqn. (A-2), the most-probable-point-locus (MPPL) of Z_{MV}, shown in Figure A-1, is a line passing through the origin and perpendicular to the limit states. The equation of this MPPL:

$$u_1 = u_2/2. \tag{A-7}$$

AMV Analysis

The AMV approximation is expressed as (Section 4.8):

$$Z_{AMV} = Z_{MV} + H(Z_{MV}) \tag{A-8}$$

where $H(Z_{MV})$ is defined as the difference between the values of Z_{MV} and Z calculated at points on the MPPL of Z_{MV}.

For the example under consideration, substitute Eqn. (A-7) into Eqn. (A-5) to get:

$$Z_{MV} = 100u_1 + 200 \tag{A-9}$$

For a given Z_{MV} value (limit state), the corresponding MPP can be calculated, based on Eqn. (A-9), as follows:

$$u_1^* = (Z_{MV} - 200)/100, \qquad u_2^* = 2u_1 = (Z_{MV} - 200)/50 \tag{A-10}$$

The corresponding X-values, using Eqns. (A-2) and (A-3), are:

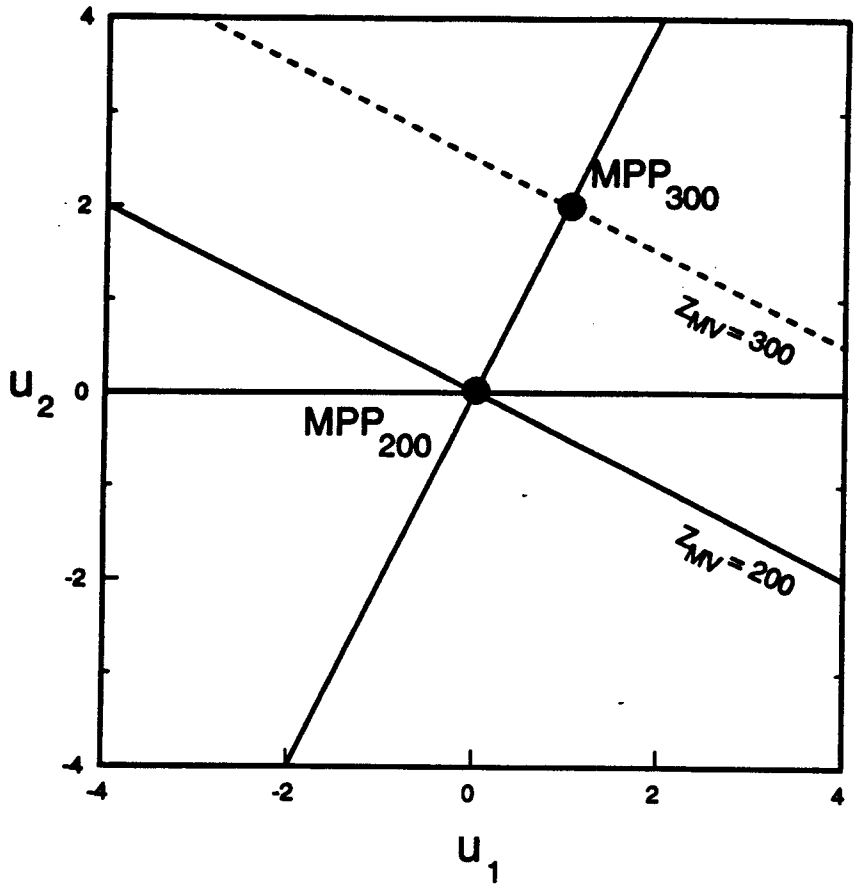

Figure A-1. Most probable point locus (MPPL) of Z_{MV}

$$X_1^* = (Z_{MV} + 800)/100, \qquad X_2^* = (Z_{MV} + 50)/25 \tag{A-11}$$

The exact value of Z at the MPP is $Z_{EXACT} = (X_1^*)^2 + (X_2^*)^2$. Therefore, the function H of Eqn. (A-8) is:

$$H(Z_{MV}) = Z(X_1, X_2) - Z_{MV} \tag{A-12}$$

In general, the H-function could be evaluated numerically assuming an algorithm, possibly a computer code, for calculating the Z-function is available. In the present example closed-form H-function can be obtained by substituting Eqn. (A-11) into Eqn. (A-12), i.e.,

$$H(Z_{MV}) = (X_1^{*2} + X_2^{*2}) - Z_{MV}$$
$$= 17(Z_{MV} - 200)^2/10000 \tag{A-13}$$

Using Eqn. (A-2) and (A-13), the AMV approximation [Eqn. (A-8)] can be written explicitly in terms of X_1 and X_2:

$$Z_{AMV} = Z_{MV} + 17(Z_{MV} - 200)^2/10000$$
$$= (20x_1 + 20X_2 - 200) + 17(20X_1 + 20X_2 - 400)^2/10000 \tag{A-14}$$

Note that Z_{AMV} is a nonlinear function.

CDF Analysis by the AMV Method

To establish the CDF of Z based on the AMV procedure, select several z values and perform the following analysis for each value.

1. Compute the CDF based on Z_{MV} [Eqn. (A-2)].

2. Compute the MPP (X_1^*, X_2^*) based on $Z_{MV} = z$.

3. Calculate $Z(X_1^*, X_2^*)$ [Eqn. (A-1)].

4. For the calculated CDF in step 1, update z-value by the value calculated in step 3 and repeat.

For the example under consideration, CDF's with nine z-values are shown in Figure A-2. Observe that the CDF based on the AMV diverges from the one based on MV in the tail regions.

Based on the AMV solution, iteration procedures (discussed in Section 4.8) may be applied to search for the exact MPP, which is used to develop a better linearization function. Further enhancement of the CDF results may be achieved efficiently by using the adaptive importance sampling method (see Section 4.9).

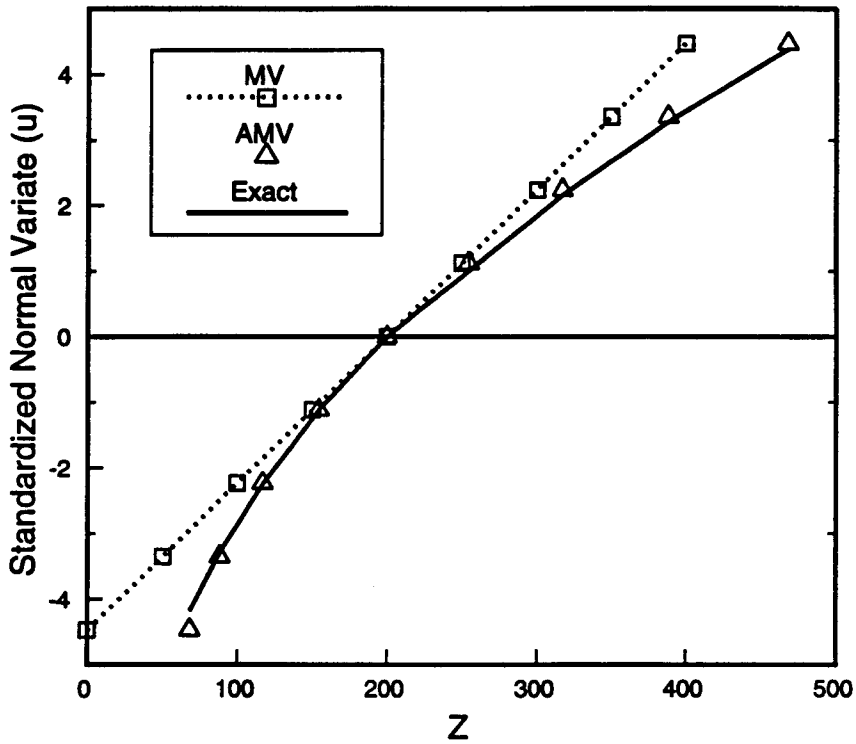

Figure A-2. CDFs of Z_{MV} and Z_{AMV}

APPENDIX B

A SENSITIVITY UPDATING ALGORITHM FOR PROBLEMS WITH LARGE NUMBER OF INPUT RANDOM VARIABLES

A Sensitivity Updating Algorithm for Problems with Large Number of Input Random Variables

Introduction

As described in Section 4, the required number of performance function calculations is approximately proportional to the number of random variables times the number of sensitivity updates during an iteration process. Consequently, when the number of input random variables (n) is large and the sensitivities are calculated by a numerical differentiation scheme, the efficiency advantage over the Monte Carlo approach may disappear. This appendix proposes a sensitivity updating algorithm to minimize the required number of full-size sensitivity calculations. The feature of this algorithm is to keep all the random variables during the iteration analysis, but the sensitivities are updated only for the important random variables until the last step where the sensitivities for all the random variables are updated. The algorithm seems to be most suitable for problems where the number of dominant random variables is small relative to n.

Proposed Method

The basic idea is to use approximate sensitivities for less-sensitive random variables until the final stage in the iteration process. This is based on the fact that for less-sensitive random variables, the MPPs are close to the median values, and therefore the mean or median-based sensitivities might provide good approximations.

Assume that the AMV model (see Section 4) is obtained and that the random variables are separated into two groups: important and less important. The AMV model can be written as follows:

$$Z_{AMV} = Z_{MV} + H(Z_{MV}) = a_0 + \sum_{i=1}^{n_A} a_i(X_i - \mu_i) + \sum_{j=1}^{n_B} a_j(X_j - \mu_j) + H(Z_{MV}) \qquad \textbf{(B-1)}$$

where $\mathbf{Z}_{AMV}$ is the linear approximation to the exact **Z**-function, $\mathbf{Z}_{MV}$ is the mean-value approximation, $\mathbf{H}(\mathbf{Z}_{MV})$ is the correction function, n_A is the number of important random variables, and $n_B = n - n_A$ is the number of less-important random variables. The selection of important random variables can be based on the probabilistic sensitivity factors (s) described in Section 4. A cut-off s value may be selected.

In the AMV method, for every limit state $\mathbf{Z}_{MV} = z$, an MPP, x^* can be found. Based on x^*, an updated linear function needs to be established to generate the next MPP. The process is repeated until the exact MPP is found.

A first-order Taylor expansion of **Z** at x^* is:

$$Z = Z(X^*) + \sum_{i=1}^{n} \frac{\partial Z}{\partial X_i}\Big|_{x^*} (X_i - x_i^*) \tag{B-2}$$

where each $\partial Z/\partial X_i$ can be computed by perturbing x_i from x_i^* to $x_i^* + \Delta x_i$ and computing the change in **Z**.

Rewriting the linearized function as:

$$Z = a_0^k + \sum_{i=1}^{n_A} a_i^k (X_i - x_i^*) + \sum_{j=1}^{n_B} a_j^k (X_j - x_j^*) \quad ; \quad n_A + n_B = n \tag{B-3}$$

where **k** is the iteration number. The regular sensitivity updating procedure will be applied only to $\mathbf{n_A}$ random variables. For the $\mathbf{n_B}$ random variables, the sensitivities, $\mathbf{a_j^k}$, are estimated based on the AMV model, Eqn. (B-1), as follows:

$$a_j^k = a_j + \frac{\partial H}{\partial X_j} \tag{B-4}$$

where

$$\frac{\partial H}{\partial X_j} = \frac{\partial H}{\partial Z_{MV}} \cdot \frac{\partial Z_{MV}}{\partial X_j} = \frac{\partial H}{\partial Z_{MV}} a_j \tag{B-5}$$

With

$$c = \frac{\partial H}{\partial Z_{MV}} \tag{B-6}$$

$$a_j^k = a_j (1 + c) \tag{B-7}$$

The factor c accounts for nonlinear effect, i.e., when the Z-function is linear, **H** and c would be zero. According to the definition of **H**, the c value can be compute by evaluating a perturbed **H** value at a perturbed probability level. This requires one Z-function calculation regardless of the number of random variables.

Substituting Eqn. (B-7) for Eqn. (B-3), the updated linearized function is

$$Z = a_0^k + \sum_{i=1}^{n_A} a_i^k (X_i - x_i^*) + \sum_{j=1}^{n_B} a_j (1+c)(X_j - x_j^*) \tag{B-8}$$

which is then used to obtain the next MPP. The coefficients a_j will remain constant while a_0^k and a_i^k will be updated for each new MPP until the MPP converges.

In general, the less-important random variable defined in the MV or AMV analyses may turn out to be important after the MPP converges. Thus, after the above analysis, the a_j^K terms should generally be updated for a full iteration. If a significant change in the sensitivity factor is found for an unimportant random variable, this variable may be re-defined as an important random variable. Similarly, a previously important random variable may be re-defined as a less important variable. Subsequently, the sensitivities are updated only for the new set of n_A random variables until the exact MPP is obtained. The process may be repeated as needed before the final full sensitivity analysis starts.

Since at the final stage of the iteration procedure all the random variables will be included, the selection of the cut-off s-value, while affecting the computational efficiency, should not influence the final result.

It can be anticipated that the algorithm will be most effective when only a small number of random variables are dominant (i.e, $0 < n_A << n_B$). In the extreme cases where all the random variables are roughly of equal importance (therefore, $n_B = n$), the proposed algorithm converges to the regular procedure and no efficiency improvement can be made. However, in such cases, since the MPP's for all the random variables will be close to the mean or median points, it is likely that the required number of iterations will be small, as shown below.

At the MPP,

$$\beta^2 = \sum_{i=1}^{n} (u_i^*)^2 \tag{B-9}$$

If the sensitivity factors, s_i, are equal, all u_i^* are equal. Then, from Eqn. (B-9),

$$u_i^* = \frac{\beta}{\sqrt{n}} \tag{B-10}$$

Thus, when n increases, u_i^* will approach zero, meaning x_i^* will approach the median and the initial sensitivity results at the mean values will remain accurate.