UNCERTAINTY ANALYSIS OF NUCLEAR WASTE PACKAGE CORROSION

Robert E. Kurth and Stephen L. Nicolosi
Battelle's Columbus Laboratories
Columbus, Ohio

ABSTRACT

This paper describes the results of an evaluation of three uncertainty analysis methods for assessing the possible variability in calculating the corrosion process in a nuclear waste package. The purpose of the study is the determination of how each of three uncertainty analysis methods, Monte Carlo, Latin hypercube sampling (LHS) and a modified discrete probability distribution method, perform in such calculations. The purpose is not to examine the absolute magnitude of the numbers but rather to rank the performance of each of the uncertainty methods in assessing the model variability. In this context it was found that the Monte Carlo method provided the most accurate assessment but at a prohibitively high cost. The modified discrete probability method provided accuracy close to that of the Monte Carlo for a fraction of the cost. The LHS method was found to be too inaccurate for this calculation although it would be appropriate for use in a model which requires substantially more computer time than the one studied in this paper.

INTRODUCTION

This study examines different techniques which can be applied to the evaluation of uncertainties in waste package performance. In particular, the uncertainty in the corrosion depth in a nuclear waste package is evaluated using three different techniques. The corrosion model selected for use in the current analysis is one suggested by Sastre and Pescatore[1]. It is important to note that the results presented here, represented by mean values, do not necessarily correspond to the "engineering" best estimate value of the corrosion depth of a nuclear waste package. The model was selected because there is probabilistic data available for the model inputs. Thus, this model provides a good test of the uncertainty analysis techniques since it is not overly simplistic from a physical process standpoint and some data is available. Similarly, the upper and lower bounds of the corrosion depth given below only are meaningful relative to each other--the absolute value is not to be taken as presenting reality. The reasons for this is, that while these physical models are based on the physical phenomena, they are not complete and they are idealized. The model is appropriate for assessing the applicability and performance of uncertainty analysis methods, however, it is not appropriate to attach significance to the actual numeric values which result.

The three uncertainty analysis methods chosen for study are Monte Carlo simulation, Latin Hypercube Sampling (LHS), and Discrete Probability Distribution (DPD) methods. These methods were selected because they are currently the most widely used techniques in the nuclear industry. The DPD method has been found to be deficient in many analyses because of computer storage limitations. This paper describes an application of a new algorithm which allows the mechanisms of the DPD technique to be applied without regard to computer storage requirements. While the method is derived from the underlying principles inherent to the DPD method it is closer to a hybrid method of Monte Carlo, DPD, and Latin square sampling in which the unique features of each method have been retained in the development of a new program which has been called RASCAL[2].

The remainder of this paper is structured as follows. First, the corrosion model is described and the necessary input variables are defined. The results of the time dependent uncertainty analysis are given next. The evaluation of the performance of the Monte Carlo, LHS, and RASCAL methods is then presented. The next section presents a qualitative discussion on the use of these methods in probabilistic and sensitivity analysis. Finally, a brief summary is given.

CORROSION MODEL FOR WASTE PACKAGE EVALUATION

The corrosion model used for this study is one suggested by Sastre and Pescatore[1]. The general form is given by

where

- C_D: corrosion (mm)
- Ku: uniform corrosion factor
- Kp: pitting corrosion factor
- T: temperature (K)
- O: oxygen concentration (ppm)
- Cl: chlorine concentration (ppm)
- t: time (years)
- a,b,c,n: empirical constants

The constants in Eq. (1) are determined by a regression analysis of data developed by Westinghouse[2]. The calculation of the temperature, T, is performed using a one-dimensional coordinate system in which the package, overpack, and backfill regions are represented by concentric circles.

In order to check the computer model developed for this study, the 0.1 to 99.9 percentile spread for the uniform corrosion facts, Ku, were calculated and compared to the results of reference 1. The range computed in this reference for Ku was (0.00147, 676) while the computer model developed for this study gave (0.00157, 636). This range was found to be within the needed accuracy for this study since the ultimate goal is the evaluation of uncertainty analysis methods not an accurate prediction of the corrosion depth.

Input Variable Definition

For the corrosion model presented in the previous chapter, there are a total of 16 inputs to the model.

Of these 16 inputs, 10 are considered to be random variables. Eight of the random variables are described by uniform distributions, one by a normal distribution, and the last is described by a special form of a Weibull distribution, called a Rayleigh distribution in the probabilistic literature. Each of these variables, both constant and stochastic, are shown in Table I with their associated values or distribution type and parameters.

In Table I, the first parameter listed is the lower limit for a uniform distribution, and the minimum value for the Rayleigh distribution. The discretization of each of the random variable input distributions was done in the same way for both the LHS and RASCAL analyses, and is shown in Table II for the case when the number od discrete intervals is equal to 10. When the number of discrete points is different from 10, the same procedure is used to generate results corresponding to Table II.

Time Dependent Uncertainty Analysis Results

The first effect to be examined is that of pitting corrosion. Because of the variation associated with the uniform pitting corrosion factors, rate equations were used for both the uniform and pitting corrosion models. Thus,

$$U(T + t) = U(T) + r\ U(t)/\ t \qquad (2)$$

$$P(T + t) = P(T) = t\ P(t)/\ t \qquad (3)$$

where U(t) and P(t) are the predictive equations for uniform and pitting corrosion, respectively, and U(0) and P(0) are both equal to zero. In Eq. (2), the exponent of time, i.e., variable 15 from Table I, is held constant at its mean value, while it is normally distributed, with parameters given in Table I, for use in Eq. (3).

The predicted corrosion depth variability is examined by each of the three uncertainty analysis methods. For the sake of comparison, a Monte Carlo analysis was performed using 1000 sample runs for the calculations. Since it is straightforward to calculate confidence bands for the mean value of the corrosion depth, this analysis is assumed to represent "truth", i.e., the correct value. The mean corrosion depth as calculated by the Monte Carlo method for years 100 to 1,000, and the 95% confidence limits are given in Table II.

The DPD mehtod suffers from the drawback that for the analysis currently being performed there are 16 variables and a minimum of 10 discrete points. Since many of these variables are dependent and, therefore, condensation cannot be performed, this implies that at least 2.0E+10 storage points are required. Obviously, this is outside the limits of most fixed computer storage requirements. However, the RASCAL algorithm is capable of analyzing this problem and is used in this case.

The results of the RASCAL and LHS calculations are shown in Table III and IV, respectively. The result of the RASCAL case when 10 discrete points are used, denoted RASCAL-10, are very good but were actually too good to believe. As can be seen from the RASCAL run when 20 points are used, denoted RASCAL-20, the results when 10 points are used were evidently a result of fortuitous circumstances. However, as was expected, the RASCAL method is always better than the LHS method, since no replicates of the LHS design were used.

In the LHS analysis, there appears to be a bias being introduced into the results. For the 10 strata cases, the mean prediction gets closer to the Monte Carlo result as time increases, while for the 20 strata analysis, the reverse is true. One possible explanation of this result is that the uniform corrosion factor is causing the conditional mean of each strata to be skewed to the right, i.e., towards higher values. Because the RASCAL analysis is providing an estimate of the distribution within each strata while the LHS method is not, the bias will be introduced until enough combinations have been made for the distribution to approach a normal distribution or until enough strata have been used to reduce the bias. Alternatively, replicate designs using LHS can be used to assimilate the effects of such skewness.

While some indication of the RASCAL and LHS methods effectiveness can be obtained from an examination of the mean values, the ultimate goal is to assess each method's ability to indicate the uncertainty in the model. Figure 1 presents a plot of the standard deviation calculated by Monte Carlo, RASCAL, and LHS analysis. From this figure, it is apparent that the RASCAL method of analysis done considerably better than the LHS method in predicting the standard deviation which is one measure of the uncertainty in the response of the model. For the reasons discussed above, this result is to be expected since the RASCAL method uses many more sample points than the LHS method.

UNCERTAINTY ANALYSIS PERFORMANCE EVALUATION

Because of the complex nature of the function being evaluated, there is no analytic solution to predict what the possible spread in the distribution of the corrosion depth may be. Therefore, the Monte Carlo results will be assumed to represent "truth" in the estimate of the spread in the distribution. This has the added advantage that, for by using the results of the Monte Carlo analysis, estimates of the confidence levels for various statistical measures can be made.

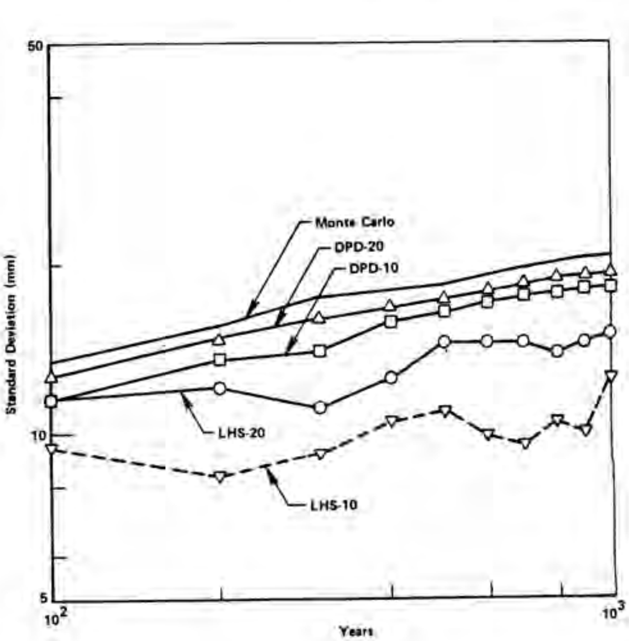

Fig. 1. Standard Deviation Predictions.

TABLE I

CORROSION MODEL INPUT DATA DESCRIPTION

Variable	Distribution Type
Rock Parameters	
1.Geothermal Temperature (C)	Uniform
2.Thermal Conductivity (W/M/K)	Uniform
3.Rock Density (Kg/M**3)	Uniform
4.Specific Heat (J/Kg/K)	Uniform
Waste Parameters	
5.Waste Age (Years)	Constant
6.Initial Power (W)	Constant
7.Backfill Outer Diameter (M)	Constant
8.Backfill Conductivity (W/M/K)	Uniform
9.Overpack Outer Diameter (M)	Constant
10.Canister Wall Thickness (M)	Constant
11.Canister Length (M)	Constant
Corrosion Parameters	
12.Pitting Corrosion Factor	Uniform
13.Chlorine Concentration (ppm)	Uniform
14.Oxygen Concentration (ppm)	Uniform
15.Exponent of Time	Normal
16.Uniform Corrosion Parameter	Rayleigh

Variable Number	Parameter 1	Parameter 2
1	54.0	60.0
2	1.25	2.5
3	390.0	2410.0
4	820.0	1160.0
5	0.0	NA
6	2100.0	NA
7	0.686	NA
8	0.40	1.4
9	0.325	NA
10	0.053	NA
11	4.10	NA
12	1.00	6.0
13	1.00	101.0
14	0.01	3.0
15	0.469	0.0349
16	0.01	118.643

TABLE II

MEAN CORROSION DEPTH AS PREDICTED BY MONTE CARLO SIMULATION

Year	Lower Bound (mM)	Mean (mM)	Upper Bound (mM)
100	15.4	15.9	16.3
200	25.1	25.7	26.5
300	33.1	33.8	34.4
400	39.8	40.2	40.7
500	45.5	46.0	46.6
600	50.8	51.6	52.3
700	55.3	56.2	56.7
800	60.2	60.8	61.5
900	64.3	64.9	65.6
1,000	68.4	68.9	69.9

TABLE III

MEAN CORROSION DEPTH AS PREDICTED BY RASCAL SIMULATION

Year	Monte Carlo	RASCAL-10	RASCAL-20
100	15.9	15.6	15.7
200	25.7	25.1	25.4
300	33.8	33.3	33.0
400	40.2	40.4	39.3
500	46.0	46.5	45.2
600	51.2	51.3	50.6
700	56.2	56.3	55.3
800	60.8	60.6	59.7
900	64.9	64.7	63.8
1,000	68.9	68.7	67.9

TABLE IV

MEAN CORROSION DEPTH AS PREDICTED BY LHS SIMULATION

Year	Monte Carlo	LHS-10	LHS-20
100	15.9	14.1	11.6
200	25.7	22.6	32.9
300	25.8	29.7	32.9
400	40.2	35.5	39.4
500	46.0	41.1	44.5
600	51.6	46.4	49.2
700	56.2	52.1	53.8
800	60.8	57.2	58.2
900	64.9	60.4	62.5
1,000	68.9	65.1	66.3

This immediately raises the question of, "What measure should be used to represent the uncertainty in the spread of the distribution of corrosion depths?" In many uncertainty analyses (see, for example, reference 3), the measure of the range of possible responses is the variance. In this evaluation, variances are not used for two reasons. First, the Monte Carlo estimate is known to be less accurate for estimating higher order moments; and since this is to represent truth, a more reasonable measure should be examined. Secondly, since so few points are used in the LHS method, it would automatically be at a disadvantage in such an evaluation since the number of points used to estimate the variance would not be sufficient to ever produce a reasonable estimate of the variance. Therefore, the measure of the uncertainty in the response to be used in this evaluation is percentiles. The percentile measure of uncertainty offers two distinct advantages: (1) confidence level bounds can be obtained for the Monte Carlo results, and (2) it is known that for both the RASCAL and LHS methods as the number of discrete points is increased, the percentile values must approach the Monte Carlo values.

Percentile Measure of Uncertainty and Confidence Levels

The result of a Monte Carlo analysis will be a vector of N responses where N is the number of Monte Carlo simulations. If this vector of responses, denoted V, is ordered then an empirical cumulative distribution function (CDF) can be constructed. The first value of V, v(1), is the value of the 1.0/N percentile value. For example, in the Monte Carlo analysis performed for this study, N is equal to 1,000. Thus, v(1) is the estimate of the 0.1 percentile value of the distribution of V, in this case the corrosion depth. Similarly, v(10) is the estimate of the first percentile value, v(100), the estimate of the tenth percentile value, and so on. If the M percentile value is desired, then it can be shown that, at a confidence level of g, this value is bounded by

Mu = int[m + Z((1 + g)/2)*(m*(1 - m/N))**0.5] + 1

Ml = int[m - Z((1 + g)/2)*(m*(1 - m/N))**0.5] + 1

where m = MN/100, int[ ] is the integer part of the term inside the square brackets, and Z(x) is the value of the standard normal distribution at the point x. For a confidence level of 95%, i.e., we are 95% confident that the actual value of the M percentile value lies between Ml and Mu, Z(0.975) is equal to 0.83525. Given these parameters, the values and confidence levels of the Monte Carlo analysis for the 5th, 10th, 90th, and 95th percentile values are given in Table V.

Table VI gives the results of the percentile calculations for the RASCAL-20 and LHS-20 cases, together with the Monte Carlo result, at year 1,000. As the figures show, the RASCAL over estimates the spread. The RASCAL analysis does predict the 10th and 90th percentile value within the accuracy of the Monte Carlo result while the LHS result only falls within the range of the 95th percentile value. Given the poor performance of the LHS prediction for the other values examined, it is not believed that this result is an indication of anything other than good fortune. Additionally, in the analysis, the response is the corrosion depth and, therefore, an over estimate of the spread represents a conservative result.

On method for guaging the effectiveness of each method's ability to predict the uncertainty is to assume one is interested in estimating the spread between the 10th and 90th percentile values. Table VII gives the predicted values of each percentile obtained from this table, the RASCAL calculation reaches the "correct" value much more quickly than the LHS calculation. The indication is that if the 5th and 95th percentile values are important, then a RASCAL calculation using 40 discrete points would be more appropriate.

The final consideration for recommending one method over another is the amount of computer time taken for each of the analyses. Table VIII gives the execution time used for each analysis on a CDC-855 series computer. The actual cost of the runs will vary from these figures because of different input and output requirements, plotting, etc. However, the values in Table VIII represent a fair comparison of the computational efficiencies. The results indicate that the RASCAL analysis could use up to 45 discrete points before it would become as expensive as the Monte Carlo analysis while the LHS method could use up to 81 strata. However, given the increased accuracy of the RASCAL method at estimating the tail values of the response for the number of discrete intervals examined, 20, and its greater flexibility, it is believed that the RASCAL method is a better method for performing uncertainty analysis.

Probabilistic and Sensitivity Analysis

The ability of the Monte Carlo analysis to predict classic statistical quantities of interest, e.g., mean, variance, confidence levels, etc., gives it a great appeal. However, it must be recalled that the corrosion model being studied has been greatly simplified and is relatively cheap to evaluate. When more accurate or better physical descriptions become available, it is expected that each evaluation will require a substantial amount of computer time. At that point, it may become prohibitively costly to perform a Monte Carlo analysis even using stratified or importance sampling algorithms in conjunction with the Monte Carlo scheme. Then it will be crucial to have in place a method to not only examine the uncertainty of these new models, but also to use for probabilistic modeling and/or sensitivity studies.

TABLE V

PERCENTILE VALUES AND CONFIDENCE LEVELS FROM MONTE CARLO SIMULATION

Percentile	Lower Bound	Value	Upper Bound
5	37.66	38.07	38.89
10	42.72	43.45	43.98
90	94.96	96.16	97.83
95	103.5	103.9	105.9

TABLE VI

PERCENTILE ESTIMATES FROM EACH OF THE UNCERTAINTY METHODS

%TILE	Monte Carlo Value	RASCAL-20 Value	LHS-20 Value
5	38.07	33.22	49.15
10	43.45	43.62	50.14
90	96.16	96.45	88.68
95	103.9	116.7	104.6

TABLE VII

PREDICTION OF SELECTED PERCENTILES USING DIFFERENT NUMBERS OF DISCRETE POINTS OR STRATA

Method Used	10th %tile	90th %tile
Monte Carlo	43.45	96.16
RASCAL-10	42.86	108.16
RASCAL-20	43.62	96.45
LHS-10	50.06	83.43
LHS-20	50.14	88.68

TABLE VIII

COMPUTER TIME REQUIRED FOR EACH UNCERTAINTY ANALYSIS

Method	Time (Seconds)
Monte Carlo	572
RASCAL-10	96
RASCAL-20	254
LHS-10	51
LHS-20	69

The sensitivity analysis of a function or computer program, as was discussed above, is not overly concerned with the prediction of the PDF of the response but rather is concerned with assessing the variability of the response due to changes in the input variables when these changes are small and near the design points. For such instances, the LHS method provides the best technique for determining the sensitivity, for two reasons. First, the sample space which the LHS scheme uses is relatively small about the design points; and, therefore, the information obtained will be most accurate near this point. Secondly, the substantially lower cost of performing a LHS analysis makes it very attractive to use whenever possible. The probabilistic analysis of a system or subsystem should be performed with a Monte Carlo analysis if it is possible from a cost standpoint.

If it is not economically possible to use a Monte Carlo method then, since one is interested in determining the PDF of the response, the RASCAL method provides the best alternative of those examined in this study. In using the RASCAL method, one must insure that either enough discrete points are being used to adequately describe the range of the PDF in which there is interest or the intervals must be chosen so that the input PDFs are weighted to cover the region of interest. The RASCAL method is the only option to Monte Carlo in shifting probability weights so that different regions of the probability density can be examined more closely than others.

SUMMARY

The inclusion of stochastic variation or uncertainty in parameters or models is best performed with Monte Carlo simulation methods, whether the analysis is probabilistic, uncertainty, and/or sensitivity. If the Monte Carlo is too costly then the preferred methods are discrete probability distribution (RASCAL) methods for probabilistic and uncertainty analysis, and LHS for sensitivity analysis. It was found during this study that the mean corrosion depth could be predicted within 2% of the Monte Carlo results throughout a 1,000 year calculation, using RASCAL analysis at one-half the computer cost. The LHS mean calculation came within 5% of the Monte Carlo prediction at one-fifth the cost. The prediction of the standard deviation, one measure of the uncertainty, showed an even larger discrepancy with the RASCAL method predicting it within 5% while the LHS prediction only came within 30%. Finally, the RASCAL method provided results within the confidence bands of the Monte Carlo method in predicting the 10th and 90th percentile values of the corrosion distribution at 1,000 years while the LHS was 8-15% off in its predicted value.

ACKNOWLEDGEMENT

This work was supported by the Nuclear Regulatory Commission under contract number 04-82-015.

REFERENCES

1. Sastre, C. and Pescatore, C., "Draft Technical Position on Waste Package Reliability," NUREG-0097/R, 1983.

2. Kurth, R. E., "The Development and Application of a New Probabilistic Analysis Technique for Nuclear Risk Calculations," PhD Dissertation, Ohio State University, 1985.

3. Westinghouse, AESD-TME-3113

4. Cox, D. C. and P. Baybutt, "Methods for Uncertainty Analysis: A Comparative Survey," Risk Analysis, 1, No. 4 (1981).