Uncertainty Analysis Associated with Radioactive Waste Disposal: A Discussion Paper*

Robert M. Cranwell
Sandia National Laboratories
Albuquerque, New Mexico 87185

Jon C. Helton
Arizona State University
Tempe, AZ 85281

This paper discusses the problem of incorporating and representing uncertainty in the analysis of risk associated with the geologic disposal of radioactive waste. There are many collections of related events, features and processes which might affect the long-term performance of a disposal site; for convenience, such a collection is referred to as a scenario. Uncertainty in the analysis of the risk associated with a disposal site generally arises from two major sources: (1) the inexactness with which the occurrence of various scenarios can be predicted and (2) the inexactness with which the consequences associated with individual scenarios can be predicted. The inexactness in (2) generally arises from the inability to completely characterize the physical processes associated with individual scenarios. Questions of the following type arise: How can a deterministic model be converted into a probabilistic model and the resultant probabilistic predictions be analyzed? How does one compile the various scenarios which could affect a disposal site and describe the uncertainty in the quantification of these scenarios? What is an appropriate way to calculate the consequences, and to represent the effects of uncertainty,

---

*Work supported by the U.S. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research under memorandum of understanding DOE 40-550-75 with the U.S. Department of Energy.

for a single scenario? How should individual scenarios be selected for inclusion in a comprehensive disposal site analysis? What is an appropriate way to calculate the consequences, and to represent the effects of uncertainty, for a collection of scenarios selected to represent a disposal site? The preceding questions, and possible approaches to their solution, are discussed in the context of a project at Sandia National Laboratories to develop a methodology to assess the risk associated with the geologic disposal of radioactive waste.

1. Introduction

The management and long-term disposal of high-level radioactive waste and spent fuel produced by nuclear power generation and national defense activities is becoming an increasingly important issue. The waste inventory is already large and difficult to manage, and the future use of nuclear power is being questioned at least partly on the basis of waste disposal capability. Many options have been proposed for disposal of radioactive waste, including burial in deep excavations or boreholes, emplacement in the seabed, ejection into space, and others. The current option being pursued most actively is that of burial in mined depositories in deep, geologic formations. The formations under present consideration include granite, shale, basalt, domed salt and bedded salt. The time period over which the performance of such repositories must be assessed is long -- from at least a few thousand years to perhaps a few hundred thousand years. Obviously, experiments and monitoring to gain information on system behavior cannot be carried out over such time periods. The only available way to evaluate candidate sites, waste forms and repository designs and to assess the safety of repositories is predictive modeling. An important part of such modeling efforts will be the determination and representation of uncertainty in model input and model predictions.

This paper considers the determination and representation of uncertainty in the analysis of the geologic disposal of radioactive waste. It is derived from a project at Sandia National Laboratories to develop a methodology to assess the risk associated with such disposal. The paper will present the major aspects in organizing the information which is available for a potential disposal site and in obtaining a measure, with appropriate uncertainty bounds, of the consequences associated with the site. The following organization is used. For the remainder of this section, the influence of models on the development and application of uncertainty analysis techniques in the context of radioactive waste disposal is discussed. Then, five questions which arise in uncertainty analysis for geologic waste disposal are presented. The presentation for each question includes a discussion of its importance in the context of waste disposal and possible techniques for its treatment. Specifically, the following questions are considered:

-- Section 2: How can a deterministic model be converted into a probabilistic model and the resultant probabilistic predictions be analyzed?

-- Section 3: How does one compile the various scenarios which could affect a disposal site and describe the uncertainty in the quantification of these scenarios?

-- Section 4: What is an appropriate way to calculate the consequences, and to represent the effects of uncertainty, for a single scenario?

-- Section 5: How should individual scenarios be selected for inclusion in a comprehensive disposal site analysis?

-- Section 6: What is an appropriate way to calculate the consequences, and to represent the effects of uncertainty, for a collection of scenarios selected to represent a disposal site?

The paper concludes with a brief summary section.

As already noted, the ideas presented in this paper are derived from a project at Sandia National Laboratories. This project consists of three major parts: (1) the development of models to represent physical processes associated with the disposal of waste in geologic formations, (2) the development of techniques for the assessment and use of these models, and (3) the application of these models and techniques to a hypothetical waste repository. The development of uncertainty analysis techniques belongs in the second category. Specifically, the designation "uncertainty analysis techniques" is used to mean methods by which the inexactness of our knowledge with respect to the occurrence of events and processes at a disposal site and the inexactness of our capability to describe such events and processes can be translated into probabilistic statements (e.g., expected values, variances, distributions, confidence intervals) about their consequences.

The development and application of uncertainty analysis techniques for the assessment of waste disposal sites is very closely tied to the models which represent physical processes at these sites. The obtainable information for a site generally does not provide immediate insight into the consequences associated with a repository at that site. Rather, models must be used to process the obtainable information into forms that do provide insight with respect to the repository. Thus, uncertainty analysis techniques must operate in conjunction with the models to determine the uncertainty associated with model predictions from the uncertainty associated with the data supplied to the models as input. Much of the data which can be obtained for disposal sites will be represented probabilistically while the models which have been developed to represent processes at these sites are generally deterministic. In essence, uncertainty analysis techniques are used to develop probabilistic models (i.e., random variables as input and output) from deterministic models (i.e., individual variables as input and output) and then to analyze the predictions made by the new probabilistic models.

Due to their importance in shaping the uncertainty analysis techniques presented in this paper and to help indicate the potential complexity of an uncertainty analysis, the models which were developed in the Sandia program for use in the assessment of waste disposal sites are now briefly indicated. The processes which these models represent can be divided into four major categories: (1) near-field repository behavior, (2) groundwater transport, (3) pathways-to-man and (4) dosimetry and health effects. The two major models available to represent near-field behavior are SWIFT, a three-dimensional, finite difference model which solves a coupled system of partial differential equations to represent fluid flow, heat transfer, brine migration and radionuclide transport (Di78), and DNET, a quasi two-dimensional network model which simulates waste/host rock interactions and feedback effects in the vicinity of a depository (Cr80b). Two major models are also available to represent groundwater transport. One of these is SWIFT. The other is NWFT/DVM, a network flow and transport model which solves radionuclide transport equations by a distributed velocity scheme (Ca80a, Ca80b). A two part model is available to perform pathways-to-man calculations (He80b). The first part represents radionuclide movement in the surface environment by a system of differential equations; the second part represents radionuclide movement from the surface environment to a human population by use of concentration ratios. A model is available which performs dosimetry calculations on the basis of 70-year dose factors derived from the ICRP-II model and which estimates health effects from individual doses and latent cancer risk factors proposed by the BEIR-II committee (Ru80).

This section concludes with a brief discussion of the importance of representing uncertainty in calculated consequences for geologic waste disposal. First, it is not possible to exactly predict the sets of events, features and processes (i.e., scenarios) which may affect a disposal site. Most likely, the best that can be obtained will be probabilistic statements (possibly quite crude) about the occurrence of individual scenarios. Second, it is not possible to exactly describe the conditions (i.e., input data for computer models) necessary to predict the consequences associated with specific scenarios. Again, the best that can be obtained may be probabilistic statements (possibly quite crude) about the variables which define model input. To ignore the effects of such uncertainties on model predictions could produce very misleading results.

2. How can a deterministic model be converted into a probabilistic model* and the resultant probabilistic predictions be analyzed?

This is a very general question but one which is at the center of uncertainty analysis for geologic waste disposal. As already indicated, uncertainty analysis is used to mean the application of methods by which the inexactness of our knowledge with respect to the occurrence of events and processes at a disposal site and the inexactness of our capability to describe such events and processes can be translated into probabilistic statements about their consequences. The analysis of a disposal site involves the use of several large and complex deterministic models. However, the input values for these models are actually the realizations of many different random variables. Thus, uncertainty analysis for a particular consequence can be interpreted as the study of a random variable which describes the behavior

---

*The designation "deterministic model" is used to indicate a function whose input is a sequence of real numbers and whose output is one or more real numbers. The designation "probabilistic model" is used to indicate a function whose input is a sequence of random variables and whose output is one or more random variables.

of this consequence; in turn, the character of this random variable results from the deterministic models involved and the random variables which describe input values for these models. This leads to the following question: How can a "more complex" model be constructed from the deterministic models which takes random variables and their associated joint distributions as input and generates one or more random variables as output? Then, uncertainty analysis becomes the study of the dependent random variables.

In this section, a technique for converting a deterministic model into a probabilistic model is discussed. Further, analysis techniques for the resultant dependent random variables are also indicated. Then, application of these techniques to a single scenario is considered in Section 4, and application of these techniques to an overall assessment of a disposal site is considered in Section 6.

In practice, it probably will not be possible to develop a model which takes actual random variables (i.e., functions defined on probability spaces) as input and in turn actually generates such functions as output. Indeed, the random variables under consideration will often be estimated by empirical distributions or other approximations. Thus, it is proposed that a deterministic model be converted into a probabilistic model by using some scheme to generate a sequence of values (actually, a sequence of vectors, where each vector contains one value from each random variable) sampled from the random variables according to their distributions (or, the best approximations to them which are available). Then, the individual variable values associated with each of these vectors can be supplied to the deterministic model as input. This generates a sequence of model outputs which, in conjunction with the probabilities associated with the individual sets of input for the deterministic model, can be used to generate a random variable (actually, an approximation to a random variable) which is the result of the applying the new "probabilistic" model to the original random variables. Then, suitable analysis of this new random variable and the values used in its generation provide the desired uncertainty analysis.

The discussion of the preceding paragraph is now elaborated on. In geologic waste disposal and many other problems, one initially has a deterministic model D which is a function of the variables v_1, ..., v_n. However, it is difficult to justify specific choices for the v_i; in reality, the values of these variables are realizations of associated random variables V_1, ..., V_n. Thus, what is desired is an assessment of the behavior of $D(v_1, ..., v_n)$ which reflects the behavior of V_1, ..., V_n. In essence, it is desired to replace the deterministic model D with a probabilistic model P which has the random variables V_1, ..., V_n as input and has a random variable as output. Then, the dependent random variable can be analyzed to gain insight with respect to the behavior of predictions made by the original deterministic model D. However, for most situations it is not possible to actually construct the probabilistic model P; instead, the best that can be done is to construct a model P_A which takes approximations to the random variables $V_1, \ldots, V_n$ as input and in turn generates an approximation to a random variable as output. Then, this approximation can be analyzed to gain information with respect to the uncertainty associated with the predictions made by the original deterministic model D.

The preceding discussion leads to two questions: What is a suitable way to sample from the random variables which are to be used as input for the new probabilistic model? What is a suitable way to represent and analyze the random variable which is generated as output by this model? Before possible solutions are proposed, several desirable properties for the methods are listed:

-- Provide for estimates of expected value, variance and distribution for the dependent random variable,

-- Provide for estimates of confidence intervals for the preceding entities,

-- Permit investigation of different distributions and correlations for the random variables used as input,

-- Permit determination of variables which are important in producing uncertainty in model predictions,

-- Be numerically efficient with respect to the amount of calculation required to obtain the information indicated in the preceding desiderata.

Due to the potentially large amount of computation required by some of the models used to study radioactive waste disposal, the last consideration is very important.

There exist many sampling techniques which might be used to sample from random variables to generate probabilistic input data for deterministic models. Several of these methods are compared in McKay, Conover and Beckman (Mc79). Of the possible sampling methods, the one used in the Sandia project is Latin hypercube sampling. This technique to select n different values from each of k variables $X_1, \ldots, X_k$ operates in the following manner. The range of each variable is divided into n nonoverlapping intervals on the basis of equal width or equal probability. One value from each interval is selected at random (for equal probability, random sampling means with respect to the probability density in the interval). The n values thus obtained for X_1 are paired in a random manner (equally likely combinations) with the n values of X_2. These n pairs are combined in a random manner with the n values of X_3 to form n triplets, and so on, until n k-tuples are formed. This is the Latin hypercube sample. A computer program for generating Latin hypercube samples [illegible] been developed and documented by Iman, Davenport and Ze[illegible] Im80c).

To initial[illegible]esent the random variable which is the output from the probabilistic model, a cumulative distribution function can be used. Then, the questions of uncertainty analysis become questions of describing this random variable. Important aspects of such an analysis are now indicated. Specifically, with the assumption that Latin hypercube sampling has been used to convert a deterministic model into a probabilistic model, the following statements can be made:

-- Unbiased estimates of the mean, other moments and the distribution function are possible (Im80b, Theorem 1, p. 11).

-- Variances of the preceding estimators are usually smaller than variances of estimators arising from other sampling techniques; however, this result may be closely related to monotonicity properties of the model. The sample variance provides a biased estimate of the population variance; however, the bias is often small (Im80b, Sections 2.5 and 2.6).

- Sensitivity analysis techniques based on partial correlation and stepwise regression can be used to determine the dominant variables with respect to the dependent random variabl (Im78).

-- The effects of different distribution assumptions for the input random variables on the dependent random variable can be investigated without rerunning the model (Im80b, Chapter 3).

-- A variation of Latin hypercube sampling known as replicated Latin hypercube sampling can be used to obtain confidence intervals for estimators with respect to the dependent random variable (personal communication from R. L. Iman with respect to ongoing work).

-- Determination of the preceding information is efficient in that it can be accomplished with less calculation (i.e., the generation of fewer sample values for the dependent random variable) than with other sampling techniques which have been considered (Mc79; Im80a, Chapter 4).

It is felt that all of the preceding considerations fall under the general heading of uncertainty analysis.

3. How does one compile the various scenarios which could affect *a disposal site and describe the uncertainty in the quantification* of these scenarios?

The word scenario is used to denote a collection of related events, features and processes which might affect a disposal site. Uncertainty in the analysis of a waste disposal site arises from two major sources: (1) the inexactness with which the occurrence of various scenarios can be predicted and (2) the inexactness with which the consequences associated with individual scenarios can be predicted. The inexactness in (2) generally arises from the inabilty to completely characterize the physical processes associated with individual scenarios. The problem of incorporating uncertainty when converting from a deterministic model to a probabilistic model is considered in Section 2. This is the central problem in uncertainty analysis for geologic waste disposal. However, the first problem which must be dealt with in an uncertainty analysis of a site is how to organize the scenarios referred to in (1) and the computationally related descriptions of these scenarios referred to in (2) into a form which permits application of the uncertainty analysis techniques indicated in Section 2. This organizational technique must operate to group events, features and processes into scenarios in a manner that facilitates use of the models which are available to predict consequences. In this regard, it is emphasized that it is usually not possible to immediately ascertain the consequences associated with individual scenarios. Instead, it is necessary to describe these scenarios in sufficient detail and then to use models to predict their consequences.

The organizational technique that is being used in the Sandia project is now briefly indicated. The events, features and processes that are important with respect to the behavior of a repository are organized into four categories: (1) those which could influence release of radionuclides from the depository to a nearby aquifer system, (2) those which could influence release of radionuclides from the depository directly to the biosphere, (3) those which could influence radionuclide movement in groundwater to some surface discharge location, and (4) those which could influence radionuclide movement in the surface environment and resultant human exposure. Each of the preceding categories has a number of sets of conditions (i.e., subscenarios) associated with it. Appropriate unions of these sets are referred to as *scenarios* and are the basic *organizational units* for the analysis of a disposal site. This technique is intended to operate in conjunction with the physical models indicated in Section 1. Additional discussion of scenario development, definition and application is provided in Cranwell, et al (Cr80c).

Two observations are now made with respect to the scenario development technique just indicated. First, it is felt that this organizational method is preferable to event-tree, fault-tree techniques for studying uncertainty in the context of radioactive waste disposal. This statement is made for the following reasons: (1) many of the so-called "events" in waste disposal do not represent immediate or abrupt changes in the repository system but rather slow continuous changes over 100's to 1000's of years (e.g., salt dissolution, shaft or borehole seal degradation, formation of a geologic dike), and hence their occurrence cannot be represented by a simple "yes" or "no" statement, (2) the events and processes of concern in waste disposal do not necessarily have to occur in a particular sequence and hence organizational problems develop when one attempts to force an ordering onto them, and (3) the existence of feedback loops may be important (e.g., for disposal in bedded salt, thermal expansion and resultant cracking can lead to salt dissolution which can lead to subsidence and collapse which can lead to additional dissolution which can lead to further subsidence and collapse). The preceding observations lead back to a point which has already been stated: for a repository analysis, it is necessary to organize conditions and then use predictive modeling to study the consequences associated with these conditions. The second observation is that it is not possible to prove "completeness" in the sense of unequivocally establishing that all possible scenarios have been compiled. Through care in scenario development and appropriate independent review, assurance can be sought that a collection of scenarios is acceptably complete. However, this cannot be proved.

The probabilistic description of each scenario involves two types of random variables: a random variable which describes the probability of occurrence for the scenario, and a collection of random variables which describes the inexactness with which the information needed to model the scenario is known. The first type of random variable is needed (1) to calculate consequences and uncertainty across all scenarios, (2) to perform risk calculations, and (3) to screen scenarios on the basis of probability. The second type of random variable is needed (1) to calculate consequences and uncertainty for individual scenarios (2) to perform risk calculations, and (3) to screen scenarios on the basis of consequence.

Determination of the random variables needed to describe the scenarios associated with a disposal site is dependent on both the individual scenarios and the particular site. It is difficult to give specific techniques for their determination in a general paper such as this; indeed, the thrust of this paper is, given that these random variables can be determined, how can the uncertainty which they impose on assessments of a site be studied? However to provide a feeling for what is involved in the determination of these random variables, the following six general approaches are discussed briefly: (1) application of known physical relationships, (2) laboratory measurements of properties and processes, (3) field measurements of geologic conditions and processes, (4) investigation and interpretation of past historic and geologic records, (5) synthesis of expert opinion, and (6) deliberate conservatism. All of these techniques are most useful when their application is as site-specific as possible.

The first technique is based on estimating probabilities through use of accepted scientific principles. For example, heat flow calculations can be used to determine temperature distributions in the vicinity of a depository. This information can be used to estimate the probability of thermally induced fracturing in the rock strata surrounding the depository. In turn, the preceding probabilities can be used to estimate the probabilty of water flow through the depository. As another example, chemical and physical characteristics of an area might be used to estimate ranges and distributions for the partitioning of radionuclides between the liquid and solid phases of a system (i.e., distribution coefficients).

The second technique is based on laboratory measurements. Many important variables can be investigated through properly designed laboratory experiments. Such variables include ones which relate to distribution coefficients, leach rates, shaft-seal failure rates, concentration ratios and development of fracture systems.

The third technique involves direct field measurements of geologic conditions and processes to determine the probabilistic nature of important varaibles. For example, field observation can be used to determine the distribution of erosion rates with respect to climate and topography. As other examples, direct observation can yield information on processes such as uplift and subsidence, dissolution, and growth of fracture systems and on conditions as characterized by porosities, conductivities and distribution coefficients.

The fourth technique is similar to the third technique in that it uses data from past human and geologic activity. Yet it is different in that it infers probabilities from past records rather than from field measurements designed to determine them directly. For example, past rates and distributions for tectonic and volcanic activity might be used to infer such rates for the future. However, as both techniques are based on extrapolation of past and present conditions into the future, care must be taken to assure that the time periods used in these projections are appropriate. Use of time periods that are too short can produce misleading results. For example, the historical records on which to base estimates of inadvertent intrusion by drilling only extend back to 50 to 100 years -- a short time period in comparison to the few thousand to possibly a few hundred thousand years involved in a repository assessment. Similarly, the use of time periods that are too long can also produce misleading results. For example, consideration of average occurrence probabilities over very long time periods can obscure significant fluctuations in geologic behavior which take place over shorter time periods.

The fifth technique is the use of expert opinion to quantify uncertainties and estimate probabilites when experimental and observational data are scarce or imprecise. Expert opinion usually involves many of the other techniques in the list. A group of experts tries to put together the inadequate information gained through other techniques. If such a group can reach a consensus and if other groups agree with their conclusions, then this may provide the best estimates of probabilities and uncertainties that can be obtained. Expert opinion can be developed in several ways: the Delphi method is a formally structured procedure, but informal debate and argument are a more common method. All methods involve examination of data, expression of opinion, and interaction among expert participants.

The sixth technique is deliberate conservatism. While conservatism is not a technique for meaningfully quantifying uncertainty, it does provide a starting place in an analysis for a variable on which there is limited information. Specifically, an analysis can be carried out with probabilities and ranges which are conservative in the sense that they tend to cause overestimation of the variable's significance. Thus, if an analysis errs, it is by overestimating rather than by underestiamting the consequences associated with a disposal site. Further, if such a variable emerges as being important, then additional work is required to determine a more realistic distribution for its value. Unfortunately, the ranges and distributions for many variables in a waste disposal site analysis will involve estimates. In developing such estimates, it is important to avoid excessive use of conservatism, as this can produce very misleading uncertainty analysis results.

Although application of the preceding techniques is not discussed in this paper, various of them have been applied in the Sandia waste isolation project. Such applications can be found in the following papers and reports: Beckman and Johnson (Be80), Cranwell (Cr80a), Cranwell and Donath (Cr80d, Cr80e, Cr80f), Donath, Schwartz and Cranwell (Do80), and Helton and Iman (He80b).

As is evident, many of the random variables needed to describe a waste disposal site will be ill-defined. For many, the best obtainable descriptions may be empirical distribution functions based on inadequate laboratory or field observations. For others, expert opinion, and possibly several divergent expert opinions, may be all that exists.

4. What is an appropriate way to calculate the consequences, and to represent the effects of uncertainty, for a single scenario?

As already discussed, an individual scenario consists of a collection of related events, features and processes potentially affecting radionuclide movement away from a depository and eventual human exposure to these radionuclides. Further, associated with each scenario are random varibles which represent the available knowledge with respect to the input variables required by the models in use to predict the consequences associated with the scenario. As several models may be required and several decay chains may be involved, the potential number of random variables is large (i.e., 10's to 100's). For this discussion, it is assumed that a scenario has been described on a gross scale as indicated in Section 3. Further, it is also assumed that a suitable grouping of deterministic models has been formed to calculate the consequences associated with the scenario and that random variables which describe the uncertainty in the input data for these models have been defined.

The question on which this section is based is a special case of the question considered in Section 2. How can a deterministic model be converted into a probabilistic model and the resultant probabilistic predictions be analyzed? When the individual deterministic models required to represent the individual parts of the scenario are combined, the result is a single "large" deterministic model D and a collection of random variables $V_1, \ldots, V_n$ which represent the uncertainty in the input variables $v_1, \ldots, v_n$ for D. Then, it is desired to represent the uncertainty in the predictions made by D which results from the uncertainty described by the random variables $V_1, \ldots, V_n$. As discussed in Section 2, this can be accomplished by constructing a probabilistic model P from the deterministic model D. The new probabilistic model P will use the random variables $V_1, \ldots, V_n$ as input and will generate a random variable $P(V_1, \ldots, V_n)$ as output which can be analyzed to determine the uncertainty in the predictions made by D. In reality, a model P_A will be constructed which takes approximations to $V_1, \ldots, V_n$ as input and generates an approximation to $P(V_1, \ldots, V_n)$ as output.

Before considering the information which might be sought in an uncertainty analysis for an individual scenario, the nature of D, P and P_A is indicated. It is conceptually convenient to think of D, P and P_A as individual models, which they are. However, in actual applications, they will not generally be individual models in the sense of a computer program that can be executed as a single entity. Instead, D will consist of a number of individual programs which are executed separately. The probabilistic model P would never be defined. The approximation P_A to P would be a program for sampling from the random varibles $V_1, \ldots, V_n$ which operated in conjunction with the individual programs which constituted D. Specifically, the sampling program would initially generate a sequence of samples $(v_{i1}, \ldots, v_{in})$, i = 1, ..., m, from the random variables $V_1, \ldots, V_n$. These values would be stored on file. Then, the values in each n-tuple would be used in conjunction with the individual models in D to calculate a consequence. In practice, this would involve a sequence of calculations and file generations. Generally, D would consist of submodels $D_1, \ldots, D_k$ which operate in sequence in the sense that D_i generates input for D_{i+1}. The model D_1 would be managed by a program P_1 which read the samples from the random variables, converted them into input for D_1, supplied this input to D_1, and recorded the results generated by D_1. Similarly, D_2 would be managed by a program P_2 which read the samples from the random variables and the results generated by D_1, converted this information into input for D_2, supplied this input to D_2, and recorded the results generated by D_2. This process would continue until a file containing the actual information of interest (i.e., an approximation of the random variable $P(V_1, \ldots, V_n)$) was generated. Thus, the probabilistic model P_A would operate through the sampling program, the data handling programs $P_1, \ldots, P_k$, and the original deterministic models $D_1, \ldots, D_k$.

Information which can be sought about the consequences, and associated uncertainty, for an individual scenario is now indicated. The random variables which define input for the deterministic models can be sampled by Latin hypercube sampling. In turn, this sample can be used in generating an approximation to a random variable which incorporates the uncertainty in consequence predictions for the scenario. Then, information on this random variable can be sought as discussed in Section 2. Unbiased estimates of the mean, other moments and the distribution function can be calculated. The preceding estimates generally have small variances relative to estimates provided by other sampling methods; the sample variances can be used to provide biased estimates of these variances. Sensitivity analysis techniques can be used to determine the dominant variables in affecting the behavior of the selected consequence. The effects of different assumptions with respect to the original random variables can be investigated without rerunning the models. If replicated Latin hypercube sampling is used, confidence intervals can be estimated. It is felt that investigations of the preceding nature fall under the general heading of uncertainty analysis.

This section has considered the analysis of consequences associated with an individual scenario. Results obtained in such an analysis are conditional; they are based on the assumption that the scenario occurs. Section 6 considers the analysis of consequences for a disposal site when the probabilities of occurrence for individual scenarios are included.

5. How should individual scenarios be selected for inclusion in a comprehensive disposal site analysis?

When a potential disposal site is analyzed, the number of possible scenarios may be too large to permit their complete inclusion in a consequence/uncertainty analysis. Thus, the problem arises of how to select the scenarios which will be used in the analysis. Actually, this is a two part problem: (1) How should the scenarios be selected for consideration? and (2) How should the variables which are used to describe the scenarios be selected? The preceding are very important questions; they are also site and model specific. Therefore, the following discussion of them is general and brief. However, in the analysis of a particular site, resolution of these questions will probably involve significant effort.

The scenario selection problem is considered first. As already noted, the scenario generation technique will probably generate more scenarios than can be incorporated into the final analysis of a site. Indeed, the first effort at scenario development will probably be to generate as comprehensive a collection of scenarios as possible. Then, a suitable subcollection of these scenarios must be selected for use in a comprehensive site analysis. With the assumption that the scenario development process disallows physically unreasonable scenarios, there are two criteria left which can be used to screen scenarios for inclusion in the final site analysis: consequence and probability. Possible scenarios with very low consequences can be omitted because of their small potential to affect risk and to cause uncertainty in the analysis of risk. Similarly, possible scenarios with very low probabilities can also be omitted. It is also possible that scenarios with "intermediate" consequences and probabilities may be screened on the basis of risk. Due to the large computational effort required to perform a site analysis as indicated in Section 6, it is important to reduce the number of scenarios as much as possible in a manner which is consistent with an overall goal of a meaningful consequence/uncertainty analysis. An additional technique that may be useful is to seek out scenarios which are "similar" and to find ways to pool such collections into single scenarios.

The second question is now considered. That is, how should the variables which are used to describe the scenarios be selected? There is generally some knowledge of the variables to be used; if this were not the case, there would not be models for physical processes associated with a site and there would not be a scenario development technique based on organizing events and processes in a manner suitable for use with these models. Actually, the question can be restated as follows: For a model which is capable of using many variables to describe a process, how can the variables be identified which, due to the uncertainty in their values, dominate the uncertainty in predictions made by the model? Some variables may be very uncertain in their values; however, if they have little effect on model predictions, then this uncertainty is of limited concern. However, it is important to recognize and include variables which do have significant effects on the uncertainty of model predictions. Identification of important variables can be accomplished with sensitivity analysis techniques designed to assess the effects that individual variables and their associated distributions have on model predictions.

In the Sandia project, an approach to sensitivity analysis based on partial correlation coefficients and stepwise regression has been found to be productive. The basic idea is to (1) select a group of potentially important variables which define input for a model, (2) choose ranges and distributions for the variables, (3) sample from the variables according to their assigned ranges and distributions, (4) generate input values for the model from the sampled values of the [illegible] with the generated input, and (6) assess the relationships between the original variables and model output by partial correlation coefficients and stepwise regression. Latin hypercube sampling has been successfully used as the sampling technique; also, the rank transform has been helpful in reducing the effects of nonlinearity in model predictions. This approach to sensitivity analysis is discussed in Iman, Helton and Campbell (Im78), and examples of its application are contained in Campbell, Iman and Reeves (Ca80a), Helton, Brown and Iman (He80a), and Helton and Iman (He80b).

This section ends with the observation that the proper selection of variables for both modeling and field investigation will be one of the recurrent problems in the analysis of a disposal site. The appropriate use of sensitivity analysis techniques provides a systematic way to investigate and then select such variables. A very important aspect of sensitivity analysis is that it can be used to identify the variables which cause the most uncertainty in model predictions and hence to identify the areas where additional study will do the most to reduce uncertainty in the analysis of a site.

6. What is an appropriate way to calculate the consequences, and to represent the effects of uncertainty, for a collection of scenarios selected to represent a disposal site?

For the following, it is assumed that previously discussed techniques have been used to select a collection of scenarios, to assign probability distributions for the occurrence of these scenarios, to select a set of variables needed to predict the consequences associated with the scenarios, and to define probability distributions for these variables. Then, the question on which this section is based is a special case of the question considered in Section 2: How can a deterministic model be converted into a probabilistic model and the resultant probabilistic predictions be analyzed?

For use in later notation, it is assumed that m scenarios and n variables are under consideration. Further, it is assumed that U_i, i = 1, ..., m, is a sequence of random variables such that U_i represents the probability of occurrence for the i^{th} scenario and that V_i, i = 1, ..., n, is a sequence of random variables such that V_i represents the probability of occurrence for the i^{th} variable required to model the site. In reality, the models will require many more variables than just those associated with the V_i; however, the V_i correspond to those variables which are sufficiently important in affecting the uncertainty for a site to be considered as random variables rather than as fixed input values. When the individual deterministic models required to represent the various parts of the i^{th} scenario are combined, a single deterministic model M_i is produced; this model is a function of the form $M_i(v_1, \ldots, v_n)$, where the v_i are individual realizations of the random variables V_i. Some of the V_i may be required in the modeling of all scenarios; others may be required for only a few scenarios. However, it is organizationally convenient to assume that all scenarios involve all the V_i. Next, a model D can be defined which, for specified occurrence probabilities for each scenario, gives an expected consequence for the disposal site. Specifically, if u_i is a realization of U_i for i = 1, ..., m and v_i is a realization of V_i for i=1, ..., n, then D is the model given by

$$D(u_1, \ldots, u_m, v_1, \ldots, v_n) = \sum_{i=1}^{m} u_i \, M_i(v_1, \ldots, v_n).$$

The problem of uncertainty analysis for the disposal site now becomes how to determine and represent the uncertainty in predictions made by D which results from the characteristics of $U_1, \ldots, U_m, V_1, \ldots, V_n$. As already noted, this is precisely the problem discussed in Section 2. Specifically, a probabilistic model P is constructed from the deterministic model D; the new probabilistic model uses the random variabes $U_1, \ldots, U_m, V_1, \ldots, V_n$ as input and generates a random variable $P(U_1, \ldots, U_m, V_1, \ldots, V_n)$ as output. This new random variable can then be analyzed to determine the uncertainty in predictions made by D. The general nature of the construction process for P in the context of waste disposal is indicated in Section 4. However, that description is for the analysis of a single scenario; one additional level of complexity exists in a site analysis due to the consideration of many scenarios.

Information which can be sought about the consequences, and associated uncertainty, for a collection of scenarios selected to represent a disposal site is the same as indicated at the ends of Sections 2 and 4.

7. Summary

The problem of incorporating and representing uncertainty in the analysis of geologic waste disposal has been discussed. The approach has been to view uncertainty analysis in the context of the problem of how to convert from a deterministic model (i.e., a function whose input is a sequence of real numbers) to a probabilistic model (i.e., a function whose input is a sequence of random variables and whose output is one or more random variables). Then, uncertainty analysis becomes the study of how the properties of the output random variable are determined by the properties of the input random variables. In the context of this approach, various questions which relate to uncertainty analysis for geologic waste disposal have been discussed and the manner in which the problems associated with these questions are being treated in the Sandia project has been indicated.

Acknowledgements

The authors wish to thank the many individuals who have contributed to the Sandia National Laboratories Waste Isolation Project from which this paper is derived. In particular, R. L. Iman and W. J. Conover are recognized for their contributions related to sensitivity and uncertainty analysis.

References

Be80 R. J. Beckman and M. E. Johnson, 1980, A Probabilistic Model for Estimating the Effects of Volcanism on a Nuclear Waste Repository, Report No. LA-UR80-2166, Los Alamos Scientific Laboratory, Los Alamos, NM.

Ca78 J. E. Campbell, R. T. Dillon, M. S. Tierney, H. T. Davis, P. E. McGrath, F. J. Pearson, H. R. Shaw, J. C. Helton and F. A. Donath, 1978, Risk Methodology for Geologic Disposal of Radioactive Waste: Interim Report, Report No. SAND78-0029, Sandia National Laboratories, Albuquerque, NM.

Ca80a J. E. Campbell, R. L. Iman and M. Reeves, 1980, Risk Methodology for Geologic Disposal of Radioactive Waste: Transport Model Sensitivity Analysis, Report No. SAND80-0644, Sandia National Laboratories, Albuquerque, NM.

Ca80b J. E. Campbell, P. C. Kaestner, B. S. Langkopf and R. B. Lantz, 1980, Risk Methodology for Geologic Disposal of Radioactive Waste: The Network Flow and Transport (NWFT) Model, Report No. SAND79-1920, Sandia National Laboratories, Albuquerque, NM.

Ca80c J. E. Campbell, D. E. Longsine and M. Reeves, 1980, Risk Methodology for Geologic Disposal of Radioactive Waste: The Distributed Velocity Method of Solving the Convective-Dispersion Equation, Report No. SAND80-0717, Sandia National Laboratories, Albuquerque, NM.

Cr80a R. M. Cranwell, 1980, Risk Methodology for Geologic Disposal of Radioactive Waste: Probabilistic Study of a Waste Repository Being Penetrated by Drilling Activity, to appear as a Sandia National Laboratory technical report.

Cr80b R. M. Cranwell and J. E. Campbell, 1980, Risk Methodology for Geologic Disposal of Radioactive Waste: A Model for Incorporating Feedback Effects in Salt Dissolution Processes, Report No. SAND80-0067, Sandia National Laboratories, Albuquerque, NM.

Cr80c R. M. Cranwell, J. E. Campbell, F. A. Donath, J. C. Helton and N. C. Finley, 1980, Risk Methodology for Geologic Disposal of Radioactive Waste: Scenario Development, Report No. SAND 80-1429, Sandia National Laboratories, Albuquerque, NM.

Cr80d R. M. Cranwell and F. A. Donath, 1980, Application of Geometric Probability to the Existence of Faults in Anisotropic Media, to appear in Proceedings of the International Symposium on the Scientific Basis for Nuclear Waste Management.

Cr80e R. M. Cranwell and F. A. Donath, 1980, Probabilistic Treatment of Faulting in Geologic Media, J. Amer. Geophys. Un., Special Handin Volume.

Cr80f R. M. Cranwell and F. A. Donath, 1980, Risk Methodology for Geologic Disposal of Radioactive Waste: Probability of Faulting in a Repository System, Report No. SAND80-0164, Sandia National Laboratories, Albuquerque, NM.

Di78 R. T. Dillon, R. B. Lantz and S. B. Pahwa, 1978, Risk Methodology for Geologic Disposal of Radioactive Waste: The Sandia Waste Isolation Flow and Transport (SWIFT) Model, Report No. SAND78-1267, Sandia National Laboratories, Albuquerque, NM.

Do80 F. A. Donath, F. W. Schwartz and R. M. Cranwell, 1980, Probabilistic Aspects of Faulting and Its Possible Influence on the Isolation of Nuclear Waste, 26th International Geologic Congress, Paris, France.

He80a J. C. Helton, J. B. Brown and R. L. Iman, 1980, Risk Methodology for Geologic Disposal of Radioactive Waste: Asymptotic Properties of the Environmental Transport Model, Report No. SAND79-1908, Sandia National Laboratories, Albuquerque, NM.

He80b J. C. Helton and R. L. Iman, 1980, Risk Methodology for Geologic Disposal of Radioactive Waste: Sensitivity Analysis of the Environmental Transport Model, Report No. SAND79-1393, Sandia National Laboratoreis, Albuquerque, NM.

He80c J. C. Helton and P. C. Kaestner, 1980, Risk Methodology for Geologic Disposal of Radioactive Waste: Model Description and User Manual for Pathways Model, Report No. SAND78-1711, Sandia National Laboratories, Albuquerque, NM.

Im78 R. L. Iman, J. C. Helton and J. E. Campbell, 1978, Risk Methodology for Geologic Disposal of Radioactive Waste: Sensitivity Analysis Techniques, Report No. SAND78-0912, Sandia National Laboratories, Albuquerque, NM.

Im80a R. L. Iman and W. J. Conover, 1980, Risk Methodology for Geologic Disposal of Radioactive Waste: A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables for Simulation Studies, Report No. SAND80-0157, Sandia National Laboratories, Albuquerque, NM.

Im80b R. L. Iman, W. J. Conover and J. E. Campbell, 1980, Risk Methodology for Geologic Disposal of Radioactive Waste: Small Sample Sensitivity Analysis Techniques for Computer Models, with an Application to Risk Assessment, Report No. 80-0020, Sandia National Laboratories, Albuquerque, NM.

Im80c R. L. Iman, J. M. Davenport and D. K. Zeigler, 1980, Latin Hypercube Sampling (Program User's Guide), Report No. SAND79-1473, Sandia National Laboratories, Albuquerque, NM.

Mc79 M. D. McKay, W. J. Conover and R. J. Beckman, 1979, A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, Technometrics 21, pp. 239-245.

Ru80 G. E. Runkle, R. M. Cranwell and J. D. Johnson, 1980, Risk Methodology for Geologic Disposal of Radioactive Waste: Dosimetry and Health Effects, Report No. SAND80-1372, Sandia National Laboratories, Albuquerque, NM.