12/10/91 ①

SANDIA REPORT

SAND90-2510 • UC-721

Unlimited Release

Printed December 1990

**Constructing Probability Distributions of
Uncertain Variables in Models of the
Performance of the Waste Isolation Pilot Plant:
The 1990 Performance Simulations**

Martin S. Tierney

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550
for the United States Department of Energy
under Contract DE-AC04-76DP00789

MASTER

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
Office of Scientific and Technical Information
PO Box 62
Oak Ridge, TN 37831
Prices available from (615) 576-8401, FTS 626-8401

Available to the public from
National Technical Information Service
US Department of Commerce
5285 Port Royal Rd
Springfield, VA 22161
NTIS price codes
Printed copy: A03
Microfiche copy: A01

SAND90-2510
Unlimited Release
Printed December 1990

SAND--90-2510
DE91 006755

**CONSTRUCTING PROBABILITY DISTRIBUTIONS OF UNCERTAIN VARIABLES IN
MODELS OF THE PERFORMANCE OF THE WASTE ISOLATION PILOT PLANT:
THE 1990 PERFORMANCE SIMULATIONS**

Martin S. Tierney
Safety and Reliability Analysis Division
Sandia National Laboratories
Albuquerque, New Mexico

ABSTRACT

A five-step procedure was used in the 1990 performance simulations to construct probability distributions of the uncertain variables appearing in the mathematical models used to simulate the Waste Isolation Pilot Plant's (WIPP's) performance. This procedure provides a consistent approach to the construction of probability distributions in cases where empirical data concerning a variable are sparse or absent and minimizes the amount of spurious information that is often introduced into a distribution by assumptions of nonspecialists. The procedure gives first priority to the professional judgment of subject-matter experts and emphasizes the use of site-specific empirical data for the construction of the probability distributions when such data are available. In the absence of sufficient empirical data, the procedure employs the Maximum Entropy Formalism and the subject-matter experts' subjective estimates of the parameters of the distribution to construct a distribution that can be used in a performance simulation.

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

RELATED WIPP PERFORMANCE ASSESSMENT DOCUMENTS

PUBLISHED

Bertram-Howery, S. G., and R. L. Hunter, eds. 1989. *Plans for Evaluation of the Waste Isolation Pilot Plant's Compliance with EPA Standards for Radioactive Waste Management and Disposal*. SAND88-2871. Albuquerque, NM: Sandia National Laboratories.

Bertram-Howery, S. G., and R. L. Hunter, eds. 1989. *Preliminary Plan for Disposal-System Characterization and Long-Term Performance Evaluation of the Waste Isolation Pilot Plant*. SAND89-0178. Albuquerque, NM: Sandia National Laboratories.

Bertram-Howery, S. G., and P. N. Swift. 1990. *Status Report: Potential for Long-Term Isolation by the Waste Isolation Pilot Plant Disposal System*. SAND90-0616. Albuquerque, NM: Sandia National Laboratories.

Bertram-Howery, S. G., M. G. Marietta, R. P. Rechard, P. N. Swift, D. R. Anderson, B. L. Baker, J. E. Bean Jr., W. Beyeler, K. F. Brinster, R. V. Guzowski, J. C. Helton, R. D. McCurley, D. K. Rudeen, J. D. Schreiber, and P. Vaughn. 1990. *Preliminary Comparison with 40 CFR Part 191, Subpart B for the Waste Isolation Pilot Plant, December 1990*. SAND90-2347. Albuquerque, NM: Sandia National Laboratories.

Bertram-Howery, S. G., M. G. Marietta, D. R. Anderson, K. F. Brinster, L. S. Gomez, R. V. Guzowski, and R. P. Rechard. 1989. *Draft Forecast of the Final Report for the Comparison to 40 CFR Part 191, Subpart B for the Waste Isolation Pilot Plant*. SAND88-1452. Albuquerque, NM: Sandia National Laboratories.

Guzowski, R. V. 1990. *Preliminary Identification of Scenarios That May Affect the Escape and Transport of Radionuclides From the Waste Isolation Pilot Plant, Southeastern New Mexico*. SAND89-7149. Albuquerque, NM: Sandia National Laboratories.

Hunter, R. L. 1989. *Events and Processes for Constructing Scenarios for the Release of Transuranic Waste from the Waste Isolation Pilot Plant, Southeastern New Mexico*. SAND89-2546. Albuquerque, NM: Sandia National Laboratories.

Marietta, M. G., S. G. Bertram-Howery, D. R. Anderson, K. F. Brinster, R. V. Guzowski, H. J. Iuzzolino, and R. P. Rechard. 1989. *Performance Assessment Methodology Demonstration: Methodology Development for Evaluating Compliance with EPA 40 CFR 191, Subpart B, for the Waste Isolation Pilot Plant*. SAND89-2027. Albuquerque, NM: Sandia National Laboratories.

Rechard, R. P. 1989. *Review and Discussion of Code Linkage and Data Flow in Nuclear Waste Compliance Assessments*. SAND87-2833. Albuquerque, NM: Sandia National Laboratories.

Rechard, R. P., H. J. Iuzzolino, J. S. Rath, R. D. McCurley, and D. K. Rudeen. 1989. *User's Manual for CAMCON: Compliance Assessment Methodology Controller*. SAND88-1496. Albuquerque, NM: Sandia National Laboratories.

Rechard, R. P., H. J. Iuzzolino, and J. S. Sandha. 1990. *Data Used in Preliminary Performance Assessment of the Waste Isolation Pilot Plant (1990)*. SAND89-2408. Albuquerque, NM: Sandia National Laboratories.

Rechard, R. P., W. Beyeler, R. D. McCurley, D. K. Rudeen, J. E. Bean, and J. D. Schreiber. 1990. *Parameter Sensitivity Studies of Selected Components of the Waste Isolation Pilot Plant Repository/Shaft System*. SAND89-2030. Albuquerque, NM: Sandia National Laboratories.

IN PREPARATION

Berglund, J., and M. G. Marietta. *A Computational Model for the Direct Removal of Repository Material by Drilling*. SAND90-2977. Albuquerque, NM: Sandia National Laboratories.

Brinster, K. F. *Preliminary Geohydrologic Conceptual Model of the Los Medaños Region near the Waste Isolation Pilot Plant for the Purpose of Performance Assessment*. SAND89-7147. Albuquerque, NM: Sandia National Laboratories.

Chen, Z., and H. L. Schreyer. *Formulation and Computational Aspects of Plasticity in Damage Models for Geological Materials with Emphasis on Concrete*. SAND90-7102. Albuquerque, NM: Sandia National Laboratories.

Guzowski, R. V. *Applicability of Probability Techniques to Determining the Probability of Occurrence of Potentially Disruptive Intrusive Events at the Waste Isolation Pilot Plant*. SAND90-7100. Albuquerque, NM: Sandia National Laboratories.

Helton, J. C. *Sensitivity Analysis Techniques and Results for Performance Assessment of the Waste Isolation Pilot Plant*. SAND90-7103. Albuquerque, NM: Sandia National Laboratories.

Hora, S., D. Scott, and K. Trauth. *Expert Judgment on Inadvertent Human Intrusion into the Waste Isolation Pilot Plant*. SAND90-3063. Albuquerque, NM: Sandia National Laboratories.

Marietta, M. G., D. R. Anderson, D. Scott, and P. N. Swift. *Review of Parameter Sensitivity Studies for the Waste Isolation Pilot Plant*. SAND89-2028. Albuquerque, NM: Sandia National Laboratories.

Marietta, M. G., P. N. Swift, B. L. Baker, K. F. Brinster, and P. J. Roache. *Parameter and Boundary Conditions Sensitivity Studies Related to Climate Variability and Scenario Screening for the Waste Isolation Pilot Plant*. SAND89-2029. Albuquerque, NM: Sandia National Laboratories.

Marietta, M. G., R. P. Rechard, P. N. Swift, and others. *Preliminary Probabilistic Safety Assessment of the Waste Isolation Pilot Plant*. SAND90-2718. Albuquerque, NM: Sandia National Laboratories.

Rechard, R. P., A. P. Gilkey, D. K. Rudeen, J. S. Rath, W. Beyeler, R. Blaine, H. J. Iuzzolino, and R. D. McCurley. *User's Reference Manual for CAMCON: Compliance Assessment Methodology Controller Version 3.0*. SAND90-1983. Albuquerque, NM: Sandia National Laboratories.

Rechard, R. P., P. J. Roache, R. L. Blaine, A. P. Gilkey, and D. K. Rudeen. *Quality Assurance Procedures for Computer Software Supporting Performance Assessments of the Waste Isolation Pilot Plant*. SAND90-1240. Albuquerque, NM: Sandia National Laboratories.

Roache, P. J., R. Blaine, and B. L. Baker. *SECO User's Manual*. SAND90-7096. Albuquerque, NM: Sandia National Laboratories.

Tierney, M. S. *Combining Scenarios in a Calculation of the Overall Probability Distribution of Cumulative Releases of Radioactivity from the Waste Isolation Pilot Plant, Southeastern New Mexico*. SAND90-0838. Albuquerque, NM: Sandia National Laboratories.

CONTENTS

EXECUTIVE SUMMARY	ES-1
I. INTRODUCTION	I-1
Purpose of This Report	I-2
Issues Not Addressed in This Report	I-3
II. PROCEDURES FOR CONSTRUCTING PROBABILITY DISTRIBUTIONS	II-1
An Outline of the Procedures	II-1
Empirical Cumulative Distribution Functions	II-3
Piecewise-Linear Cumulative Distribution Functions	II-5
The Maximum Entropy Formalism	II-7
An Application of the Procedures	II-11
III. LIMITATIONS ON THE 1990 PROBABILITY DISTRIBUTIONS	III-1
The Effects of Spatial Averaging	III-1
Correlations Between Model Variables	III-2
REFERENCES	R-1
GLOSSARY	G-1

FIGURES

Figure

E-1 The Five-Step Procedure Used to Construct Cumulative Distribution Functions (CDFs) for the 1990 Performance Simulations	E-1
II-1 Empirical and Piecewise-Linear CDFs for Tortuosity Data	II-5
II-2 Typical PDF Showing the Different Measures of Location	II-15
II-3 Piecewise-Linear CDF Based on Range and Median Value	II-15

TABLES

Table

II-1 Probability Distributions for Variables Sampled in Current WIPP Performance Simulations	II-12
--	-------

EXECUTIVE SUMMARY

A five-step procedure was used in the 1990 performance simulations to construct probability distributions of the uncertain variables appearing in the mathematical models used to simulate the Waste Isolation Pilot Plant's (WIPP's) performance. Figure E-1 summarizes the steps in the procedure.

This procedure provides a consistent approach to the construction of probability distributions in cases where empirical data concerning a variable are sparse or absent and minimizes the amount of spurious information that is often introduced into a distribution by assumptions of nonspecialists. The procedure gives first priority to the professional judgment of subject-matter experts and emphasizes the use of site-specific empirical data for the construction of the probability distributions when such data are available. In the absence of sufficient empirical data, the procedure employs the Maximum Entropy Formalism and the subject-matter experts' subjective estimates of the parameters of the distribution to construct a distribution that can be used in a performance simulation.

TRI-6342-634-0

Figure E-1. The Five-Step Procedure Used to Construct Cumulative Distribution Functions (CDFs) for the 1990 Performance Simulations. RI refers to responsible investigator (i.e., subject-matter expert); MEF refers to the Maximum Entropy Formalism.

I. INTRODUCTION

The Waste Isolation Pilot Plant (WIPP) is a research and development facility authorized by Congress (Public Law 96-164 [1980]) for the purpose of demonstrating the safe management, storage, and eventual disposal of those defense-generated transuranic (TRU) wastes that the U.S. Department of Energy (DOE) may designate as requiring deep geologic disposal. The DOE has established a program (hereinafter called the WIPP Project) to conduct the scientific and engineering investigations that are necessary for the demonstrations authorized by Congress. Further background on the WIPP and the WIPP Project can be found in U.S. DOE (1980) and U.S. DOE (1990).

The DOE will dispose of designated TRU wastes at the WIPP repository only after demonstrating compliance with the requirements of the U.S. Environmental Protection Agency's (EPA's) *Environmental Standards for the Management and Disposal of Spent Nuclear Fuel, High-Level and Transuranic Radioactive Wastes; Final Rule, 40 CFR Part 191*, (the Standard, EPA, 1985). The part of the Standard most relevant to this report, Subpart B or the "Environmental Standards for Disposal," sets qualitative and numerical requirements on the postclosure performance of the WIPP. (Although Subpart B of the Standard was remanded to the EPA by the United States Court of Appeals for the First Circuit, the WIPP Project will continue to respond to the Standard as first promulgated until a new Standard is in place [U.S. DOE and State of New Mexico, 1981].) In particular, the "Containment Requirements" in § 191.13 of Subpart B set numerical limits on the likelihoods that cumulative releases of radioactivity from the WIPP System to the accessible environment, for 10,000 years after closure of the system, will exceed certain prescribed levels. Demonstrating compliance with the Standard is the same as establishing a reasonable assurance that the numerical limits on the likelihoods of the prescribed levels of release specified in the Containment Requirements will not be exceeded. Further background on the Containment Requirements can be found in the Standard and in Tierney (in prep.).

In addition to specifying numerical limits, the Containment Requirements also suggest a general approach to the testing of compliance with the numerical limits on the likelihoods of cumulative releases of radioactivity from the disposal system. The EPA calls this general approach "performance assessment" and suggests that, if practicable, its end-product should be an overall probability distribution of cumulative releases of radioactivity to the accessible environment. The published guidance for interpreting and implementing the Containment Requirements suggests that the overall probability distribution should take the form of a "... 'complementary cumulative distribution function' that indicates the probability of exceeding

various levels of release" (EPA, 1985, Appendix B). In practice, estimators of such complementary cumulative distribution functions (CCDFs) are constructed by Monte Carlo simulations of the behavior of the total system during its period of performance. Background on the uses of Monte Carlo simulation in performance assessment can be found in Tierney (in prep.).

Monte Carlo simulations of the WIPP System require three things: (1) a suite of mathematical models (usually implemented on a computer) that can predict the amount of radioactivity released from the WIPP System when it is subject to the geologic, anthropogenic, and climatic conditions that could prevail during the period of performance; (2) an identification of the independent variables that appear in the mathematical models; and (3) the assignment of probability distributions to the sensitive independent variables in a manner that reflects the state of knowledge about the likelihood of the actual values these variables may have in the real system (Tierney, in prep.). Background on the models used in the WIPP simulations can be found in Lappin et al. (1989), Marietta et al. (1989), Rechard et al. (1990a) and other documents cited in these reports. Background on sensitivity studies of selected variables of WIPP-system models can be found in Rechard et al. (1990a). The present report is concerned with the procedures that were used in 1990 to provide item 3, an assignment of probability distributions to the important independent variables of the WIPP performance models.

Purpose of This Report

The WIPP Project has performed preliminary simulations of the WIPP System with the purpose of demonstrating the applicability of the methods and models it has developed for testing compliance with the Containment Requirements (Marietta et al., 1989). Rechard et al. (1990a, Appendix A) listed the approximately 240 distinct independent variables that could appear in the mathematical or computer-based models used in these simulations. Most of these variables specify the physical, chemical, or hydrologic properties of the rock formations in which the WIPP is placed; a substantial number of the variables specify physical or chemical properties of engineered materials and the waste form; some are the dimensions of engineered features of the facility, and some pertain to future climatic variability or future episodes of exploratory drilling at the WIPP. About 60 of the 240 variables are judged to warrant uncertainty analysis; preliminary ranges of variability are given for these variables in Appendix A of Rechard et al. (1990a).

Preliminary simulations of WIPP performance (Marietta et al., 1989) included up to 40 of the approximately 60 uncertain variables in the Latin hypercube sampling (LHS) scheme currently being used by the WIPP Project in its Compliance Assessment Methodology Controller (CAMCON, see Rechard et al.,

1989). Background on the assignment of probability density functions (PDFs) to these variables can be found in Appendix C of Marietta et al. (1989). No systematic procedures were used to assign PDFs to these variables: the distributions were assigned by WIPP analysts largely on the basis of limited data from Lappin et al. (1989), data from analogous (non-WIPP context) situations described in the literature and, in a few instances, on the basis of the professional judgment of subject-matter experts. Because the simulations of Marietta et al. (1989) were primarily made for demonstrational purposes, the lack of defensible and systematic procedures for the assignment of probabilities in these studies was not a serious flaw. Subsequent review of this work clarified the need for such procedures in future simulations that would be used to test compliance with the Containment Requirements.

The present, brief report describes and rationalizes the systematic procedure that was used in 1990 by the WIPP Project to construct probability distributions (cumulative distribution functions [CDFs] or probability density functions [PDFs]) for the uncertain independent variables in the WIPP performance models. The procedure is described and applied to variables currently being sampled in the WIPP performance models in Chapter II. Technical details of the procedure are also provided in Chapter II.

The 1990 procedure is described in this report to elicit reviewer's comments and start the review cycle. The WIPP Project has been asked to perform iterative performance assessments semiannually, with annual documentation of these assessments. A widely acceptable final compliance assessment depends on constructive feedback from peer reviewers of each annual assessment. This brief report is intended to focus some of the review efforts on a critical component of the performance-assessment process: construction of CDFs or PDFs.

Issues Not Addressed in This Report

Owing to limited information and time constraints, it has not been possible to address all the issues that are normally associated with the construction of probability distributions for a set of model variables. Important issues not treated or only mentioned here are

- (a) The effects of possible dependencies among the different kinds of model variables on the assignment of probability distributions to those variables;
- (b) The role of spatial correlations in constructing probability distributions for the variables of a lumped-parameter model;
- (c) The assignment of extreme-value probabilities to a variable on the basis of a limited number of observations of the variable;

- 1 (d) The assignment of numerical probabilities to parameters of natural and
2 anthropogenic phenomena that may occur in the far future.
3

4 Because of the lack of information, WIPP Project analysts have assumed that
5 all of the approximately 60 uncertain variables in their mathematical models
6 are independent (though not identically distributed) random variables. With
7 one exception (the lumped parameters specifying WIPP room hydraulic
8 conductivities and porosities), the possible effects of spatial correlations
9 on reducing the variances of the variables in certain lumped-parameter
10 performance models have been ignored. Owing to limited data, the extreme-
11 value probabilities of most of the sensitive variables cannot be estimated
12 with great confidence. Finally, the problem of assigning probabilities to the
13 parameters of processes and events that may occur at the WIPP in the far
14 future is only beginning to be addressed. The demonstrational performance
15 simulations (Marietta et al., 1989) considered scenarios for climatic change
16 and human intrusion at the WIPP in which the climatic and intrusion parameters
17 were assigned fixed values. Current performance simulations have attempted to
18 introduce uncertainty in these parameters in the simplest possible ways. For
19 the parameters of the human-intrusion scenarios, see Appendix C of Tierney (in
20 prep.).
21

22 The fact that issues (a) and (b) were not addressed in the 1990 performance
23 simulations severely limits the validity of some of the CDFs that were
24 constructed by the procedure described in this report; further discussion of
25 these issues is provided in Chapter III.
26

II. PROCEDURES FOR CONSTRUCTING PROBABILITY DISTRIBUTIONS

An Outline of the Procedures

In 1990, the WIPP Project constructed probability distributions for the uncertain variables appearing in performance models of the WIPP System by following steps 1 through 5 described below. Explanations of the meaning of underlined terms appearing in descriptions of the steps are deferred until later sections of this chapter. The acronym RI, "responsible investigator," will hereinafter mean the Sandia National Laboratory investigator who is judged to be the expert in the subject matter of the variable.

STEP 1

Determine the existence of site-specific empirical data for the variable in question; i.e., find a documented set of site-specific sample values of the variable. If empirical data sets exist, go to Step 3; if no empirical data sets are found, go to Step 2.

STEP 2

Request that the RIs supply a specific shape (e.g., normal, lognormal, etc.) and associated numerical parameters for the distribution of the variable. If the RIs assign a specific shape and numerical parameters, go to Step 5; if the RIs cannot assign a specific shape, go to Step 4.

STEP 3

Determine the size of the combined empirical data sets. If the number of values in the combined data set is >3 , use the combined data to construct an empirical cumulative distribution function or, alternatively, a piecewise-linear cumulative distribution function, and then go to Step 5. If the number of variables in the combined data set is ≤ 3 , go to Step 4.

STEP 4

Request that the RIs provide subjective estimates of (a) the range of the variable (i.e., the minimum and maximum values taken by the variable) and (b) if possible, one of the following (in decreasing order of preference): (1) percentile points for the distribution of the variable (e.g., the 25th, 50th, and 75th percentiles), (2) the mean value and standard deviation of the distribution, or (3) the mean value. Then, as justified by the Maximum

1 Entropy Formalism (MEF), construct one of the following distributions
2 depending upon the kind of subjective estimate that has been provided and go
3 to Step 5.

4
5 A uniform distribution (PDF) over the range of the variable.

6
7 A piecewise-linear CDF based on the subjective percentiles.

8
9 A truncated normal distribution based on the subjective range, mean value,
10 and standard deviation.

11
12 A truncated exponential distribution based on the subjective range and mean
13 value.

14
15 **STEP 5**

16
17 End of procedures; distribution is assigned.

18
19 This five-step procedure was motivated by a desire to maintain as close a
20 connection between situation-specific data/information and model parameters as
21 possible. Though obviously not unique, the formulation of the procedure was
22 guided by two axioms: (1) a probability distribution describing a variable
23 should, to the maximum extent practicable, be constructed from empirical data
24 and information that are site specific, and (2) if numerical data (i.e.,
25 sample values for the quantity) are few or nonexistent, probability
26 distributions for that quantity should be constructed using *only* those
27 subjective but quantified judgments about the quantity that are made by
28 experts in the subject matter pertaining to the quantity. It is assumed that
29 a subject-matter expert will take account of all relevant information, site-
30 specific or generic, in making subjective but quantified judgements about the
31 shape of a variable's distribution.

32
33 Axiom 1 recognizes that empirical, system-specific data — combined with
34 proven theoretical concepts and informed, professional interpretation of the
35 data — are the only link between the real system and the mathematical models
36 that are being used to study the real system's behavior. The need for Axiom 2
37 arises when, for various reasons, numerical data for an independent variable
38 of a model are few or entirely absent (unfortunately, this is the situation
39 for the majority of the uncertain independent variables in current WIPP
40 performance models). When data are lacking, professional judgment is all that
41 is left; Axiom 2 ensures that only subjective information provided by persons
42 with specialized knowledge of the variable (usually, persons other than the
43 performance-assessment analyst) will be included in determining the form of
44 the probability distribution. Adherence to Axiom 2 practically dictates the
45 use of a particular method called the Maximum Entropy Formalism (MEF, see

below) for constructing probability distributions from quantifiable subjective judgments.

Empirical Cumulative Distribution Functions

Suppose that one is given $N > 3$ sample values of an uncertain independent variable X that appears in a WIPP performance model,

$$X_1, X_2, X_3, \dots, X_N.$$

In the remainder of this chapter, it is assumed that the X_n s are independent, identically distributed random variables with a common (but unknown) CDF that is here denoted by $F(x)$. Furthermore, since all of the WIPP performance-model variables are positive, it will be assumed that X is a non-negative variable; i.e., $X \geq 0$. (The reader should nevertheless keep in mind the ways the assumption of independence could fail, e.g., the possibility of a biased sample arising from intervariable and spatial correlations among different kinds of variables.)

Upon ordering the sample data, one gets

$$X_1^*, X_2^*, X_3^*, \dots, X_N^*, \text{ with } X_n^* \leq X_{n+1}^*, n = 1, 2, 3, \dots, N-1.$$

If X is an intrinsically discrete variable, or if X is intrinsically continuous and some of the X_n^* s are identical (perhaps owing to the precision with which the original X_n s were measured), the ordered sample data can be grouped into $M \leq N$ ordered pairs,

$$(x_1, f_1), (x_2, f_2), (x_3, f_3), \dots, (x_M, f_M),$$

where $(x_1, x_2, \dots, x_M)$ is the ordered set of distinct values among the X_n s and the f_m s denote the multiplicities of the X_m s. For example, if X_6 appears three times in the data set, then $f_6 = 3$. Clearly, $1 \leq f_m < N$, and

$$\sum_{m=1}^M f_m = N.$$

As an example, one can take the 15 sample values of Culebra tortuosity cited in Table E-9 of Lappin et al. (1989); these values become the 12 ordered pairs: $(0.03, 1)$, $(0.04, 1)$, $(0.08, 1)$, $(0.09, 3)$, $(0.10, 1)$, $(0.12, 1)$, $(0.13, 1)$, $(0.14, 1)$, $(0.16, 1)$, $(0.21, 2)$, $(0.29, 1)$, $(0.33, 1)$.

The empirical percentiles p_m associated with the sample data are defined as the ratio of the number of values in the set $\{X_n; 1 \leq n \leq N\}$ that are less than or equal to x_m , $1 \leq m \leq M$, to the total number of values in the set ($= N$). Using this definition, it follows that

$$p_{m+1} = p_m + (1/N)f_{m+1}, \quad p_1 = f_1/N; \text{ and so}$$

$$p_m = (1/N) \sum_{i=1}^m f_i, \quad 1 \leq m \leq M.$$

The p_m s are a nondecreasing sequence of numbers ≤ 1 with $p_M = 1$.

The empirical cumulative distribution function (empirical CDF) associated with the sample data $X_1, X_2, \dots, X_N$ is the piecewise constant function here denoted by $F_c(\xi)$ and defined for $\xi \in [0, \infty)$ by

$$F_c(\xi) = \begin{cases} 0 & \text{if } \xi \leq x_1, \\ p_m & \text{if } x_m < \xi \leq x_{m+1}, \quad m = 1, 2, \dots, M-1, \\ 1 & \text{if } \xi > x_M. \end{cases}$$

The empirical CDF associated with the 15 sample values of tortuosity from Table E-9 of Lappin et al. (1989) is drawn as the dotted curve on Figure II-1.

The empirical CDF $F_c(\xi)$ is an unbiased estimator (see Blom, 1989, p. 194) of the unknown distribution of the variable X (Blom, 1989, p. 216).

The mean value or expected value of the variable X with respect to the empirical CDF $F_c(\xi)$ is here denoted by $\langle X \rangle_c$ and is the same as the usual sample mean, that is,

$$\langle X \rangle_c = (1/N) \sum_{m=1}^M f_m x_m;$$

hence $\langle X \rangle_c$ is an unbiased estimator of the expected value of the unknown distribution $F(x)$. The expected value associated with the empirical CDF for tortuosity in Figure II-1 is 0.14.

The variance of the variable X with respect to the empirical CDF $F_c(\xi)$ is here denoted by s_c^2 and can be computed as follows:

$$s_c^2 = (1/N) \sum_{m=1}^M f_m [x_m - \langle X \rangle_c]^2.$$

This is not an unbiased estimator of the variance of X , but the quantity $[N/(N-1)]s_c^2$ (the usual sample variance) is an unbiased estimator. The s_c^2 associated with the empirical CDF for the tortuosity data in Figure II-1 is 6.9×10^{-4} (hence the standard deviation $s_c \approx 0.083$).

TRI-6342-579-0

Figure II-1. Empirical and Piecewise-Linear CDFs for Tortuosity Data. Dotted line is empirical CDF; solid line is Piecewise-Linear CDF.

Piecewise-Linear Cumulative Distribution Functions

Use of an empirical CDF in practical Monte Carlo calculations may have some drawbacks. All of the sampling techniques used in Monte Carlo simulation (e.g., random sampling, LHS) require the drawing of a number of random variates from each of the distribution functions for uncertain model variables. Inspection of the example empirical CDF shown in Figure II-1 reveals that drawing random variates from an empirical CDF will only give back the discrete data points $x_1, x_2, \dots, x_M$ with respective frequencies $f_1/N, f_2/N, \dots, f_M/N$ as $N \rightarrow \infty$. Of course, this is the intended result when the variable X is an intrinsically discrete random variable (e.g., $X_n = n$ could be the number of times an event occurs in a fixed period of time). But if the variable X is an intrinsically continuous variable (e.g., the spatial average of tortuosity or porosity) and the points of the empirical data set $\{X_n, 1 \leq n \leq N\}$ are few and sparsely placed on the real line, it is possible that the sampled variates used in the simulations will always "miss" one or more of those critical values of X at which the output of the performance model could be particularly sensitive. For this reason, performance-assessment analysts

prefer to sample from continuous CDFs for those variables that are known to be continuously distributed.

The empirical CDF described above can be modified and cast into a continuous distribution in several ways. Perhaps the simplest way is to draw straight lines between the vertices of the empirical CDF, i.e., the points $(0,0)$, (x_1, p_1) , (x_2, p_2) , ..., (x_M, p_M) on the graph of the CDF (for example, see the solid lines so drawn on Figure II-1 for the tortuosity data). The piecewise-linear CDF constructed in this way is here denoted by $F_\ell(\xi)$ and is analytically expressed by

$$F_\ell(\xi) = \begin{cases} 0 & \text{if } \xi \leq 0, \\ P_{m-1} + \frac{f_m(\xi - x_{m-1})}{N(x_m - x_{m-1})} & \text{if } x_{m-1} < \xi \leq x_m, \quad m = 1, 2, \dots, M, \\ 1 & \text{if } \xi > x_M, \end{cases}$$

where $p_0 = 0$ and $x_0 = 0$.

Inspection of the example shown on Figure II-1 reveals that drawing random variates from a piecewise-linear CDF will give back a random selection of all of the values of the variable X that lie between 0 and x_M , not just the original values $x_1, x_2, \dots, x_M$. The author has not found or been able to develop a proof that a piecewise-linear CDF constructed in this way is an unbiased estimator of the unknown distribution of the variable X .

The mean value or expected value of the CDF $F_\ell(\xi)$ is here denoted by $\langle X \rangle_\ell$ and can be expressed as

$$\langle X \rangle_\ell = (1/N) \sum_{m=1}^M f_m (x_m + x_{m-1})/2.$$

The variance of the CDF $F_\ell(\xi)$ is denoted by s_ℓ^2 and can be expressed as

$$s_\ell^2 = (1/N) \sum_{m=1}^M f_m (x_m^2 + x_m x_{m-1} + x_{m-1}^2)/3 - \langle X \rangle_\ell^2.$$

The author has not found or been able to develop a proof that $\langle X \rangle_\ell$ and s_ℓ^2 are unbiased estimators of the respective mean and variance of the unknown distribution $F(x)$. For the CDF for the tortuosity data shown on Figure II-1, $\langle X \rangle_\ell = 0.13$ and $s_\ell^2 = 5.0 \times 10^{-2}$.

It is somewhat surprising that the piecewise-linear CDF obtained by simply drawing straight lines between empirical-percentile points of an empirical CDF is the same distribution that is obtained by using the Maximum Entropy Formalism (MEF; to be discussed in the next section) and constraints specified by empirical percentile points.

The Maximum Entropy Formalism

The literature on the Maximum Entropy Formalism (MEF) is now vast; the reader should consult the reviews edited by Levine and Tribus (1978), or the recent monograph by Jumarie (1990), for thorough discussions of the foundations and areas of application of this subject. The MEF has been used before to construct prior probability distributions of uncertain variables in nuclear-risk assessment models: See Cook and Unwin (1986) and Unwin et al. (1989).

In this report, the MEF is simply regarded as a consistent mathematical procedure for the derivation of a probability distribution function for an uncertain variable, X , from a set of quantitative constraints on the form of that distribution; e.g., quantitative statements about the range, the mean, the variance, or the percentiles of the distribution. The quantitative constraints may be empirical constraints, i.e. constraints based on sample values of the variable, or subjective constraints based on professional judgment.

The central problem of the MEF is the determination of extrema of the so-called entropy functional, defined by

$$S(f) = - \int_a^b f(x) \ln[f(x)] dx,$$

over the set of all probability density functions, $f(x)$, which are nonzero in the range $[a,b]$ and which satisfy prescribed, quantitative constraints.

The entropy functional is the continuous version of the information-theoretic entropy

$$S = -k \sum_i P_i \ln P_i,$$

i.e., it is the expected value of Shannon's measure,

$$I(X_i) = -k \ln P_i, \text{ } k \text{ a constant,}$$

of the amount of information gained by observing the outcome of an experiment in which a random variable X_i is observed to take on the value x_i with probability P_i (Hamming 1991; Ch. 7). The entropy functional has also been interpreted as a measure of the amount of uncertainty inherent in a PDF or as a measure of the amount of information that would be required to specify completely the value of a random variable X (for the idea that entropy is "missing" information, see Baierlein, 1971). Thus, finding an extremum of the entropy functional subject to prescribed constraints can be construed as finding the PDF, within the set of all PDFs that incorporate the information inherent in the constraints, which maximizes the amount of remaining information that must be supplied in order to completely specify the value of the uncertain variable X . Use of the MEF can minimize the amount of spurious information that often enters into the construction of a PDF from sparse data or limited quantitative information.

The prescribed informational constraints are best expressed as integral constraints, i.e., they should take the form

$$\int_a^b g_m(x)f(x)dx = C_m, \quad m = 0, 1, 2, \dots, M,$$

where the g_m s are given, integrable functions of x on the interval $[a,b]$ and the C_m s are given constants. One necessary constraint on a PDF is that its integral over $[a,b]$ must equal one; thus one conventionally takes $g_0 = 1$ and $C_0 = 1$. By expressing the constraints in this way, one can derive a general solution to the problem (in the calculus of variations) of maximizing $S(f)$ subject to the $M+1$ constraints (see, for example, Tribus, 1969). The maximizing PDF, here denoted by $f^*(x)$, is given by

$$f^*(x) = Z^{-1}(\lambda_1, \lambda_2, \dots, \lambda_M) \exp \left[- \sum_{m=1}^M \lambda_m g_m(x) \right],$$

where Z^{-1} is the reciprocal of Z and

$$Z(\lambda_1, \lambda_2, \dots, \lambda_M) = \int_a^b \exp \left[- \sum_{m=1}^M \lambda_m g_m(x) \right] dx.$$

The λ_m , $1 \leq m \leq M$, are constants (Lagrange multipliers) to be determined by solving the following set of M equations in M unknowns:

$$-(\partial/\partial\lambda_m) \ln Z = C_m, \quad 1 \leq m \leq M.$$

The special forms of $f^*(x)$ that arise from this formalism when the constraints mentioned in the outline of the five-step procedure are applied are of particular interest:

- a. When only the range of X is given (i.e., no constraints other than normalization of the PDF), then $f^*(x)$ is the uniform distribution on the interval $[a,b]$. Obviously, this makes sense only if $|b-a| < \infty$, i.e., the range of the variable X is bounded.
- b. When the range and M percentile points of the CDF are given, then $f^*(x)$ is a weighted sum of M uniform distributions that vanishes outside the range $[a,b]$ and the associated CDF is piecewise linear. In this case, the $M \geq 1$ constraints are of the form

$$g_m(x) = u(x_m - x), \quad C_m = p_m, \quad m = 1, 2, 3, \dots, M,$$

where $u(\cdot)$ is the unit step function (Abramowitz and Stegun, 1964, p. 1020, 29.1.3), the x_m s are given percentile points in the interval $[a,b]$, and the p_m s are the corresponding percentiles.

- c. When the range, the mean value, and the variance (or coefficient of variation) of the variable X are given, then $f^*(x)$ is a truncated normal distribution that vanishes outside the interval $[a,b]$. In this case, the two constraints are of the form

$$g_1(x) = x, \quad C_1 = \mu; \quad g_2 = (x - \mu)^2, \quad C_2 = \sigma^2,$$

where μ and σ^2 are respectively the given mean value and variance.

- d. When the range and only the mean value of the variable X are given, then $f^*(x)$ is a truncated exponential distribution that vanishes outside the interval $[a,b]$. In this case, $g_1 = x$ and $C_1 = \mu$.

Proofs of Cases a, c, and d can be found in Tribus (1969). The author has not been able to locate a proof of Case b and has therefore supplied his own proof below.

Let the empirical or subjective percentile points be the given as $M \geq 1$ ordered pairs (x_1, p_1) , (x_2, p_2) , ..., (x_M, p_M) with

$$0 \leq a < x_1 < x_2 < \dots < x_M < b < \infty; \quad p_1 < p_2 < \dots < p_M$$

and $0 < p_m < 1$ for all $m > 0$. For convenience, define

$$x_0 = a, p_0 = 0; \quad x_{M+1} = b, p_{M+1} = 1.$$

The constraints on the candidate PDFs, $f(x)$, may then be written as

$$\int_a^b u(x_m - s)f(s)ds = p_m, \quad m = 0, 1, 2, \dots, M+1,$$

where $u(\cdot)$ is the unit step function (Abramowitz and Stegun, 1964, p. 1020). The PDF that maximizes the entropy functional is therefore

$$f^*(x) = \exp \left[\sum_{m=1}^M \lambda_m u(x_m - x) - 1 \right],$$

where the λ_m s are constants to be determined from the constraints. Inspection of this PDF shows that it is a piecewise-constant function on the interval $[a, b]$; i.e., $f^*(x) = A_m$, if $x_{m-1} < x \leq x_m$, with A_m a different constant for each $m = 1, 2, \dots, M+1$. The constants A_m are simply related to the constants λ_m , and it is easier to determine the A_m s from the constraints. For example, consider the integral of $f^*(x)$ between x_{m-1} and x_m . This integral is $(x_m - x_{m-1})A_m$, but by the constraints it is also equal to $(p_m - p_{m-1})$. It follows that

$$A_m = (p_m - p_{m-1})/(x_m - x_{m-1}), \quad m = 1, 2, \dots, M+1.$$

By integrating $f^*(x) = A_m$, $m = 1, 2, \dots, M+1$, between $x_0 = a$ and a point $\xi > a$, one finds the CDF associated with $f^*(x)$:

$$F^*(\xi) = \begin{cases} 0 & \text{if } 0 < \xi \leq a, \\ p_{m-1} + (p_m - p_{m-1})(\xi - x_{m-1})/(x_m - x_{m-1}), & \text{if } x_{m-1} < \xi \leq x_m, \\ 1 & \text{if } \xi > b. \end{cases}$$

This result is a piecewise-linear CDF of the kind described earlier in this chapter.

Once again, the reader should take note that in using the MEF, the ranges, percentiles and percentile points, mean values, and variances to be supplied in Cases a through d can be either empirical or subjective numbers; that is, they can be numbers derived from measurements of the variable X , or they can be furnished as the "best estimates" of the RIs. Of course, if only subjective estimates are used to form the parameters of an MEF distribution,

it is meaningless to inquire whether that distribution is an unbiased estimator of the unknown distribution, $F(x)$. The resulting distribution is purely subjective and can only reflect the accuracy of the PIs' best estimates of the distribution's parameters.

An Application of the Procedures

The most recent simulations of WIPP performance used probability distributions obtained by the five-step procedure described above. The results of this first, informal trial of the procedure are summarized in Table II-1: column 1 of the table names the 29 variables that were sampled in the recent simulations and gives their physical units; column 2 names the kind of distribution that was ultimately assigned; and column 3 briefly states the source of information and the basis for the assignment of the distribution named in column 2.

In this first trial of the procedures, no formal elicitation of expert judgment of the type suggested by Bonano et al. (1990) was used. A memo was sent to WIPP Project RIs in Department 6340 of Sandia National Laboratories asking that they provide any information they might have concerning each of the 29 variables; the requested information was to be supplied in one or more of the following forms and listed in order of decreasing preference on the part of the performance-assessment analyst:

- (1) A table of WIPP-specific, measured values of the variable.
- (2) Reasoned estimates of percentile points for the variable; i.e. the provision of statements like "90 percent of solubility values for radionuclide species A lie below 10^{-4} molar."
- (3) Reasoned estimates of the mean value and standard deviation of the variable.
- (4) Reasoned estimates of only the mean value of the variable.
- (5) At minimum, and always in addition to information of types 1 through 4, reasoned estimates of the maximum and minimum values (range) that the variable could assume in the context of the WIPP system.

In addition to a written request for information, informal meetings were held with the PIs in order to explain the purpose of the request for information and to help their understanding of some of the statistical terms used in the memorandum. These informal meetings revealed that some of the RIs were

TABLE II-1. PROBABILITY DISTRIBUTIONS FOR VARIABLES SAMPLED IN CURRENT WIPP PERFORMANCE SIMULATIONS*

Variable Name and Units	Type of Distribution	Source or Basis for Distribution†
1. Salado Capacitance (Pa ⁻¹)	Lognormal	Assigned by RI.
2. Salado Permeability (m ³)	Piecewise Linear	MEF-empirical percentiles from data provided by RI.
3. Salado Pressure (MPa)	Uniform	MEF-bound provided by RI.
4. Room-Waste Solubility (all radionuclide species, kg/kg)	Loguniform	Assigned by RI.
5. Room-Time of First Intrusion	Modified Exponential	Appendix C of Tierney (in prep.).
6. Brine Pocket Initial Pressure (MPa)	Piecewise Linear	MEF-bounds and median provided by RI.
7. Borehole Permeability m ²	Lognormal	Freeze and Cherry, 1979.
8. Borehole Porosity (dimensionless)	Normal	Freeze and Cherry, 1979.
9. Brine Pocket Bulk Volume (m ³)	Uniform	MEF-bounds provided by RI.
10. Culebra Tortuosity (dimensionless)	Piecewise Linear	MEF-empirical percentiles from data in Tables E-9 of Lappin et al., 1989.
11. Culebra Diffusion Coefficient (all radionuclide species, m ² /s)	Uniform	MEF-bounds are maximum and minimum of values given in Table A-8 of Rechard et al., 1990a.
12. Culebra Fracture Spacing (m)	Piecewise Linear	MEF-bounds and median provided by RI.

* A complete description of the probability distributions for all variables used in the 1990 performance simulations can be found in Rechard et al. (1990b).

† The RIs' responses that provided the sources or basis for each distribution are documented in Memos 3-11 and Letters 1a and 1b of Appendix A of Rechard et al. (1990b).

TABLE II-1. PROBABILITY DISTRIBUTIONS FOR VARIABLES SAMPLED IN CURRENT WIPP PERFORMANCE SIMULATIONS
(concluded)

Variable Name and Units	Type of Distribution	Source or Basis for Distribution
13. Culebra Recharge Factor (dimensionless)	Uniform	Marietta et al., in prep.
14. Culebra Precipitation Factor (dimensionless)	Uniform	Marietta et al., in prep.
15. Borehole cross-sectional area (m ²)	Empirical	Data provided by RI.
16-19. Culebra - Matrix Retardation Factors for Plutonium, Americium, Neptunium and Uranium (dimensionless)	Piecewise Linear	MEF-subjective percentiles (0, 25, 50, 75, 100) provided by RI.
20-23. Culebra - Fracture Retardation Factors for Plutonium, Americium, Neptunium and Uranium (dimensionless)	Piecewise Linear	MEF-subjective percentiles (0, 25, 50, 75, 100) provided by RI.
24-29. Culebra Hydraulic Conductivity for Zones 1-7 (m/s)	Piecewise Linear	MEF-empirical percentiles from data provided by RI.

1 confused about the meanings of the several measures of the shape of a
2 probability distribution (Figure II-2). In subsequent meetings, most of the
3 RIs agreed that, in the absence of data, they could not supply reasoned
4 estimates of the mean value, μ , or standard deviation, σ , of the unknown
5 distribution and that the measures of location they had previously called
6 "expected values" were more likely to be estimates of the median value, x_{50} ,
7 or the mode, $x_{\max}$, of the distribution.

8
9 If the RI provided the range (a,b) and an estimate of the median, x_{50} , the MEF
10 yielded the simple, piecewise-linear CDF illustrated in Figure II-3.
11 Providing a subjective estimate of the mode of an unknown PDF was discouraged.
12 In the absence of additional information about the value of the PDF at the
13 mode (information usually not known to an RI), the use of a subjective mode as
14 a constraint in the MEF only gives back the uniform distribution over the
15 range (a,b), the same distribution that arises if the range alone is
16 specified.

TRI-6342-580-0

2

Figure II-2. Typical PDF Showing the Different Measures of Location.

TRI-6342-667-0

5

Figure II-3. Piecewise-Linear CDF Based on Range and Median Value.

III. LIMITATIONS ON THE 1990 PROBABILITY DISTRIBUTIONS

The major limitations on the validity of the probability distributions constructed for the 1990 performance simulations are believed to be the consequence of two things:

- (1) The effects of spatial averaging on the variance of model variables used in lumped-parameter models were ignored.
- (2) Possible correlations between model variables were ignored.

The Effects of Spatial Averaging

Since most of the WIPP performance models are lumped-parameter models, many of the variables to be assigned CDFs in the WIPP performance models are actually spatial averages of physical quantities that can only be measured on spatial scales that are small compared with the spatial scale used in the models. For example, the effective hydraulic conductivity and porosity of a WIPP waste room (a structure having a volume of the order of 1000 m³) are actually volumetric averages over the local hydraulic conductivity and porosity of approximately 1000 consolidated waste units (collapsed waste barrels) each having volumes of the order of one cubic meter. The RI usually provides information about variability of a quantity on the smaller of the two spatial scales. It is easy to show that use of this small-scale variability to reflect directly the variance in the lumped-parameter model variable will result in unnecessarily conservative CDFs. Very roughly, the following relationship holds between the variance of a volumetric average and the variance of the "local," small-scale quantity:

$$\sigma_{ave}^2 \approx (v/V)\sigma_{loc}^2$$

where v is a correlation volume and V is the volume over which the local physical quantity is to be averaged (analogous relationships hold for linear and areal averages). Although the precise size of the correlation volume is not known in every case, it is usually known that $v \ll V$. It follows that the variance of a volumetric average may be much smaller than the apparent variance of the local quantity. On the other hand, the mean value of the volumetric average should be equal to the mean value of the local quantity. The picture of the PDF for a spatial average that emerges from these remarks

1 is one of a distribution that is sharply peaked about the mean value of the
2 local quantity. In the absence of other kinds of information indicating
3 uncertainty in the mean value of the local quantity, it would be inefficient
4 to sample from such a highly peaked distribution; the variable in question
5 would simply be assigned the best estimate of the mean value of the local
6 quantity.

7
8 Thus, in seeking more information about those model variables that are known
9 to be spatial averages of local quantities, it may be necessary to ask that
10 experts provide scales of measurements and correlation lengths, and state
11 their estimate of the uncertainty in the mean value of the local quantity, in
12 addition to providing the observed or perceived variability of the local
13 quantity itself.

14 15 16 **Correlations Between Model Variables**

17
18 All of the uncertain variables studied during the 1990 performance simulations
19 were assumed to be independent random variables, although it was known in
20 advance that many of them were interdependent, i.e. correlated in some way.
21 Correlations of the model variables may arise from the fact that there are
22 natural correlations between the local quantities used to determine the form
23 of the model variable (e.g., local porosity could be strongly correlated with
24 local permeability); or correlations of model variables may be implicit in the
25 form of the mathematical model in which they are used. As an example of the
26 latter circumstance, the current model for predicting WIPP-room hydraulic
27 conductivity and porosity (see Rechar 1990b, Chapter III) makes these
28 variables depend upon the volume fractions of specific waste forms (i.e.,
29 fractions of combustibles, metallics, sludges, etc.) contained in the entire
30 waste inventory. These volume fractions are obviously uncertain variables
31 themselves even though they were not treated as variables in the 1990
32 performance simulations. Taking account of the uncertainty in volume
33 fractions would change estimates of the uncertainty in the mean value of the
34 WIPP-room hydraulic conductivity and porosity.

35
36 Correlations among the important variables of the WIPP performance models need
37 to be examined in detail since these model-dependent correlations may either
38 increase or decrease the variance of a particular variable, and therefore
39 effectively change the shape of that variable's CDF.

REFERENCES

- Abramowitz, M., and I. A. Stegun. 1964. *Handbook of Mathematical Functions*, AMS 55. Washington, DC: National Bureau of Standards, U.S. Department of Commerce.
- Baierlein, R. 1971. *Atoms and Information Theory*. San Francisco, CA: W. H. Freeman and Co.
- Blom, Gunnar. 1989. *Probability and Statistics: Theory and Applications*. New York: Springer-Verlag.
- Bonano, E. J., S. C. Hora, R. L. Keeney, D. von Winterfeldt. 1990. *Elicitation and Use of Expert Judgement in Performance Assessment for High-Level Radioactive Waste Repositories*. NUREG/CR-5411, SAND89-1821. Albuquerque, NM: Sandia National Laboratories.
- Cook, Ian, and S. D. Unwin. 1986. "Controlling Principles for Prior Probability Assignments in Nuclear Risk Assessment," *Nuclear Science and Engineering*: 94, 107-119.
- Freeze, R. A., and J. A. Cherry. 1979. *Groundwater*. Englewood Cliffs, NJ: Prentice Hall, Inc.
- Hamming, R. W. 1991. *The Art of Probability for Scientists and Engineers*. Reading, MA: Addison-Wesley Publishing Company, Inc.
- Jumarie, G. 1990. *Relative Information: Theories and Applications*. New York: Springer-Verlag.
- Lappin, A. R., R. L. Hunter, D. P. Garber, and P. B. Davies, assoc. eds. 1989. *Systems Analysis, Long-Term Radionuclide Transport, and Dose Assessments, Waste Isolation Pilot Plant (WIPP), Southeastern New Mexico; March 1989*. SAND89-0462. Albuquerque, NM: Sandia National Laboratories.
- Levine, R. D., and M. Tribus. eds. 1978. *The Maximum Entropy Formalism*. Cambridge, MA: The MIT Press.
- Marietta, M. G., S. G. Bertram-Howery, D. R. Anderson, F. K. Brinster, R. V. Guzowski, H. Iuzzolino, and R. P. Rechard. 1989. *Performance Assessment Methodology Demonstration: Methodology Development for Evaluating Compliance With EPA 40 CFR 191, Subpart B, for the Waste Isolation Pilot Plant (WIPP)*. SAND89-2027. Albuquerque, NM: Sandia National Laboratories.
- Marietta, M. G., P. N. Swift, B. L. Baker, K. F. Brinster, and P. J. Roache. 19___. *Parameter and Boundary Conditions Sensitivity Studies as Related to Climate Variability and Scenario Screening for the Waste Isolation Pilot Plant*. SAND89-2029. Albuquerque, NM: Sandia National Laboratories. In preparation.

References

- 1 Public Law 96-164. 1980. Department of Energy National Security and
2 Military Applications of Nuclear Energy Authorization Act of 1980, Title II-
3 General Provisions: Waste Isolation Pilot Plant, Delaware Basin, NM.
4 Pub. L. No. 96-164, 93 Stat. 1259.
5
6 Rechard, R. P., H. J. Iuzzolino, J. S. Rath, A. P. Gilkey, R. D. McCurley,
7 and D. K. Rudeen. 1989. *User's Manual for CAMCON: Compliance Assessment*
8 *Methodology Controller*. SAND88-1496. Albuquerque, NM: Sandia National
9 Laboratories.
10
11 Rechard, R. P., W. Beyeler, R. D. McCurley, D. K. Rudeen, J. E. Bean, and J.
12 D. Schreiber. 1990a. *Parameter Sensitivity Studies of Selected Components*
13 *of the Waste Isolation Pilot Plant Repository/Shaft System*. SAND89-2030.
14 Albuquerque, NM: Sandia National Laboratories.
15
16 Rechard, R. P., H. J. Iuzzolino, and J. S. Sandha. 1990b. *Data Used in*
17 *Preliminary Performance Assessment of the Waste Isolation Pilot Plant*
18 *(1990)*, SAND89-2408. Albuquerque, NM: Sandia National Laboratories.
19
20 Tierney, M. S., 19___. *Combining Scenarios in a Calculation of the Overall*
21 *Probability Distribution of Cumulative Releases of Radioactivity from the*
22 *Waste Isolation Pilot Plant, Southeastern New Mexico*. SAND90-0838.
23 Albuquerque, NM: Sandia National Laboratories. In preparation.
24
25 Tribus, Myron. 1969. *Rational Descriptions, Decisions and Designs*. New
26 York: Pergamon Press.
27
28 Unwin, S. D., E. G. Cazzoli, R. E. Davis, M. Khatib-Rahbar, M. Lee, H.
29 Nourbakhsh, C. K. Park, and E. Schmidt. 1989. "An Information-Theoretic
30 Basis for Uncertainty Analysis: Application to the QUASAR Severe Accident
31 Study," *Reliability Engineering and System Safety*: 26, 143-162.
32
33 U.S. DOE (Department of Energy). 1980. *Final Environmental Impact*
34 *Statement, Waste Isolation Pilot Plant*. DOE/EIS-0026. Washington, DC:
35 United States Government Printing Office.
36
37 U.S. DOE (Department of Energy). 1990. *Final Supplement Environmental*
38 *Impact Statement, Waste Isolation Pilot Plant*. DOE/EIS-0026-FS.
39 Washington, DC: U.S. Department of Energy, Office of Environmental
40 Restoration.
41
42 U.S. Department of Energy and State of New Mexico. 1981, as modified.
43 "Agreement for Consultation and Cooperation between the U.S. Department of
44 Energy and the State of New Mexico on the Waste Isolation Pilot Plant,"
45 modified 11/30/84, 8/4/87, and 3/22/88.
46
47 U.S. Environmental Protection Agency. 1985. "Environmental Standard for the
48 Management and Disposal of Spent Nuclear Fuel, High-Level and Transuranic
49 Waste; Final Rule." 40 CFR 191. *Federal Register*, 50: 38066-38089.
50

GLOSSARY

ccdf - see complementary cumulative distribution function.

cdf - see cumulative distribution function.

complementary cumulative distribution function (CCDF) - One minus the cumulative distribution function.

Culebra Dolomite Member - The lower of two layers of dolomite within the Rustler Formation that are locally water bearing.

cumulative distribution function - The sum (or integral as appropriate) of the probability of those values of a random variable that are less than or equal to a specified value.

empirical - Relying explicitly upon or derived explicitly from observation or experiment.

exponential distribution - A probability distribution whose PDF is an exponential function defined on the range of the variable in question.

hydraulic conductivity - The measure of the rate of flow of water through a unit cross-sectional area under a unit hydraulic gradient.

lognormal distribution - A probability distribution in which the logarithm of the variable in question follows a normal distribution.

loguniform distribution - A probability distribution in which the logarithm of the variable in question follows a uniform distribution.

mean - The expectation of a random variable; i.e., the sum (or integral) of the product of the variable and the PDF over the range of the variable.

median - That value of a random variable at which its CDF takes the value 0.5; i.e., the 50th percentile point.

mode - That value of a random variable at which its PDF takes its maximum value.

normal distribution - A probability distribution in which the PDF is a symmetric, bell-shaped curve of bounded amplitude extending from minus infinity to plus infinity.

Glossary

1 PDF - see probability density function.

2
3 porosity - The percentage of total rock volume occupied by voids.

4
5 probability density function - For a continuous random variable X , the
6 function giving the probability that X lies in the interval x to $x + dx$
7 centered about a specified value x .

8
9 solubility - The equilibrium concentration of a solute when undissolved
10 solute is in contact with the solution.

11
12 subjective - The opposite of empirical: not supported by explicit records of
13 measurements or experiments.

14
15 tortuosity - A measure of the actual length of the path of flow through a
16 porous medium.

17
18 truncated distribution - A probability distribution whose curve is defined on
19 a range of variable values that is smaller than the range normally associated
20 with the distribution: e.g., a normal distribution defined on a finite range
21 of variable values.

22
23 uniform distribution - A probability distribution in which the PDF is
24 constant over the range of variable values.

25
26 variance - The square of the standard deviation of a probability
27 distribution; the standard deviation is a measure of the amount of spreading
28 of a PDF about its mean.

END

DATE FILMED

02 / 19 / 91