Uncertainty and Sensitivity Analyses Plan

**Hanford Environmental
Dose Reconstruction Project**

**J. C. Simpson
J. V. Ramsdell, Jr.**

April 1993

**Letter report prepared for
review and approval by the
Technical Steering Panel and the
Centers for Disease Control and Prevention
under Contract 200-92-0503(CDC)/18620(BNW)**

**Battelle
Pacific Northwest Laboratories
Richland, Washington 99352**

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MAST

LEGAL NOTICE

This report was prepared by Battelle as an account of sponsored research activities. Neither Sponsor nor Battelle nor any person acting on behalf of either:

MAKES ANY WARRANTY OR REPRESENTATION, EXPRESS OR IMPLIED, with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, process, or composition disclosed in this report may not infringe privately owned rights; or

Assumes any liabilities with respect to the use of, or damages resulting from the use of, any information, apparatus, process, or composition disclosed in this report.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply endorsement, recommendation, or favoring by Sponsor or Battelle.

Printed in the United States of America

**Available to DOE and DOE contractors from the
Office of Scientific and Technical Information, P.O. Box 62, Oak Ridge, TN 37831;
prices available from (615) 576-8401. FTS 626-8401.**

**Available to the public from the National Technical Information Service,
U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161.**

Preface

In 1987, the U.S. Department of Energy (DOE) directed the Pacific Northwest Laboratory, operated by Battelle Memorial Institute, to conduct the Hanford Environmental Dose Reconstruction (HEDR) Project. The DOE directive to begin project work followed a 1986 recommendation by the Hanford Health Effects Review (HHER) Panel. The HHER Panel was formed to consider the potential health implications of past Hanford-Site releases of radioactive materials.

A December 1990 Memorandum of Understanding between the Secretaries of the DOE and the U.S. Department of Health and Human Services (DHHS) transferred responsibility for managing the DOE's dose reconstruction and exposure assessment studies to the DHHS. This transfer resulted in the current contract between Battelle, Pacific Northwest Laboratories (BNW) and the Centers for Disease Control and Prevention (CDC), an agency of the DHHS.

An 18-member Technical Steering Panel (TSP) was selected to direct the project work. The TSP consists of experts in the various technical fields relevant to HEDR Project work and representatives from the states of Washington, Oregon, and Idaho; Native American Tribes; and the public.

The purpose of the HEDR Project is to estimate the radiation dose that individuals may have received as a result of emissions since 1944 from DOE's Hanford Site near Richland, Washington. One major objective of the HEDR project is to estimate doses to the thyroid of individuals who have been exposed to iodine-131 released into the air from the Hanford facilities. A second major objective is to estimate doses that an individual may have received from the release of other radionuclides into the Columbia River. Mathematical models are being developed to estimate the radiation doses.

The purpose of this Uncertainty and Sensitivity Analyses Plan is to provide the TSP with the information necessary to make a decision about the appropriate method for analyzing uncertainty in the doses that individuals may have received from operations at Hanford. The proposed uncertainty analyses allow researchers to make quantitative statements about the likelihood that an unknown dose is within a specified interval. The proposed sensitivity analyses allow researchers to determine which individual parameters have uncertainties with the largest impact on the uncertainty of the predicted dose.

This draft plan will be reviewed by an international panel prior to the uncertainty/sensitivity analysis peer review workshop to be held May 24-25, 1993 in Richland, Washington. The report is being submitted now for technical peer review and comments on the proposed approach. This report meets HEDR Milestone 0803A. Final editorial review will be completed upon receipt of the peer review panel's report.

Executive Summary

Hanford Environmental Dose Reconstruction (HEDR) Project staff are developing mathematical models to be used to estimate the radiation dose that individuals may have received as a result of emissions since 1944 from the U.S. Department of Energy's (DOE) Hanford Site near Richland, Washington. An uncertainty and sensitivity analyses plan is essential to understand and interpret the predictions from these mathematical models. This is especially true in the case of the HEDR models where the values of many parameters are unknown.

This plan gives a thorough documentation of the uncertainty and hierarchical sensitivity analysis methods recommended for use on all HEDR mathematical models. The documentation includes both technical definitions and examples. In addition, an extensive demonstration of the uncertainty and sensitivity analysis process is provided using actual results from the Hanford Environmental Dose Reconstruction Integrated Codes (HEDRIC). This demonstration shows how the approaches used in the recommended plan can be adapted for all dose predictions in the HEDR Project.

From the documentation and demonstration in this plan, the following recommendations are made for the use of uncertainty and sensitivity analyses in the HEDR Project.

- All dose predictions calculated using mathematical models developed by the HEDR Project should be accompanied by a description of the uncertainty in the prediction.
 - The Monte Carlo method should be the only approach used to estimate uncertainty in estimated doses.
 - When a deterministic model is used, commercially available software should be used to calculate the uncertainties using the Monte Carlo approach for a selection of representative locations and individuals. The uncertainty estimates should be extended to specific cases by analogy.
 - The uncertainties in predicted doses should be presented graphically using boxplots.
- Hierarchical sensitivity analyses should be performed on all mathematical models developed by the HEDR Project.
 - Measures of sensitivity derived from multiple regression should be used in the sensitivity analyses, whenever possible.
 - When the multiple regression approach is not possible, the uncertainties should be evaluated by holding subsets of the parameters constant.
 - Sensitivity analyses should be performed for a selection of representative locations and individuals, using a hierarchical approach.

An advantage of this approach is that it can be applied across all the dose predictions and mathematical models developed by the HEDR Project. The consistent use of one unified approach enhances the interpretation and communication of the results of the uncertainty and sensitivity analyses.

Glossary

Average cow	Surrogate for a herd of cows where milk is averaged across cows
BNW	Battelle, Pacific Northwest Laboratories
CIDER	Calculation of Individual Doses from Environmental Radionuclides
CDC	Centers for Disease Control and Prevention
CDF	Cumulative Distribution Function
Ci	curies
CPD	Coefficient of Partial Determination
CV	Coefficient of Variation
DESCARTES	Dynamic Estimates of Concentrations and Radionuclides in Terrestrial Environments
deterministic answer of a model	A singular answer obtained by using a single number for each model parameter
DHHS	U.S. Department of Health and Human Services
DOE	U.S. Department of Energy
groundshine	Reflection back to individual resulting from iodine-131 in the soil
HEDR	Hanford Environmental Dose Reconstruction Project
HEDRIC	Hanford Environmental Dose Reconstruction Integrated Codes
HHER	Hanford Health Effects Review
Individual cow	Single cow whose milk is the only source for an entire year
LHS	Latin hypercube sampling
Monte Carlo	A computer simulation method of estimating distributions of model predictions when model input is uncertain
PDF	Probability Density Functions
rad	A unit of energy absorbed from ionizing radiation (equal to 100 ergs per gram of irradiated material)
RATCHET	Regional Atmospheric Transport Code for Hanford Emissions Tracking Source Term Release Model

RM	Reactor model
RS	Random sampling
SRC	Standardized Regression Coefficient
SPD	Standardized Partial Differential
SRS	Stratified random sampling
STRM	Source Term Release Model
TSP	Technical Steering Panel

Contents

Preface	iii
Executive Summary	v
1.0 Introduction	1.1
1.1 Uncertainty and Sensitivity Analyses in the HEDR Project	1.1
1.2 Uncertainty and Sensitivity Analyses Plan Document	1.2
1.3 Structure of this Document	1.2
2.0 Hanford Environmental Dose Reconstruction Conceptual and Computational Models	2.1
2.1 Conceptual Model: Elements that Determine Dose	2.1
2.1.1 Concentration in Food	2.3
2.1.2 Quantity of Iodine-131 in Plants and Soil	2.3
2.1.3 Surface Deposition and Air Concentration of Iodine-131	2.6
2.1.4 Puff Transport—Transport of Iodine from the Source to a Given Location	2.6
2.1.5 Iodine-131 Releases from the Hanford Site	2.6
2.2 Overview of the Hanford Environmental Dose Reconstruction Integrated Codes	2.10
3.0 Uncertainty Analysis Approaches	3.1
3.1 Identification of Assessment Question and Model	3.1
3.1.1 Assessment Question 1	3.2
3.1.2 Assessment Question 2	3.4
3.1.3 Assessment Question 3	3.5
3.1.4 Assessment Question 4	3.6
3.2 Strategy for Specifying Parameter Uncertainties	3.7
3.2.1 Estimating the Maximum Conceivable Parameter Range	3.8
3.2.2 Obtaining Subjective Information	3.8
3.2.3 Selecting a Probability Density Function	3.10
3.2.4 Parameter Selection Frequency	3.11

3.3	Estimation of Uncertainty Using Mathematical Techniques	3.11
3.3.1	Analytical Analysis of the Model	3.11
3.3.2	Differential Analysis of the Model	3.15
3.4	Estimation of Uncertainty Using Monte Carlo Technique	3.25
3.4.1	Example of Monte Carlo Technique	3.25
3.4.2	Statistical Sampling Strategies	3.27
3.5	Description of Uncertainty in Prediction	3.31
3.5.1	Variance, Standard Deviation, and Coefficient of Variation	3.31
3.5.2	Subjective Confidence Interval	3.31
3.5.3	Boxplot	3.33
3.6	Comparison of Mathematical and Monte Carlo Techniques	3.34
4.0	Sensitivity Analysis Approaches	4.1
4.1	Sensitivity Analysis of Mathematically Estimated Uncertainty	4.1
4.2	Sensitivity Analysis of Monte Carlo Estimated Uncertainty	4.3
4.2.1	Measures from Multiple Regression	4.4
4.2.2	Alternative Measures When Multiple Regression is Inappropriate	4.6
4.3	Hierarchical Uncertainty and Sensitivity Analyses	4.6
5.0	Demonstration of Uncertainty and Sensitivity Analysis Methods	5.1
5.1	Demonstration Overview	5.2
5.2	Uncertainty and Sensitivity Analyses of Dose from Milk Ingestion	5.6
5.2.1	Model and Parameters	5.6
5.2.2	Uncertainty Analysis	5.7
5.2.3	Sensitivity Analysis	5.10
5.2.4	Summary of Results from the Uncertainty and Sensitivity Analyses ...	5.11
5.3	Uncertainty and Sensitivity Analyses of Iodine-131 Concentration in Milk	5.16
5.3.1	Model and Parameters	5.16
5.3.2	Uncertainty Analysis	5.20
5.3.3	Sensitivity Analysis	5.21

5.3.4	Summary of Results from the Uncertainty and Sensitivity Analyses . . .	5.28
5.4	Uncertainty and Sensitivity Analyses of Iodine-131 Concentration in Pasture	5.29
5.4.1	Model and Parameters	5.29
5.4.2	Uncertainty Analysis	5.29
5.4.3	Sensitivity Analysis	5.32
5.4.4	Summary of Results from the Uncertainty and Sensitivity Analyses . . .	5.33
5.5	Uncertainty and Sensitivity Analyses of Iodine-131 Concentration in Harvested Plant Products	5.36
5.5.1	Model and Parameters	5.37
5.5.2	Uncertainty Analysis	5.38
5.5.3	Sensitivity Analysis	5.38
5.5.4	Summary of Results from the Uncertainty and Sensitivity Analyses . . .	5.40
5.6	Uncertainty and Sensitivity Analyses of Surface Deposition Iodine-131	5.42
5.6.1	Model and Parameters	5.42
5.6.2	Uncertainty Analysis	5.44
5.6.3	Sensitivity Analysis	5.44
5.6.4	Summary of Results from the Uncertainty and Sensitivity Analyses . . .	5.48
5.7	Conclusions	5.49
6.0	Recommendations, Data Quality Objectives, and Conclusions	6.1
6.1	Recommended Approach for Uncertainty Analyses of HEDR Models	6.1
6.2	Recommendations for Sensitivity Analyses of HEDR Mathematical Models . . .	6.2
6.3	Extent to Which Sensitivity Analyses Should be Performed	6.2
6.4	Data Quality Objectives for Uncertainty and Sensitivity Analyses	6.3
6.5	Conclusions	6.3
7.0	References	7.1
	Appendix A - Statistical Terminology Used in Uncertainty and Sensitivity Analyses	A.1
	Appendix B - Probability Distributions Used in Uncertainty and Sensitivity Analyses	B.1

Figures

2.1	Conceptual Model of the Elements that Determine Total Dose	2.2
2.2	Conceptual Model of the Concentration in Food and Biomass of Plant Product	2.4
2.3	Conceptual Model of the Quantity of Iodine-131 in Plants and Soil	2.5
2.4	Conceptual Model of the Surface Deposition of Iodine-131	2.7
2.5	Conceptual Model of the Puff Transport	2.8
2.6	Conceptual Model of the Source Term for Iodine-131	2.9
2.7	Components of Hanford Environmental Dose Reconstruction Integrated Codes	2.11
3.1	Stratified Random Sampling Combinations	3.28
3.2	Means and Standard Deviations from Two Kinds of Sampling	3.30
3.3	Envelopes Containing Cumulative Distribution Functions	3.30
3.4	Estimated Cumulative Probability Distribution Functions and Boxplots of Predicted Dose Using Mathematical and Monte Carlo Techniques	3.32
5.1	HEDR Atmospheric Model Domain	5.3
5.2	Average of 100 Predicted Weekly Total Surface Depositions and the Harvest/Pasture Seasons - 1945	5.5
5.3	Boxplots of the Annual Dose to a Reference Adult Male from Ingestion of Cow's Milk Containing Iodine-131 - 1945	5.9
5.4	Sensitivity Measures for Annual Dose from Ingestion of Milk from Individual Cow, Feeding Regime 1	5.12
5.5	Sensitivity Measures for Annual Dose from Ingestion of Milk from Individual Cow, Feeding Regime 3	5.13
5.6	Sensitivity Measures for Annual Dose from Ingestion of Milk from Average Cow, Feeding Regime 1	5.14
5.7	Sensitivity Measures for Annual Dose from Ingestion of Milk from Average Cow, Feeding Regime 3	5.15
5.8	Boxplots of the Average Daily Iodine-131 Concentration in Cow's Milk - 1945	5.22
5.9	Sensitivity Measures for Average Daily Iodine-131 Concentration in Milk from Individual Cow, Feeding Regime 1 - 1945	5.24
5.10	Sensitivity Measures for Average Daily Iodine-131 Concentration in Milk from Individual Cow, Feeding Regime 3 - 1945	5.25

5.11	Sensitivity Measures for Average Daily Iodine-131 Concentration in Milk from Average Cow, Feeding Regime 1 - 1945	5.26
5.12	Sensitivity Measures for Average Daily Iodine-131 Concentration in Milk from Average Cow, Feeding Regime 3 - 1945	5.27
5.13	Boxplots of the Average Daily Concentration of Iodine-131 in Pasture Season - 1945	5.31
5.14	Sensitivity Measures for Average Daily Concentration of Iodine-131 in Pasture - 1945	5.34
5.15	Sensitivity Measures for Average Daily Concentration of Iodine-131 in Pasture - 1945	5.35
5.16	Boxplots of the Concentration of Iodine-131 in Grain at Harvest - 1945	5.39
5.17	Sensitivity Measures for Concentration of Iodine-131 in Grain at Harvest - 1945	5.41
5.18	Boxplots of the Annual Surface Deposition - 1945	5.43
5.19	Sensitivity Measures for Annual Surface Deposition - 1945	5.46
5.20	Ranking of Parameters by the Importance of Their Uncertainty on the Uncertainty in Dose from Ingestion of Milk from Cow's on Feeding Regimes 1 and 3	5.50

Tables

4.1	Sensitivity Analysis Measures for Model of $Dose_{milk\ ingestion}$ in Examples 12 and 13	4.4
5.1	Average Predicted Annual Source Term and Surface Contamination of Iodine-131 - 1945	5.4
5.2	Mean and Coefficient of Variation of 100 Predicted Concentrations of Iodine-131 in Cow's Milk - 1945	5.8
5.3	Mean and Coefficient of Variation of 100 Predicted Doses from Cow's Milk Containing Iodine-131 - 1945	5.8
5.4	5th and 95th Subjective Confidence Limits Giving 90 Percent Subjective Confidence Interval for Dose - 1945	5.10
5.5	Minimum, Mode, and Maximum of Triangular PDFs for Daily Consumption Feeding Regime	5.18
5.6	Mean and Coefficient of Variation of the 100 Annual Consumption Values at Eltopia	5.19
5.7	Mean and Coefficient of Variation of 100 Predicted Concentrations of Iodine-131 in Harvested Food - 1945	5.19
5.8	Mean and Coefficient of Variation of 100 Predicted Average Daily Concentrations of Iodine-131 in Pasture and Soil - 1945	5.20
5.9	5th and 95th Subjective Confidence Limits That Give a 90 Percent Subjective Confidence Interval for Average Daily Concentration of Iodine-131 in Milk - 1945	5.21
5.10	PDFs of Parameters Used in Coupled Differential Equations for the Concentration of Iodine-131 in Pasture	5.30
5.11	Average and Coefficient of Variation of 100 Predicted Total and Daily Surface Deposition of Iodine-131 During the Pasture Season	5.30
5.12	5th and 95th Subjective Confidence Limits Which Give a 90 Percent Subjective Confidence Interval for Predicted Average Daily Concentration of Iodine-131 in Pasture	5.32
5.13	PDFs of Parameters Used in Coupled Differential Equations for the Concentration of Iodine-131 Which are Specific to Grain	5.37
5.14	5th and 95th Subjective Confidence Limits Which Give a 90 Percent Subjective Confidence Interval for Predicted Concentration of Iodine-131 in Grain at Harvest	5.40
5.15	5th and 95th Subjective Confidence Limits Which Give a 90 Percent Subjective Confidence Interval for Predicted Annual Surface Deposition of Iodine-131	5.45

5.16	Average of Predictions of Integrated Air Concentration of Iodine-131 for April 1945 from Four RATCHET Runs with Subsets of the Parameters Held Constant	5.45
5.17	Coefficients of Variations for 50 Predictions of Integrated Air Concentration of Iodine-131 for April 1945 from Four RATCHET Runs with Subsets of the Parameters Held Constant	5.48

Examples

Example 1:	Computational Model 1: Assessment Question/Deterministic Answer	3.3
Example 2:	Computational Model 2: Assessment Question/Deterministic Answer	3.4
Example 3:	Computational Model 3: Assessment Question/Deterministic Answer	3.5
Example 4:	Computational Model 4: Assessment Question/Deterministic Answer	3.6
Example 5:	Computational Model 5: Assessment Question/Stochastic Answer	3.7
Example 6:	Additive Model	3.12
Example 7:	Multiplicative Model (with Independent Parameters)	3.14
Example 8:	Differential Analysis (Continuation of Example 6)	3.18
Example 9:	Continuation of Example 7	3.20
Example 10:	Computational Model 6: Assessment Question/Deterministic Answer	3.25
Example 11:	Monte Carlo with Simple Random Sampling to Calculate Milk Ingestion for Model in Example 10	3.26
Example 12:	Sensitivity Analysis of Example 9	4.1

1.0 Introduction

The primary objective of the Hanford Environmental Dose Reconstruction (HEDR) Project is to estimate radiation doses to specific and reference individuals from exposure to radionuclide emissions that began in 1944 at the Hanford Site. The HEDR Project will use mathematical models to generate estimates of these doses and their uncertainties. These mathematical models, collectively, form the Hanford Environmental Dose Reconstruction Integrated Codes (HEDRIC). HEDRIC (see Chapter 2.0) is the set of computer codes that estimate doses using the mathematical models.

The values of most parameters in these models are not well-known (Snyder et al. 1992) because of incomplete knowledge regarding

- amounts of radionuclides released into the environment
- airborne transport of the released radionuclides
- accumulation of radioactivity in food products
- production and distribution of food products
- lifestyles and diets of specific or reference individuals.

An important objective of the HEDR Project is to estimate the uncertainty of computed doses that arises from uncertainty about parameter values.

Section 1.1 presents an overview of the uncertainty and sensitivity analysis process being used by the HEDR Project. The objectives of this document are presented in Section 1.2. Section 1.3 outlines the structure of this document.

1.1 Uncertainty and Sensitivity Analyses in the HEDR Project

Uncertainty and sensitivity analyses are critical in effectively understanding and interpreting predictions from mathematical models (Iman and Helton 1988). These analyses help to determine the uncertainty of model predictions, as well as which pathways (Section 2.1) and/or parameters are the most important contributors to uncertainty in the predicted dose. The uncertainty and sensitivity analysis process is used to understand the uncertainties in the model predictions that are due to uncertainties in the model parameters. The steps in this process are (IAEA 1989, Finkle 1990, Morgan and Henrion 1990):

1. Identifying the problem: identifying the assessment question and what must be estimated to answer that question
2. Specifying the conceptual and computational models: determining which exposure pathways will be included in the model and what mathematical equations will be used to represent (model) the pathways
3. Parameterizing the models: determining the numerical values to be used for each parameter of the mathematical equations

4. Estimating the uncertainty in dose: propagating the parameter uncertainties into the total uncertainty in the predicted dose
5. Deriving quantitative statements of uncertainty: presenting the quantitative statement about the likelihood that an unknown dose is within a specific interval of doses
6. Sensitivity analysis: identifying the parameters and exposure pathways in the mathematical models that contribute the most to uncertainty in predicted dose
7. Combining the preceding steps to develop an understanding of the effects of parameter uncertainty on the predicted dose
8. Presenting and interpreting the results of the analysis.

Although the uncertainty and sensitivity analysis process is set forth in a linear manner, the results at any one step can often send the investigator back to one of the previous steps and the process begins again from that step.

1.2 Uncertainty and Sensitivity Analyses Plan Document

This document has the following objectives:

- To provide a recommended plan for using the uncertainty and sensitivity analyses for all HEDR mathematical models
- To provide a detailed explanation of each step in the uncertainty and sensitivity analysis process outlined in Section 1.1
- To demonstrate how the uncertainty and sensitivity analysis process works with actual results from HEDRIC.

Although the recommended plan for conducting uncertainty and sensitivity analyses is the ultimate objective of this document, a considerable effort has been made here to explain the steps of the analysis process and to demonstrate that the recommended uncertainty and sensitivity analysis process can be successfully implemented for the HEDR models. The process is demonstrated here to be sound for the complex air pathway model. The basic process will be applied to other models used in the HEDR Project such as the Columbia River model.

1.3 Structure of this Document

Section 2 gives an overview of the conceptual model which is the basis for HEDRIC. Sections 3 and 4 give a detailed explanation, with examples, of the first six steps in the uncertainty and sensitivity analysis process presented in Section 1.1. More specifically, Section 3 discusses Steps 1 through 5 of the uncertainty and sensitivity analysis process outlined in Section 1.1. Section 4 is devoted to sensitivity analysis approaches (Step 6). Section 5 demonstrates the uncertainty and sensitivity analysis process on actual results from HEDRIC. Section 6 of this report gives the recommendations for the

uncertainty and sensitivity analyses of all HEDR models. Section 7 lists the references used in developing this plan. The document has two appendices. The statistical terminology used in this report is defined in Appendix A. In Appendix B, the probability density functions used in the HEDR models are defined.

2.0 Hanford Environmental Dose Reconstruction Conceptual and Computational Models

The HEDR Project is using a series of five integrated computer models to compute doses received by individuals from exposure to radionuclides from Hanford via the air pathway. This series of computer models is called HEDRIC (Hanford Environmental Dose Reconstruction Integrated Codes). HEDRIC computes the doses on the basis of equations that express mathematically the conceptual models believed to be appropriate for the HEDR project. In this chapter, we begin in Section 2.1 by discussing the conceptual model the HEDR Project is using to answer the following assessment question:

What is the total dose due to iodine-131 received by a specific individual during a given time period and at a given location from the air pathway?

This discussion is followed by Section 2.2 which briefly outlines the computational models developed from these conceptual models. Although the calculational flow of the HEDR models is linear, beginning with the production of iodine-131 in the reactor and ending with dose, the discussion of the HEDR conceptual and computational models presented here is given in reverse order. The order of presentation is so the reader will appreciate those model components that are integral to the estimation of dose. The presentation of the conceptual model begins with a discussion of the major pathways that lead to dose. The order of presentation matches the order in which the hierarchical sensitivity analysis method in Section 5 is applied (that is, beginning with dose and working backwards through the model to determine the relative uncertainty that is induced by each model component). Detailed information of the computational models is provided in Heeb (1993), Ramsdell and Burk (1992), Snyder et al. (1992), Eslinger (1992), Eslinger et al. (1992), and Miley et al. (1992). The reader is encouraged to consult these documents for a more complete discussion of the HEDR models.

2.1 Conceptual Model: Elements that Determine Dose

The total dose of iodine-131 that an individual may have received from Hanford operations comes from four exposure pathways in increasing order of importance—groundshine, immersion, inhalation, and ingestion. These pathways are shown in Figure 2.1. For each of these pathways, a dose factor converts the intake of iodine-131 to dose in units of rad.

The dose from *groundshine* results from the irradiation of the individual by iodine-131 on the ground. The dose from groundshine depends on the time the individual may have spent outdoors, the amount of shielding (such as clothing), and the concentration of iodine-131 in the soil.

The *immersion* dose results when an individual is surrounded by air containing iodine-131. Time spent outdoors is again a factor. Because air inside a building is not expected to have the same concentration of iodine-131 as outdoor air, an air exchange (the ratio of indoor air to outdoor air) must also be factored in for time spent indoors. As in the case of groundshine, shielding also plays an important role.

The *inhalation* dose is affected by the individual's activities and lifestyle. Within this pathway, breathing rate, the fraction of time spent outdoors, and the ratio of indoor to outdoor air must be considered. Dust is inhaled, so the concentration of iodine-131 in dust needs to be determined. This

Figure 2.1. Conceptual Model of the Elements that Determine Total Dose

amount is a function of the quantity of iodine-131 in the soil, the density of the soil, and the mass of particulates in the outdoor air (mass loading factor).

Ingestion dose depends on both the concentration of radionuclides that may have been present in food products and the amount of food consumed.

The concentration of iodine-131 in food, the quantity of iodine-131 in soil, and the air concentration of iodine-131 in air are obtained from other conceptual models, as discussed in detail in Sections 2.1.1, 2.1.2, and 2.1.3, respectively.

2.1.1 Concentration in Food

As shown in Figure 2.2, concentrations of iodine-131 in food can be divided into two categories: concentration in plant products and concentration in animal food products (such as milk, eggs, and meat). For both groups, a holdup time is present between the harvest or production of the food and its consumption, during which time the concentration of iodine-131 is reduced by radiological decay. Additionally, for many plant products, food preparation (washing or peeling) may also reduce iodine-131 concentrations. Food products reach the consumer through two different routes, commercial distribution systems (for example, local food markets) and home produce (such as backyard gardens and cows).

The concentration in plant products for both human and animal consumption is the quantity of iodine-131 in the plant divided by its biomass or weight. (The biomass is modeled by a modified cosine function that takes into account the potential minimum and maximum biomass of the product and its growth and senescence, or withering, rates.)

The concentration of iodine-131 in an animal product comes from four different pathways: the plants the animal eats, the water it drinks, the soil it consumes while grazing, and the deposition of iodine-131 on its harvested feed (such as bales of hay stored outdoors). The sum across these four pathways is then multiplied by a feed-to-product transfer factor to obtain the concentration in the animal product. This feed-to-product transfer factor is the ratio of the quantity of iodine-131 found in the animal product to the quantity of iodine-131 consumed by the animal. For each pathway, the quantity of iodine is the multiple of the amount of iodine-131 consumed and the concentration of iodine in that item. In the case of plant product consumption, the holdup time between harvest and ingestion must also be considered.

2.1.2 Quantity of Iodine-131 in Plants and Soil

As Figure 2.3 illustrates, determining the quantity of iodine-131 in plants and soil is a complex process. The conceptual model for determining these parameters involves a set of four linked compartments: the quantity of iodine-131 in the inner vegetation of a plant, quantity in the outer vegetation, the quantity in the upper layer of the soil, and the quantity in the root zone.

The whole process begins with the surface deposition of iodine-131. This deposition can either be intercepted by the outer vegetation on the plant or be deposited on the upper layer of soil. If iodine-131 is intercepted by the outer vegetation, the iodine-131 can move, or translocate, to the inner vegetation; can weather off onto the upper layer of soil; or, if the plant is withering, can senesce to the upper soil layer. If the iodine-131 is in the inner vegetation, and the plant is withering, it can also senesce to the upper soil layer. If the iodine-131 lands on the soil, it can move to outer vegetation either by resuspension (for example, movement in dust) or rainsplash. It can also move to the root zone soil through percolation. After the iodine-131 reaches the root zone, it can move either to

Figure 2.2. Conceptual Model of the Concentration in Food and Biomass of Plant Product

Figure 2.3. Conceptual Model of the Quantity of Iodine-131 in Plants and Soil

the inner vegetation through root uptake or leach into the deep soil, where it is no longer available for uptake by the plant or consumption by animals. The quantity of iodine-131 in each of these compartments is also reduced through radiological decay.

Many of the processes that govern the movement of iodine-131 between compartments are functions of numerous input parameters and the plant's current biomass and growth rate, which come from the biomass model.

2.1.3 Surface Deposition and Air Concentration of Iodine-131

The rate that iodine-131 is deposited on the ground at any location within the HEDR domain is the surface deposition, which depends on a variety of factors (see Figure 2.4). The surface deposition is categorized as either wet deposition or dry deposition, both of which are a function of the air concentration of iodine-131. Dry deposition is governed by the deposition velocity, which is a function of wind speed, surface roughness (trees or bare ground), species of the iodine (organic, inorganic, or particulate), and stability. Stability is determined by wind speed, cloud cover, and time of day. Wet deposition of particles is governed by the washout coefficient, which takes into account the precipitation rate, precipitation type (such as rain or snow), and species of iodine-131. Wet deposition of gaseous forms of iodine-131 is a function of wet deposition velocity which is a function of solubility, precipitation type, precipitation rate, and temperature.

The air concentration of iodine-131 at any location is a function of the amount of iodine originally released at the source (source term) less the amount that has already been deposited (puff depletion) or decayed; the amount of iodine transported to that location (puff transport); and the diffusion coefficients (a function of wind speed, surface roughness, stability, and mixing depth).

2.1.4 Puff Transport—Transport of Iodine from the Source to a Given Location

As shown in Figure 2.5, puff transport is determined by the height of the air plume released from the stacks at the Hanford site, and the wind field during transit from the stack. The individual input parameters required to determine these calculated values are shown in the figure and are the same input values used in Figure 2.4.

2.1.5 Iodine-131 Releases from the Hanford Site (Source Term)

Iodine-131 was released during chemical separations processing at the B and T Plants at Hanford. The amount released depended on how much iodine was in the fuel elements, how long it took to dissolve these fuel elements, the release rate from the separations process, and the fraction of that release that exited to the atmosphere. These relationships are illustrated in Figure 2.6.

The iodine content at the time fuel elements were discharged from the reactor during refueling (called the *push* in the terminology of the era) is a function of the total iodine in all the fuel elements in the reactor (the pile) and the peaking factor for the given push. The amount of iodine produced within the reactor pile is a function of the reactor's thermal power and the length of time the reactor was operating at that power level. The calculation of the peaking factor for a given push uses both the reactor's thermal power and the operating time at that power level, as well as the total tonnage of fuel elements in the reactor, the maximum temperature reached within the process tubes that held the fuel elements, and the temperature of the water which cooled the reactor.

Figure 2.4. Conceptual Model of the Surface Deposition of Iodine-131

Figure 2.5. Conceptual Model of the Puff Transport

Figure 2.6. Conceptual Model of the Source Term for Iodine-131

Because iodine-131 decays with a half-life of 8 days, the amount of iodine that was originally in the fuel elements is adjusted for cooling time, which is the number of days between the time when the fuel elements were pushed out of the reactor and when they were placed in the dissolver and iodine-131 was released into the atmosphere.

2.2 Overview of the Hanford Environmental Dose Reconstruction Integrated Codes

The conceptual models discussed in Section 2.1 were used as the basis for developing the computational models used in the HEDR Project. The computations are performed by HEDRIC, a series of five integrated computer models, each of which feeds information to the next level. Using HEDRIC, scientists can estimate the iodine-131 dose that a specified individual may have received from operations at Hanford. Each code is separate, but the codes interface in such a way that information is not lost as the output is passed from one module to the next.

Figure 2.7 provides a broad view of the interactions between the major components of HEDRIC. As the figure indicates, the **Reactor Model (RM)** calculates the iodine-131 present in each discharge of fuel from the operating reactors at the Hanford Site. The **Source Term Release Model (STRM)** uses this information, as well as other inputs, to calculate on an hourly basis the release of iodine-131 to the atmosphere. The **Regional Atmospheric Transport Code for Hanford Emission Tracking (RATCHET)** calculates the concentration of iodine-131 in the air and the amount deposited on the ground within the HEDR domain. The **Dynamic ESTimates of Concentrations of Radio-nuclides in Terrestrial EnvironmentS (DESCARTES)** uses information from the previous models to determine the concentration of iodine-131 that entered food products consumed by individuals living in the study area. Finally, the **Calculation of Individual Doses from Environmental Radionuclides (CIDER)** calculates the total dose of iodine-131 that an individual may have received.

The final versions of DESCARTES and CIDER are currently being coded. For the uncertainty and sensitivity analyses described in this plan, a special code, **PILOT**, has been developed. The PILOT code uses the same computational algorithms as DESCARTES and CIDER. Thus, any results from PILOT will be identical to the results from the final HEDRIC codes.

Figure 2.7. Components of Hanford Environmental Dose Reconstruction Integrated Codes

3.0 Uncertainty Analysis Approaches

This chapter reviews the uncertainty analysis approaches that were considered by HEDR Project staff for use on HEDR Project models. Two approaches for estimating uncertainty were considered and compared: mathematical and Monte Carlo techniques. Among mathematical techniques, two were investigated: analytical analysis of additive and multiplicative models, and differential analysis of more complicated models. These discussions and comparisons are preceded by discussions of concepts and procedures that are needed for implementing the uncertainty analysis approaches. Although numerous examples are provided to aid the reader, some of the mathematical equations presented in this chapter are fairly complex. These equations are provided to document that project staff are fully cognizant of the complexities involved in using these methods. Those readers who do not wish to verify the accuracy of the equations may want to focus on the examples and the interpretation of results.

The sections of this chapter are as follows:

- Section 3.1—Discussion of the first two steps in the uncertainty and sensitivity analysis process—identification of the assessment question and the model.
- Section 3.2—Discussion of the third step in the uncertainty and sensitivity analysis process—methods for determining the values of parameters.
- Section 3.3—Discussion of some mathematical approaches available for estimating the uncertainty in dose prediction, the fourth step in the uncertainty and sensitivity analysis process.
- Section 3.4—Discussion of the Monte Carlo technique for estimating of the uncertainty in dose prediction.
- Section 3.5—Discussion of the fifth step in the uncertainty and sensitivity analysis process—the quantitative description of the uncertainty in the dose prediction.
- Section 3.6—Comparison of the mathematical and Monte Carlo approaches for the estimation of the uncertainty in the dose prediction.

The reader is referred to Appendix A for definitions of statistical terminology used in Section 3.0.

3.1 Identification of Assessment Question and Model

The first step in the uncertainty and sensitivity analysis process is to identify the problem. What is the assessment question that is to be answered? The two distinct types of assessment questions are: the one with a deterministic answer and the one with a stochastic answer.

Examples of the first type of assessment question are:

- What is the annual dose received by a specific individual?
- What is the daily dose received by a reference individual of a specified population group?

Each of these questions has only one true answer. All the uncertainty in the deterministic answer is due to lack of knowledge about items that are invariant with respect to the reference unit of the assessment question IAEA (1989). That is, all the items which contribute to the individual's dose also have only one true, but unknown, value (for example, what the individual ate, how much the individual ate, or the concentration of the radionuclide in the food the individual ate).

An example of the second type of assessment question (a question with a stochastic answer) is:

- What is the annual dose received per individual in a specified population group?

The true answer to this question varies from individual to individual. Part of the uncertainty in the answer is due to stochastic variability with respect to the reference unit of the assessment question (IAEA 1989). For example, what and how much an individual ate varies from individual to individual. The answer may also contain uncertainties due to lack of knowledge about items that are invariant to the reference unit of the assessment question. For example, the concentration in the food does not vary from individual to individual; it still has one true, but unknown, value.

The second step in the uncertainty and sensitivity analysis process is to specify the model that will be used to answer the assessment question. First, a conceptual model is developed that identifies "the number of distinct compartments, interconnections between compartments and the number of processes and mechanisms to be considered explicitly. . . . The computational model is a tool that is used to produce a quantitative result. [It] is a set of equations and parameter values" (IAEA 1989). The conceptual and computational models are another source of uncertainty due to lack of knowledge. The models may not contain all the relevant processes and the mathematical equations may not completely or correctly describe the conceptual process.

The concepts of deterministic and stochastic answers are illustrated in the following four examples. Each example asks an assessment question about the dose to thyroid from ingestion of cow's milk containing iodine-131. Conceptually, the model for predicting the dose to thyroid is simple: dose is equal to the dose factor times the amount of milk ingested times the concentration of iodine-131 in the milk. However, as the assessment question is changed, the precise definitions of the predicted dose and each of the parameters are also changed in each example.

3.1.1 Assessment Question 1

What was the dose to the thyroid of a specific individual, located in Richland, Washington, from cow's milk containing iodine-131 on June 1, 1945?

The assessment question has a deterministic answer because the specified individual received only one true dose.

Example 1: Computational Model 1: Assessment Question/Deterministic Answer

$$Dose_{milk} = DoseFactor_{ingestion} \times Intake_{milk} \times Concentration_{milk}$$

where

$Dose_{milk}$ = dose to thyroid of the specified individual, located in Richland, Washington from cow's milk contaminated with iodine-131, on June 1, 1945 (rad)

$DoseFactor_{ingestion}$ = the conversion factor which converts the Curies (Ci) of iodine-131 that is ingested into the dose to the thyroid (rad)

$Intake_{milk}$ = amount of cow's milk consumed by the specific individual on June 1, 1945 (liters)

$Concentration_{milk}$ = concentration of iodine-131 in cow's milk in Richland, Washington on June 1, 1945 (Ci / liter).

The parameters in this model, $DoseFactor_{ingestion}$, $Intake_{milk}$, and $Concentration_{milk}$, have one true value, but there is considerable uncertainty about what those true parameter values were:

The value of the parameter $DoseFactor_{ingestion}$ is based on the results of another mathematical model which takes into consideration the thyroid mass, the fraction of the circulating iodine removed from the bloodstream (which is dependent on the total amount of iodine in circulation), and the half-time over which the iodine is stored in the thyroid (Dunning and Schwarz 1981). These factors are different for each individual and are difficult to measure in an individual who was exposed to iodine-131. The HEDR project is concerned with doses to individuals that were received over 40 years ago and dose factors have changed for those individuals over time.

The value of $Intake_{milk}$ is uncertain because it is based on memory recall of what an individual consumed over 40 years ago.

The value of $Concentration_{milk}$ is also based on the results of a mathematical model because no monitoring data were collected for milk at the time of the high iodine-131 releases from Hanford. The parameters in this model include the intake by the cow of numerous plant products and soil, the concentrations of iodine-131 in each of these items, and the feed-to-milk transfer factor for iodine-131. Each of these parameters is not well-known because of the large number of years since the events occurred, variability in the experimental results used to estimate them, or the fact that they are the results of other mathematical models.

The uncertainty in these parameters is characterized by specifying a probability density function (PDF) for each parameter. The PDF for each parameter in Example 1 expresses the relative likelihood of alternative values for the parameter for the specific individual on June 1, 1945 in Richland, Washington. (The types of PDFs being used by the HEDR Project are described in Appendix B.)

3.1.2 Assessment Question 2

What was the total dose to the thyroid of a specific individual, located in Richland, Washington from cow's milk containing iodine-131 in June 1945?

The assessment question has a deterministic answer, because the specified individual received only one true dose.

Example 2: Computational Model 2: Assessment Question/Deterministic Answer

$$Dose_{milk} = DoseFactor_{ingestion} \times \sum_{day=1}^{30} (Intake_{milk,day} \times Concentration_{milk,day})$$

where

$Dose_{milk}$ = total dose to thyroid for the specific individual, located in Richland, Washington, from cow's milk contaminated with iodine-131, in June 1945 (rad)

$DoseFactor_{ingestion}$ = the conversion factor which converts the Curies (Ci) of iodine-131 that is ingested into the dose to the thyroid (rad)

$Intake_{milk,day}$ = amount of cow's milk consumed by the specific individual on a given day in June 1945 (liters)

$Concentration_{milk,day}$ = concentration of iodine-131 in cow's milk in Richland, Washington on a given day in June 1945 (Ci / liter).

Again, each of the parameters in this model, $DoseFactor_{ingestion}$, $Intake_{milk,day}$, and $Concentration_{milk,day}$, have one true (uncertain) value. However, now there are 30 intake parameters and 30 concentration parameters because the assessment question concerns the entire month of June. This model assumes the $DoseFactor_{ingestion}$ value is constant over 30 days.

It might be much easier to build a computational model that uses monthly totals and/or averages instead of 30 daily values, as in Example 3.

Example 3: Computational Model 3: Assessment Question/Deterministic Answer

$$Dose_{milk} = DoseFactor_{ingestion} \times Intake_{milk,total} \times Concentration_{milk,average}$$

where

$Dose_{milk}$ = total dose to thyroid for the specific individual, located in Richland, Washington from cow's milk contaminated with iodine-131, in June 1945 (rad)

$DoseFactor_{ingestion}$ = the conversion factor which converts the Curies (Ci) of iodine-131 that is ingested into the dose to the thyroid (rad)

$Intake_{milk,total}$ = total amount of cow's milk consumed by the specific individual in June 1945 (liters)

$Concentration_{milk,average}$ = average concentration of iodine-131 in cow's milk in Richland, Washington in June 1945 (Ci / liter).

The second computational model in Example 3 can be rewritten as

$$Dose_{milk} = DoseFactor_{ingestion} \times \sum_{day=1}^{30} Intake_{milk,day} \times \frac{1}{30} \sum_{day=1}^{30} Concentration_{milk,day}$$

This computational model will only provide the same result as the first computational model for this assessment question (Example 2) if intake and concentration are constant for June 1945.

3.1.3 Assessment Question 3

What is the dose to the thyroid of a reference adult male, located in Richland, Washington from cow's milk contaminated with iodine-131 on June 1, 1945?

The reference adult male, located in Richland, Washington is intended to be a composite individual. The assessment question is still deterministic because the reference adult male received only one true dose.

Example 4: Computational Model 4: Assessment Question/Deterministic Answer

$$Dose_{milk} = DoseFactor_{ingestion} \times Intake_{milk} \times Concentration_{milk}$$

where

$Dose_{milk}$ = dose to thyroid for the reference adult male, located in Richland, Washington from cow's milk contaminated with iodine-131, on June 1, 1945 (rad)

$DoseFactor_{ingestion}$ = the conversion factor which converts the Curies (Ci) of iodine-131 that is ingested into the dose to the thyroid (rad)

$Intake_{milk}$ = amount of cow's milk consumed by the reference adult male, located in Richland, Washington on June 1, 1945 (liters)

$Concentration_{milk}$ = concentration of iodine-131 in cow's milk in Richland, Washington on June 1, 1945 (Ci / liter).

The parameters in this model, $DoseFactor_{ingestion}$, $Intake_{milk}$, and $Concentration_{milk}$, also have one true, but uncertain, value. The PDFs for $DoseFactor$ and $Intake$ should now express the relative likelihoods of the alternative values for each of these parameters for the reference adult male, on June 1, 1945 in Richland, Washington. The shapes of the PDFs are dependent on how broad a description of the reference man is used. For example, the intake by the reference adult male could simply be set to a constant of 0.5 liter (no PDF), or it could be very broadly defined with a range of 0 to 1.5 liters. The PDF for concentration is the same PDF used in Example 1 because whether the dose is for a specific individual or an reference adult male, the uncertainty about the alternative values for concentration of iodine-131 in cow's milk in Richland, Washington on June 1, 1945 is the same.

3.1.4 Assessment Question 4

What was the dose to the thyroid per adult male, located in Richland, Washington, from cow's milk, contaminated with iodine-131 on June 1, 1945?

Now the assessment question has a stochastic answer because dose varies from individual to individual. "Since it is impractical to model each [adult male in Richland, Washington on June 1, 1945] explicitly, the dose per individual is treated as a random variable. The question is thus answered probabilistically in the form of a distribution of doses per [adult male in Richland, Washington on June 1, 1945]" (IAEA 1989).

Example 5: Computational Model 5: Assessment Question/Stochastic Answer

$$Dose_{milk} = DoseFactor_{ingestion} \times Intake_{milk} \times Concentration_{milk}$$

where

$Dose_{milk}$ = dose to the thyroid per adult male, located in Richland, Washington, from cow's milk, contaminated with iodine-131, on June 1, 1945 (rad)

$DoseFactor_{ingestion}$ = the conversion factor which converts the Curies (Ci) of iodine-131 that is ingested into the dose to the thyroid (rad)

$Intake_{milk}$ = amount of milk consumed on June 1, 1945, per adult male in Richland, Washington (liters)

$Concentration_{milk}$ = concentration of iodine-131 in milk on June 1, 1945 (Ci / liter).

The $DoseFactor_{ingestion}$ and $Intake_{milk}$ parameters are now random variables. They describe the stochastic variability of these parameters within the population of adult males in Richland, Washington on June 1, 1945. The PDFs express the relative proportion of the population with alternative values of each parameter.

The uncertainty in the concentration of iodine-131 in milk on June 1, 1945 $Concentration_{milk}$, is still due to lack of knowledge, as in the previous examples. The PDF for $Concentration_{milk}$ expresses the relative likelihood of the alternative values for concentration in cow's milk that each individual could have consumed.

The dose that is calculated from this model (Example 5) is a series of distributions of doses per adult male, located in Richland, Washington on June 1, 1945 one distribution for each alternative value of the concentration of iodine-131 in milk on June 1, 1945.

3.2 Strategy for Specifying Parameter Uncertainties

The third step in the uncertainty and sensitivity analysis process is to describe the uncertainties in the model parameters with PDFs. This section describes the process being used by the HEDR Project to assess and specify the uncertainty in model parameters.

Specifying parameter uncertainties is important because the HEDR Project is planning to use computer simulation methods to quantify dose uncertainties for important exposure pathways, such as for doses attributable to atmospheric releases of radionuclides from Hanford-Site operations, described in Section 2. For the air pathway, the computer simulation method consists of repeating calculations of the dose computer model HEDRIC 100 times to generate 100 predictions of dose for a specific individual. Each time the model calculations are repeated, a new value for each uncertain parameter in HEDRIC is randomly generated from its specified PDF (PDFs are shown in Appendix B) which express the uncertainty of the parameter. Thus, parameter uncertainties, as modeled by the PDFs, have a direct impact on the variability (spread or uncertainty) exhibited by the 100 estimates of dose computed by HEDRIC.

Because of the relationship between parameter uncertainties and dose uncertainties, the methods used to specify parameter PDFs must be well understood and documented. Snyder et al. (1992) document the PDFs that are proposed for parameters in the environmental accumulation and dose portions of the HEDRIC code. The strategy used by Snyder et al. (1992) to specify the PDFs is detailed in following paragraphs.

3.2.1 Estimating the Maximum Conceivable Parameter Range

The uncertainty of a parameter is expressed in two ways: the range of possible values and the frequency with which any value within that range is expected to occur. Both the range and frequency define the PDF for each parameter.

The maximum conceivable range of a parameter is estimated by first searching Hanford-originated and general scientific literature for relevant reports, books, and scientific papers that contain parameter information. This search has the following components:

- conduct computerized searches of the radiation protection literature citation databases
- review relevant reports, papers, and books that originated at Hanford or were authored by acknowledged Hanford or non-Hanford experts in the subject area of interest
- review recent issues of scientific journals for relevant papers and information
- review pertinent references found in relevant reports, papers, and books and those recommended by experts.

After the literature search is completed, the maximum and minimum parameter values found in the literature are determined. Then, either self assessment or informal solicitation (defined in a following paragraph) of expert opinion is used to evaluate whether these maximum and minimum values should be changed to reflect physical limitations on parameter values and conditions present in the HEDR study region.

3.2.2 Obtaining Subjective Information

Specifying the maximum and minimum value of the parameter applicable to Hanford does not completely determine the uncertainty of that parameter. In particular, the likelihood that various values between the maximum and minimum could have been the true value for Hanford must be specified. That is, the PDF of each uncertain parameter, a largely subjective process, must be determined.

Subjective determination of the PDF for most parameters is necessary because it is uncommon to have enough data to evaluate statistically which PDF should be selected. When the conceivable maximum and minimum for the parameter's value are determined, one or more of the following subjective methods [from Roberds (1990)] are used to develop additional information about the parameter so that a PDF can be specified:

- self-assessment

The analyst interprets the available information and then quantifies an assessment of the likely value of the parameter and its uncertainty (Good 1965, von Holstein 1979). The rationale behind the assessment should be documented, including a description and

evaluation of the available information to enhance defensibility of such subjective probability assessments. This method is simple, but has significant limitations:

- poor quantification of uncertainty
 - uncorrected biases or unspecified assumptions or both, possibly in spite of documentation
 - imprecision
 - lack of credibility, if the analyst cannot be considered a subject-matter expert in the technical field.
- informal solicitation of expert opinion

The analyst asks an expert to interpret the available information and quantify an assessment of the likely parameter value and its uncertainty (Morgan et al. 1979, Bernreuter 1980). The defensibility of such assessments is increased from that of self-assessment techniques, primarily because of the increased credibility of the expert involved. The expert's rationale for the assessment should be well-documented, including a description of the information available to the expert and the expert's evaluation of that information.

Although generally an improvement over self-assessment techniques, informal solicitation of expert opinion has similar significant limitations, increased cost and potentially poor problem definition.

- probability encoding

Probability encoding (Spetzler and von Holstein 1972, Zamora 1975, von Holstein and Matheson 1979, and Merkhofer and McNamee 1982) is the systematic and defensible approach to developing individual subjective probability assessments, but it is also the most expensive. The analyst, trained in probability theory, elicits, in a proper and consistent manner, a technical expert's assessment of the PDF of a parameter value, which expresses that expert's uncertainty in the value in quantified terms. This is done in a formalized way in five steps:

- 1) motivating
- 2) structuring
- 3) conditioning
- 4) encoding
- 5) verifying.

Although a general improvement over other available methods mitigates most of the potential problems), some imprecision may remain. Also, probability encoding is relatively costly because it is labor intensive.

The HEDR Project is using primarily self-assessment, informal solicitation of expert opinion, and information from the scientific literature to determine the most appropriate PDF. Probability encoding is the most systematic and defensible approach for developing additional subjective information needed to specify a particular PDF for a parameter, but it is also the most expensive. Because of this expense, probability encoding is used only for a limited set of parameters. The critically important parameters, those for which probability encoding will be used, are determined on the basis of the results from sensitivity analyses.

3.2.3 Selecting a Probability Density Function

The six PDFs being used by HEDR Project staff (uniform, piecewise uniform, loguniform, triangular, normal, and lognormal PDFs) are shown in Appendix B of this report. The following criteria are being used to select one of these distributions:

- The uniform PDF, which assigns equal probability to each possible value of the parameter between the minimum and maximum values, is selected if minimal information (actual measurements or subjective opinions of experts) is available about the parameter. In other words, the uniform PDF is used as a default PDF and is used whenever it is impossible to defend the assertion that a more complex distribution is appropriate.
- The piecewise uniform PDF describes the situation where the range of the parameter (maximum value minus minimum value) can be divided into segments or pieces. For each segment, a different level of equal probability is assigned to all values of the parameter within the segment. This PDF is selected when data are sufficient to divide the range into distinct segments, but inadequate to defend unequal probabilities within segments.
- The loguniform PDF, which assigns equal probability of occurrence to the logarithm of each possible value of the parameter, is selected if 1) the minimum and maximum values span a distance of several orders of magnitude, 2) it appears that a uniform distribution of the logarithms is a plausible model, or 3) insufficient information is available to defend the assertion that a more complex distribution is appropriate.
- The triangular PDF assigns probabilities of occurrence to the possible parameter values so the PDF is in the form of a triangle. This PDF is selected if 1) the available information indicates that one value of the parameter is more likely to occur than any other single value, and 2) insufficient information is available to defend the assertion that a PDF shape is more complex than a triangle. The most likely parameter value need not be located at the center of the range. It may also be located near or at either the maximum or minimum value.
- The normal PDF assigns probability of occurrence to the possible parameter values so the PDF is symmetric and bell-shaped. This distribution is used only when sufficient data, statistical analyses, or other information exist to defend the assertion that normal distribution is a more appropriate model than a symmetric triangular distribution.
- The lognormal PDF assigns probabilities of occurrence to the logarithms of the possible parameter values so the PDF of those logarithms is symmetric and bell-shaped. The lognormal PDF is used only when sufficient data, statistical analyses, or other information exist to defend the assertion that normal distribution of the logarithms is a more appropriate model than a symmetric triangular distribution of the logarithms.

Along with specifying the PDF of each model parameter, it is necessary to specify the relationships and dependencies among parameters. These relationships are modeled by HEDR Project staff as correlation coefficients or mathematical functions. These correlations and functions are determined on the basis of information obtained from literature searches and the elicitation of expert opinion. However, it is frequently the case that very little quantitative information about correlations or functional relationships is available from the literature. In that case, self assessment and informal solicitation of expert opinion are relied upon to subjectively derive the needed relationships. If sensitivity analyses indicate the added expense is justified, probability encoding is used for the most critical relationships.

3.2.4 Parameter Selection Frequency

In addition to selecting a PDF for each model parameter, a decision must be made about how often parameter PDFs should change over time and space. For example, for each of the 100 realizations of the HEDRIC code, the environmental concentrations of iodine-131 (dose in the soil, plants, and animal products) are computed at many different locations in the study area for each day over a 5-year period. The PDF specified for a given parameter may or may not be suitable for all times and locations. The selection of PDFs for the same parameter at different times and places is made on the basis of literature reviews, self assessment, or informal solicitation of expert opinion. In general, the PDF for a given parameter is not changed over space and time without a compelling reason.

Even though the PDF of a parameter is unchanged, it must still be decided how often a new random value is selected from that PDF. For example, as discussed by Snyder et al. (1992), the ratio of the indoor air concentration relative to the outdoor air concentration can be expected to vary on a daily basis. Thus, although the same PDF is used for all days, a new random value from the PDF is selected each day. In contrast, the maximum potential biomass of a plant would only change once each year.

3.3 Estimation of Uncertainty Using Mathematical Techniques

The fourth step in the uncertainty and sensitivity analysis process is to estimate the uncertainty in the model's prediction. This report discusses two mathematical techniques and the Monte Carlo method for estimating this uncertainty. In Section 3.3.1, analytical analysis of additive and multiplicative models is discussed. In Section 3.3.2, differential analysis of general models is discussed. The techniques described in Section 3.3.2 were used in the uncertainty analysis of the dose in Stevens et al. 1992. The estimation of uncertainty in model predictions using Monte Carlo methods is discussed in Section 3.4.

3.3.1 Analytical Analysis of the Model

In this section, additive and multiplicative models are discussed. For these simple models, the mean and variance can be calculated exactly, if each component in the equation is known. The primary source for this section is Mood et al. (1974).

3.3.1.1 Additive Model

The additive model is

$$Y = \sum_{i=1}^n X_i = X_1 + X_2 + X_3 + \cdots + X_n \quad (3.1)$$

The mean and variance for the additive model are

$$E[Y] = \sum_{i=1}^n E[X_i] \quad (3.2)$$

and

$$V[Y] = \sum_{i=1}^n V[X_i] + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n COV[X_i, X_j] \quad (3.3)$$

Example 6: Additive Model

$$Dose_{total} = Dose_{milk\ ingestion} + Dose_{leafy\ vegetable\ ingestion} + Dose_{inhalation}.$$

The mean and variance are

$$E[Dose_{total}] = E[Dose_{milk\ ingestion}] + E[Dose_{leafy\ vegetable\ ingestion}] + E[Dose_{inhalation}]$$

and

$$\begin{aligned} V[Dose_{total}] = & V[Dose_{milk\ ingestion}] + V[Dose_{leafy\ vegetable\ ingestion}] + V[Dose_{inhalation}] \\ & + 2COV[Dose_{milk\ ingestion}, Dose_{leafy\ vegetable\ ingestion}] \\ & + 2COV[Dose_{milk\ ingestion}, Dose_{inhalation}] \\ & + 2COV[Dose_{leafy\ vegetable\ ingestion}, Dose_{inhalation}] \end{aligned}$$

If the doses from milk ingestion, leafy vegetable ingestion, and inhalation are uncorrelated, the covariance terms in the model are zero. However, these doses are correlated (see Example 1, Section 3.3.2).

3.3.1.2 Multiplicative Model

In this section, only 2- and 3-parameter multiplicative models will be presented. As more parameters are included in the multiplicative model, the equations become quite cumbersome.

Two-Parameter Multiplicative Model

$$Y = X_1 X_2. \quad (3.4)$$

The mean and variance for this model are

$$E[Y] = E[X_1]E[X_2] + COV[X_1, X_2] \quad (3.5)$$

and

$$\begin{aligned} V[Y] = & (E[X_1])^2 V[X_2] + (E[X_2])^2 V[X_1] + 2COV[X_1, X_2] - (COV[X_1, X_2])^2 \\ & + E[(X_1 - E[X_1])^2 (X_2 - E[X_2])^2] + 2E[X_2]E[(X_1 - E[X_1])^2 (X_2 - E[X_2])] \\ & + 2E[X_1]E[(X_1 - E[X_1])(X_2 - E[X_2])^2] \end{aligned} \quad (3.6)$$

If X_1 and X_2 are uncorrelated, then the terms with $COV[X_1, X_2]$ in Equations (3.5) and (3.6) become zero. The stronger assumption, that X_1 and X_2 are independent, is needed to reduce Equation (3.6) for the variance to a manageable form:

$$V[Y] = (E[X_1])^2 V[X_2] + (E[X_2])^2 V[X_1] + V[X_1]V[X_2] \quad (3.7)$$

Three Parameter Multiplicative Model

$$Y = X_1 X_2 X_3. \quad (3.8)$$

The mean and variance for this model are

$$\begin{aligned} E[Y] = & E[X_1]E[X_2]E[X_3] + E[X_1]COV[X_2, X_3] + E[X_2]COV[X_1, X_3] \\ & + E[X_3]COV[X_1, X_2] + E[(X_1 - E[X_1])(X_2 - E[X_2])(X_3 - E[X_3])] \end{aligned} \quad (3.9)$$

and

$$\begin{aligned} V[Y] = & E[X_1]^2 E[X_2]^2 V[X_3] + E[X_1]^2 E[X_3]^2 V[X_2] + E[X_2]^2 E[X_3]^2 V[X_1] \\ & + 2E[X_1]E[X_2]E[X_3]^2 COV[X_1, X_2] + 2E[X_1]E[X_2]^2 E[X_3] COV[X_1, X_3] \\ & + 2E[X_1]^2 E[X_2]E[X_3] COV[X_2, X_3] + 17 \text{ other terms} \end{aligned} \quad (3.10)$$

If X_1 , X_2 , and X_3 are not independent, then the 17 other terms are generally ignored in the above equation. Consequently, the estimate of the variance is only approximate and could deviate considerably from the correct answer.

If X_1 , X_2 , and X_3 are independent then the mean and variance become

$$E[Y] = E[X_1]E[X_2]E[X_3] \quad (3.11)$$

and

$$\begin{aligned} V[Y] = & E[X_1]^2 E[X_2]^2 V[X_3] + E[X_1]^2 E[X_3]^2 V[X_2] + E[X_2]^2 E[X_3]^2 V[X_1] \\ & + E[X_1]^2 V[X_2]V[X_3] + E[X_2]^2 V[X_1]V[X_3] + E[X_3]^2 V[X_1]V[X_2] \\ & + V[X_1]V[X_2]V[X_3] \end{aligned} \quad (3.12)$$

Example 7: Multiplicative Model (with Independent Parameters)

$$Dose_{milk\ ingestion} = DoseFactor_{ingestion} \times Intake_{milk} \times Concentration_{milk},$$

where

$$E[DoseFactor] = 1.78 \text{ rad}_{thyroid} / \mu Ci_{ingested}$$

$$V[DoseFactor] = (1.37 \text{ rad}_{thyroid} / \mu Ci_{ingested})^2$$

$$E[Intake_{milk}] = 100 \text{ L}_{milk\ ingested}$$

$$V[Intake_{milk}] = (40 \text{ L}_{milk\ ingested})^2$$

$$E[Concentration_{milk}] = 0.00673 \mu Ci_{ingested} / L_{milk\ ingested}$$

$$V[Concentration_{milk}] = (0.0061 \mu Ci_{ingested} / L_{milk\ ingested})^2$$

$$E[Holdup] = 2.0 \text{ days}$$

$$V[Holdup] = \frac{1}{3} \text{ days}^2$$

$$\lambda_{rad} = 0.086 / \text{day}$$

Example 7: (contd)

then the mean, variance, and standard deviation are

$$E[Dose_{milk\ ingestion}] = 1.78 \times 100 \times 0.00673 = 1.198 \text{ rad}_{thyroid}$$

$$\begin{aligned} V[Dose_{milk\ ingestion}] &= (1.78)^2(100)^2(0.00611)^2 + (1.78)^2(0.00673)^2(40)^2 \\ &\quad + (100)^2(0.00673)^2(1.37)^2 + (1.78)^2(40)^2(0.00611)^2 \\ &\quad + (100)^2(1.37)^2(0.00611)^2 + (0.00673)^2(1.37)^2(40)^2 \\ &\quad + (1.37)^2(40)^2(0.00611)^2 \\ &= 1.1828 + 0.2296 + 0.8501 + 0.1893 \\ &\quad + 0.7007 + 0.1360 + 0.1121 \\ &= 3.401 (\text{rad}_{thyroid})^2 \end{aligned}$$

$$SD[Dose_{milk\ ingestion}] = 1.844 \text{ rad}_{thyroid}$$

As seen in the Example 7, for a simple model with independent parameters, this method is easy to implement. However, the model used in this example has been simplified by deleting a parameter, namely, the holdup time of milk (the time between when the cow produced the milk and the time the person drank the milk). The more complex model is used in Example 8 in Section 3.3.2.

3.3.2 Differential Analysis of the Model

Differential analysis of the model is a two-step process. First, a Taylor series expansion of the model is calculated. Then, the appropriate expectations of the Taylor series expansion are calculated. The primary sources for this section are Taylor and Mann (1983) and Mood et al. (1974). The reader is referred to any calculus and introductory theoretical statistics book for additional information on Taylor series expansion and expectations of Taylor series expansions.

In this section, the entire process is illustrated with a 1-parameter model. The approximations for general models are given and Examples 6 and 7 are continued.

One-Parameter Taylor Series Expansion

Let $g(X)$ denote a model with one random parameter, X . The Taylor series expansion of $g(X)$ about $X = \mu_x$ is defined as

$$\begin{aligned} g(X) &= g(\mu_x) + g^{(1)}(\mu_x)(X - \mu_x) + \frac{g^{(2)}(\mu_x)(X - \mu_x)^2}{2!} \\ &\quad + \dots + \frac{g^{(n)}(\mu_x)(X - \mu_x)^n}{n!} + R_n \end{aligned} \quad (3.13)$$

where

$$g^{(i)}(\mu_x) = \left. \frac{d^i g(X)}{dX^i} \right|_{X=\mu_x}$$

$$n! = n \times (n-1) \times (n-2) \times (n-3) \times \cdots \times 3 \times 2 \times 1$$

and R_n is the remainder. Generally, only the first three terms of the Taylor series expansion are retained. Thus, the results will only be an approximation, unless the remainder is zero.

Expectations of the Taylor Series Expansion of a 1-Parameter Model

The expected value (mean) of $g(X)$ is approximated by taking the expected value of each of the first three terms in the Taylor series expansion shown in Equation (3.13)

$$\begin{aligned} E[g(X)] &\approx E[g(\mu_x)] + E[g^{(1)}(\mu_x)(X - \mu_x)] + E\left[\frac{g^{(2)}(\mu_x)(X - \mu_x)^2}{2!}\right] \\ &\approx g(\mu_x) + \frac{g^{(2)}(\mu_x)V[X]}{2} \end{aligned} \quad (3.14)$$

The expected value of the second term in Equation (3.14) is zero. The variance of $g(X)$ is derived by using the fact that

$$V[g(X)] = E[\{g(X)\}^2] - (E[g(X)])^2 \quad (3.15)$$

where

$$\begin{aligned} \{g(X)\}^2 &\approx \{g(\mu_x)\}^2 + \{g^{(1)}(\mu_x)\}^2(X - \mu_x)^2 + \frac{\{g^{(2)}(\mu_x)\}^2(X - \mu_x)^4}{4} \\ &\quad + 2g(\mu_x)g^{(1)}(\mu_x)(X - \mu_x) + g(\mu_x)g^{(2)}(\mu_x)(X - \mu_x)^2 \\ &\quad + g^{(1)}(\mu_x)g^{(2)}(\mu_x)(X - \mu_x)^3 \\ E[\{g(X)\}^2] &\approx \{g(\mu_x)\}^2 + \left(\{g^{(1)}(\mu_x)\}^2 + g(\mu_x)g^{(2)}(\mu_x)\right)V[X] \\ &\quad + g^{(1)}(\mu_x)g^{(2)}(\mu_x)E[(X - \mu_x)^3] + \frac{\{g^{(2)}(\mu_x)\}^2 E[(X - \mu_x)^4]}{4} \end{aligned}$$

and

$$(E[g(X)])^2 \approx \{g(\mu_x)\}^2 + g(\mu_x)g^{(2)}(\mu_x)V[X] + \frac{\{g^{(2)}(\mu_x)V[X]\}^2}{4}.$$

Then

$$\begin{aligned} V[g(X)] \approx & \{g^{(1)}(\mu_x)\}^2 V[X] + g^{(1)}(\mu_x)g^{(2)}(\mu_x)E[(X - \mu_x)^3] \\ & + \frac{\{g^{(2)}(\mu_x)\}^2 \left(E[(X - \mu_x)^4] - \{V[X]\}^2 \right)}{4} \end{aligned} \quad (3.16)$$

As seen above, even for a 1-parameter model, the differential method is *algebra intensive*. Also, the approximation for the variance [see Equation (3.16)] contains higher order moments, such as $E[(X - \mu_x)^3]$ and $E[(X - \mu_x)^4]$. These moments describe the skewness and kurtosis (in this case, flatness) of a PDF, respectively, and both can be large for some distributions.

Mean and Variance for a General Model Derived from Differential Analysis

For a model with k random parameters, the approximate mean and variance are:

$$\begin{aligned} E[g(\underline{X})] \approx & g(\underline{\mu_x}) + \frac{1}{2} \sum_{i=1}^k \left(\frac{\partial^2 g(\underline{X})}{\partial^2 X_i} \bigg|_{\underline{X}=\underline{\mu_x}} \right) \times V[X_i] \\ & + \sum_{i=1}^{k-1} \sum_{j=i+1}^k \left(\frac{\partial^2 g(\underline{X})}{\partial X_i \partial X_j} \bigg|_{\underline{X}=\underline{\mu_x}} \right) \times COV[X_i, X_j] \end{aligned} \quad (3.17)$$

and

$$\begin{aligned} V[g(\underline{X})] \approx & \sum_{i=1}^k \left(\frac{\partial g(\underline{X})}{\partial X_i} \bigg|_{\underline{X}=\underline{\mu_x}} \right)^2 \times V[X_i] \\ & + 2 \sum_{i=1}^{k-1} \sum_{j=i+1}^k \left(\frac{\partial g(\underline{X})}{\partial X_i} \bigg|_{\underline{X}=\underline{\mu_x}} \right) \left(\frac{\partial g(\underline{X})}{\partial X_j} \bigg|_{\underline{X}=\underline{\mu_x}} \right) \times COV[X_i, X_j] \end{aligned} \quad (3.18)$$

If all k random parameters are independent, the approximate mean and variance for the model are

$$E[g(\underline{X})] \approx g(\underline{\mu_x}) + \frac{1}{2} \sum_{i=1}^k \left(\frac{\partial^2 g(\underline{X})}{\partial^2 X_i} \bigg|_{\underline{X}=\underline{\mu_x}} \right) \times V[X_i] \quad (3.19)$$

and

$$\begin{aligned}
 V[g(\underline{X})] = & \sum_{i=1}^k \left\{ \left(\frac{\partial g(\underline{X})}{\partial X_i} \right) \bigg|_{\underline{X}=\underline{\mu}_X} \right\}^2 \times V[X_i] \\
 & + \sum_{i=1}^{k-1} \sum_{j=i+1}^k \left\{ \left(\frac{\partial^2 g(\underline{X})}{\partial X_i \partial X_j} \right) \bigg|_{\underline{X}=\underline{\mu}_X} \right\}^2 \times V[X_i] \times V[X_j] \\
 & - \frac{1}{4} \sum_{i=1}^k \left\{ \left(\frac{\partial^2 g(\underline{X})}{\partial^2 X_i} \right) \bigg|_{\underline{X}=\underline{\mu}_X} \right\}^2 \times V[X_i] \quad (3.20)
 \end{aligned}$$

Example 8: Differential Analysis (Continuation of Example 6)

The new model for total dose is

$$\begin{aligned}
 Dose_{total} = & DoseFactor_{ingestion} (Intake_{milk} \times Concentration_{milk} + Intake_{lvg} \times Concentration_{lvg}) \\
 & + DoseFactor_{inhalation} (Intake_{air} \times Concentration_{air})
 \end{aligned}$$

where it is assumed that

$$COV[DoseFactor_{ingestion}, DoseFactor_{inhalation}] \neq 0$$

$$COV[Intake_{milk}, Intake_{lvg}] \neq 0$$

$$COV[Intake_{milk}, Intake_{air}] \neq 0$$

$$COV[Intake_{lvg}, Intake_{air}] \neq 0$$

$$COV[Concentration_{milk}, Concentration_{lvg}] \neq 0$$

$$COV[Concentration_{milk}, Concentration_{air}] \neq 0$$

$$COV[Concentration_{lvg}, Concentration_{air}] \neq 0$$

and all other covariances are equal to 0.

Example 8: (contd)

Then, the first order differentials are

$$\frac{\partial Dose_{total}}{\partial DoseFactor_{ingestion}} = Intake_{milk} \times Concentration_{milk} + Intake_{lvg} \times Concentration_{lvg}$$

$$\frac{\partial Dose_{total}}{\partial DoseFactor_{inhalation}} = Intake_{air} \times Concentration_{air}$$

$$\frac{\partial Dose_{total}}{\partial Intake_{milk}} = DoseFactor_{ingestion} \times Concentration_{milk}$$

$$\frac{\partial Dose_{total}}{\partial Intake_{lvg}} = DoseFactor_{ingestion} \times Concentration_{lvg}$$

$$\frac{\partial Dose_{total}}{\partial Intake_{air}} = DoseFactor_{inhalation} \times Concentration_{air}$$

$$\frac{\partial Dose_{total}}{\partial Concentration_{milk}} = DoseFactor_{ingestion} \times Intake_{milk}$$

$$\frac{\partial Dose_{total}}{\partial Concentration_{lvg}} = DoseFactor_{ingestion} \times Intake_{lvg}$$

$$\frac{\partial Dose_{total}}{\partial Concentration_{air}} = DoseFactor_{inhalation} \times Intake_{air}$$

Fortunately, for the nonzero covariances, the second order differentials are zero.

Then, for this example, the approximate mean and variance are

$$\begin{aligned} E[Dose_{total}] = & E[DoseFactor_{ingestion}] \left(E[Intake_{milk}] E[Concentration_{milk}] \right. \\ & + E[Intake_{lvg}] E[Concentration_{lvg}] \left. \right) \\ & + E[DoseFactor_{inhalation}] E[Intake_{air}] E[Concentration_{air}] \end{aligned}$$

and

Example 8: (contd)

$$\begin{aligned}
 V[Dose_{total}] \approx & \left(E[In_{milk}]E[Conc_{milk}] + E[In_{lvg}]E[Conc_{lvg}] \right)^2 V[DF_{ing}] \\
 & + \left(E[In_{air}]E[Conc_{air}] \right)^2 V[DF_{inh}] + \left(E[DF_{ing}]E[Conc_{milk}] \right)^2 V[In_{milk}] \\
 & + \left(E[DF_{ing}]E[Conc_{milk}] \right)^2 V[In_{lvg}] + \left(E[DF_{inh}]E[Conc_{air}] \right)^2 V[In_{air}] \\
 & + \left(E[DF_{ing}]E[In_{milk}] \right)^2 V[Conc_{milk}] + \left(E[DF_{ing}]E[In_{lvg}] \right)^2 V[Conc_{lvg}] \\
 & + \left(E[DF_{inh}]E[Conc_{air}] \right)^2 V[Conc_{air}] \\
 & + 2 \left(E[In_{milk}]E[Conc_{milk}] + E[In_{lvg}]E[Conc_{lvg}] \right) \\
 & \quad \times E[In_{air}]E[Conc_{air}]COV[DF_{ing}, DF_{inh}] \\
 & + 2E[DF_{ing}]^2 E[Conc_{milk}]E[Conc_{lvg}]COV[In_{milk}, In_{lvg}] \\
 & + 2E[DF_{ing}]E[DF_{inh}]E[Conc_{milk}]E[Conc_{air}]COV[In_{milk}, In_{air}] \\
 & + 2E[DF_{ing}]E[DF_{inh}]E[Conc_{lvg}]E[Conc_{air}]COV[In_{lvg}, In_{air}] \\
 & + 2E[DF_{ing}]^2 E[In_{milk}]E[In_{lvg}]COV[Conc_{milk}, Conc_{lvg}] \\
 & + 2E[DF_{ing}]E[DF_{inh}]E[In_{milk}]E[In_{air}]COV[Conc_{milk}, Conc_{air}] \\
 & + 2E[DF_{ing}]E[DF_{inh}]E[In_{lvg}]E[In_{air}]COV[Conc_{lvg}, Conc_{air}]
 \end{aligned}$$

As seen in this example, the calculations become quite cumbersome, even for simple models. Actually, as shown in Example 10, this model ignores the holdup parameter for the sake of simplification. To implement the differential analysis, a computer program would need to be developed to do the calculations.

Example 9: Continuation of Example 7

The model for dose from milk ingestion is

$$\begin{aligned}
 Dose_{milk\ ingestion} = & DoseFactor_{ingestion} \times Intake_{milk} \\
 & \times Concentration_{milk} \times e^{-\lambda_{rad} \times Holdup}
 \end{aligned}$$

or

$$g(\underline{X}) = X_1 X_2 X_3 e^{-cX_4}$$

Example 9: (contd)

where the radiological decay rate, λ_{rad} , is constant and all the parameters are assumed independent.

The differentials for this example are

$$\left(\frac{\partial g(\underline{X})}{\partial X_1} \right)_{\underline{X}=\underline{\mu}_X} = \left(X_2 X_3 e^{-cX_4} \right)_{\underline{X}=\underline{\mu}_X} = \mu_{X_2} \mu_{X_3} e^{-c\mu_{X_4}}$$

$$\left(\frac{\partial g(\underline{X})}{\partial X_2} \right)_{\underline{X}=\underline{\mu}_X} = \left(X_1 X_3 e^{-cX_4} \right)_{\underline{X}=\underline{\mu}_X} = \mu_{X_1} \mu_{X_3} e^{-c\mu_{X_4}}$$

$$\left(\frac{\partial g(\underline{X})}{\partial X_3} \right)_{\underline{X}=\underline{\mu}_X} = \left(X_1 X_2 e^{-cX_4} \right)_{\underline{X}=\underline{\mu}_X} = \mu_{X_1} \mu_{X_2} e^{-c\mu_{X_4}}$$

$$\left(\frac{\partial g(\underline{X})}{\partial X_4} \right)_{\underline{X}=\underline{\mu}_X} = \left(-cX_1 X_2 X_3 e^{-cX_4} \right)_{\underline{X}=\underline{\mu}_X} = -c\mu_{X_1} \mu_{X_2} \mu_{X_3} e^{-c\mu_{X_4}}$$

$$\left(\frac{\partial^2 g(\underline{X})}{\partial^2 X_1} \right)_{\underline{X}=\underline{\mu}_X} = \left(\frac{\partial^2 g(\underline{X})}{\partial^2 X_2} \right)_{\underline{X}=\underline{\mu}_X} = \left(\frac{\partial^2 g(\underline{X})}{\partial^2 X_3} \right)_{\underline{X}=\underline{\mu}_X} = 0$$

$$\left(\frac{\partial^2 g(\underline{X})}{\partial^2 X_4} \right)_{\underline{X}=\underline{\mu}_X} = \left(c^2 X_1 X_2 X_3 e^{-cX_4} \right)_{\underline{X}=\underline{\mu}_X} = c^2 \mu_{X_1} \mu_{X_2} \mu_{X_3} e^{-c\mu_{X_4}}$$

$$\left(\frac{\partial^2 g(\underline{X})}{\partial X_1 \partial X_2} \right)_{\underline{X}=\underline{\mu}_X} = \left(X_3 e^{-cX_4} \right)_{\underline{X}=\underline{\mu}_X} = \mu_{X_3} e^{-c\mu_{X_4}}$$

$$\left(\frac{\partial^2 g(\underline{X})}{\partial X_1 \partial X_3} \right)_{\underline{X}=\underline{\mu}_X} = \left(X_2 e^{-cX_4} \right)_{\underline{X}=\underline{\mu}_X} = \mu_{X_2} e^{-c\mu_{X_4}}$$

$$\left(\frac{\partial^2 g(\underline{X})}{\partial X_1 \partial X_4} \right)_{\underline{X}=\underline{\mu}_X} = \left(-cX_2 X_3 e^{-cX_4} \right)_{\underline{X}=\underline{\mu}_X} = -c\mu_{X_2} \mu_{X_3} e^{-c\mu_{X_4}}$$

Example 9: (contd)

$$\left(\frac{\partial^2 g(\underline{X})}{\partial X_2 \partial X_3} \right)_{\underline{X}=\underline{\mu}_X} = \left(X_1 e^{-cX_4} \right)_{\underline{X}=\underline{\mu}_X} = \mu_{X_1} e^{-c\mu_{X_4}}$$

$$\left(\frac{\partial^2 g(\underline{X})}{\partial X_2 \partial X_4} \right)_{\underline{X}=\underline{\mu}_X} = \left(-cX_1 X_3 e^{-cX_4} \right)_{\underline{X}=\underline{\mu}_X} = -c\mu_{X_1} \mu_{X_3} e^{-c\mu_{X_4}}$$

$$\left(\frac{\partial^2 g(\underline{X})}{\partial X_3 \partial X_4} \right)_{\underline{X}=\underline{\mu}_X} = \left(-cX_1 X_2 e^{-cX_4} \right)_{\underline{X}=\underline{\mu}_X} = -c\mu_{X_1} \mu_{X_2} e^{-c\mu_{X_4}}$$

Then the mean and variance are

$$E[g(\underline{X})] = \mu_{X_1} \mu_{X_2} \mu_{X_3} e^{-c\mu_{X_4}} + \frac{1}{2} c^2 \mu_{X_1} \mu_{X_2} \mu_{X_3} e^{-c\mu_{X_4}} \times V[X_4]$$

and

$$\begin{aligned} V[g(\underline{X})] = & \left(\mu_{X_2} \mu_{X_3} e^{-c\mu_{X_4}} \right)^2 \times V[X_1] + \left(\mu_{X_1} \mu_{X_3} e^{-c\mu_{X_4}} \right)^2 \times V[X_2] + \left(\mu_{X_1} \mu_{X_2} e^{-c\mu_{X_4}} \right)^2 \times V[X_3] \\ & + \left(c\mu_{X_1} \mu_{X_2} \mu_{X_3} e^{-c\mu_{X_4}} \right)^2 \times V[X_4] + \left(\mu_{X_3} e^{-c\mu_{X_4}} \right)^2 \times V[X_1] \times V[X_2] \\ & + \left(\mu_{X_2} e^{-c\mu_{X_4}} \right)^2 \times V[X_1] \times V[X_3] + \left(\mu_{X_1} e^{-c\mu_{X_4}} \right)^2 \times V[X_2] \times V[X_3] \\ & + \left(c\mu_{X_2} \mu_{X_3} e^{-c\mu_{X_4}} \right)^2 \times V[X_1] \times V[X_4] + \left(c\mu_{X_1} \mu_{X_3} e^{-c\mu_{X_4}} \right)^2 \times V[X_2] \times V[X_4] \\ & + \left(c\mu_{X_1} \mu_{X_2} e^{-c\mu_{X_4}} \right)^2 \times V[X_3] \times V[X_4] - \frac{1}{4} \left(c^2 \mu_{X_1} \mu_{X_2} \mu_{X_3} e^{-c\mu_{X_4}} \times V[X_4] \right)^2 \end{aligned}$$

Example 9: (contd)

Now, given the following means and variances for each parameter

$$E[DoseFactor] = 1.78 \text{ rad}_{thyroid} / \mu Ci_{ingested}$$

$$V[DoseFactor] = (1.37 \text{ rad}_{thyroid} / \mu Ci_{ingested})^2$$

$$E[Intake_{milk}] = 100 \text{ L}_{milk \text{ ingested}}$$

$$V[Intake_{milk}] = (40 \text{ L}_{milk \text{ ingested}})^2$$

$$E[Concentration_{milk}] = 0.00673 \mu Ci_{ingested} / L_{milk \text{ ingested}}$$

$$V[Concentration_{milk}] = (0.00611 \mu Ci_{ingested} / L_{milk \text{ ingested}})^2$$

$$E[Holdup] = 2.0 \text{ days}$$

$$V[Holdup] = 1/3 \text{ days}^2$$

$$\lambda_{rad} = 0.086 / \text{day}$$

The approximate mean, variance, and standard deviation are

$$\begin{aligned} E[Dose_{milk \text{ ingestion}}] &\approx 1.78 \times 100 \times 0.00673 \times e^{-0.086 \times 2.0} \\ &\quad + \frac{1}{2} (0.086)^2 \times 1.78 \times 100 \times 0.00673 \times e^{-0.086 \times 2.0} \times \frac{1}{3} \\ &= 1.009 + 0.001 \\ &= 1.010 \text{ rad}_{thyroid} \end{aligned}$$

and

Example 9: (contd)

$$\begin{aligned}
 V[Dose_{milk\ ingestion}] \approx & (100 \times 0.00673 \times e^{-0.086 \times 2.0} \times 1.37)^2 \\
 & + (1.78 \times 0.00673 \times e^{-0.086 \times 2.0} \times 40)^2 \\
 & + (1.78 \times 100 \times e^{-0.086 \times 2.0} \times 0.00611)^2 \\
 & + (0.086 \times 1.78 \times 100 \times 0.00673 \times e^{-0.086 \times 2.0})^2 \times \frac{1}{3} \\
 & + (0.00673 \times e^{-0.086 \times 2.0} \times 1.37 \times 40)^2 \\
 & + (100 \times e^{-0.086 \times 2.0} \times 1.37 \times 0.00611)^2 \\
 & + (1.78 \times e^{-0.086 \times 2.0} \times 40 \times 0.00611)^2 \\
 & + (0.086 \times 100 \times 0.00673 \times e^{-0.086 \times 2.0} \times 1.37)^2 \times \frac{1}{3} \\
 & + (0.086 \times 1.78 \times 0.00673 \times e^{-0.086 \times 2.0} \times 40)^2 \times \frac{1}{3} \\
 & + (0.086 \times 1.78 \times 100 \times e^{-0.086 \times 2.0} \times 0.00611)^2 \times \frac{1}{3} \\
 & - \frac{1}{4} \left(0.086^2 \times 1.78 \times 100 \times 0.00673 \times e^{-0.086 \times 2.0} \times \frac{1}{3} \right)^2
 \end{aligned}$$

$$\begin{aligned}
 V[Dose_{milk\ ingestion}] \approx & 0.6027 + 0.1628 + 0.8385 + 0.0025 \\
 & + 0.0964 + 0.4967 + 0.1342 + 0.0015 \\
 & + 0.0004 + 0.0021 - 0.0000 \\
 = & 2.338 (rad_{thyroid})^2
 \end{aligned}$$

$$SD[Dose_{milk\ ingestion}] \approx 1.529 rad_{thyroid}.$$

Note, if the approximate variance were limited to the first term in Equation (3.20), the estimated variance would have been 1.6065 and the standard deviation would have been 1.267. This estimate is 18 percent smaller than the estimate in Example 9.

3.4 Estimation of Uncertainty Using Monte Carlo Technique

The Monte Carlo technique is a "procedure that involves statistical sampling techniques in obtaining a probabilistic approximation to the solution of a mathematical problem" (James and James 1968). In Section 3.4.1, the Monte Carlo technique is demonstrated using Example 10. In Section 3.4.2, different statistical sampling strategies that can be used in the Monte Carlo technique are reviewed. This technique has been used in the uncertainty analysis of the predictions of fallout radionuclides in foods and subsequent ingestion as part of the offsite Radiation Exposure Review Project (Breshears et al. 1989) and the uncertainty analysis of the food chain and atmospheric dispersion modules of MARC (Crick et al. 1988).

3.4.1 Example of Monte Carlo Technique

Assessment Question:

What is the annual dose to the thyroid of a reference adult male located near Eltopia, Washington from ingestion of milk containing iodine-131 during 1945?

The model is:

Example 10: Computational Model 6: Assessment Question/Deterministic Answer

$$Dose_{milk\ ingestion} = DoseFactor_{ingestion} \times Intake_{milk} \times Concentration_{milk} \times e^{-\lambda_{rad} \times Holdup}$$

where

$DoseFactor_{ingestion}$ = the conversion factor which converts the Curies (Ci) of iodine-131 that is ingested into the dose to the thyroid (rad)

$Intake_{milk}$ = total amount of milk consumed in 1945 (liters)

$Concentration_{milk}$ = average concentration of iodine-131 in milk during 1945 ($\mu\text{Ci/liter}$)

$Holdup$ = time milk is held prior to consumption (day).

λ_{rad} = radiological decay constant for iodine-131 (0.086/day)

Each of the parameters in this model has one true value. The relative likelihoods of the alternative values for $DoseFactor_{ingestion}$, $Intake_{milk}$, and $Holdup$ are expressed as the following PDFs. The PDFs for $DoseFactor_{ingestion}$ and $Holdup$ are from Snyder et al. (1992), and the PDF for $Intake_{milk}$ is assumed arbitrarily for the purposes of this example. See Appendix B for details about these PDFs.

$DoseFactor_{ingestion}$: lognormal distribution with $\mu_x = 0.344$ and $\sigma_x^2 = (0.682)^2$, so that

$$E[DoseFactor_{ingestion}] = 1.78 \text{ rad}_{thyroid} / \mu Ci$$

$$V[DoseFactor_{ingestion}] = (1.37 \text{ rad}_{thyroid} / \mu Ci)^2$$

$Intake_{milk}$: uniform distribution with $a(\text{minimum}) = 30 \text{ liters}$ and $b(\text{maximum}) = 170 \text{ liters}$, so that

$$E[Intake_{milk}] = 100 \text{ liters}$$

$$V[Intake_{milk}] = (40 \text{ liters})^2$$

$Holdup$: uniform distribution with $a(\text{minimum}) = 1 \text{ day}$ and $b(\text{maximum}) = 3 \text{ days}$, so that

$$E[Intake_{milk}] = 2 \text{ days}$$

$$V[Intake_{milk}] = \frac{1}{3}(\text{day})^2$$

The $Concentration_{milk}$ values are those predicted from another Monte Carlo run demonstrated in Section 5.3.1 where the estimated expected value and variance from the 100 predictions are

$$E[Concentration_{milk}] = 0.00673 \mu Ci / \text{liter}$$

$$V[Concentration_{milk}] = (0.00611 \mu Ci / \text{liter})^2.$$

For this example, 10 samples are drawn from each distribution, using simple random sampling (see Section 3.4.2.1). Then, for each sample the $Dose_{milk \text{ ingestion}}$ is calculated. The following results are obtained:

Example 11: Monte Carlo with Simple Random Sampling to Calculate Milk Ingestion for Model in Example 10				
$DoseFactor_{ingestion}$	$Intake_{milk}$	$Concentration_{milk}$	$Holdup$	$Dose_{milk \text{ ingestion}}$
1.001	87.6	0.01180	2.24	0.854
0.827	39.0	0.02129	1.48	0.606
2.261	114.1	0.00469	2.86	0.944
0.427	169.1	0.00629	1.87	0.387
2.461	135.6	0.00338	2.13	0.939
2.165	113.8	0.00065	2.16	0.132
3.687	76.4	0.00596	1.26	1.507
1.736	47.1	0.00846	2.28	0.569
0.391	165.6	0.00351	1.80	0.195
0.858	74.6	0.00197	1.89	0.107

For this small Monte Carlo simulation, the sample mean is $0.624 \text{ rad}_{thyroid}$ and the sample standard deviation is $0.446 \text{ rad}_{thyroid}$. The estimated mean and standard deviation using the differential method in Example 9 is $1.010 \text{ rad}_{thyroid}$ and $1.529 \text{ rad}_{thyroid}$, respectively. This relatively large difference in results is to be expected because the number of samples is very small and the method for drawing these samples from the distribution is not optimal for such a small sampling.

3.4.2 Statistical Sampling Strategies

A number of different statistical sampling strategies can be used with the Monte Carlo technique. The choice of the appropriate strategy is dependent upon the

- assessment question
- desired level of accuracy in the estimate(s)
- computational model
- computer system.

In this section, three different statistical sampling strategies are reviewed and compared.

3.4.2.1 Simple Random Sampling

Simple random sampling (RS) is the most straightforward method of sampling. Conceptually, a number between zero and one is randomly selected. This number is the value of $F_X(x)$, the Cumulative Distribution Function (CDF) of the parameter. The value of the parameter is the value of x which corresponds to $F_X(x)$. This procedure is repeated until the desired number of random values of the parameter are drawn.

For example, consider the parameter $Intake_{milk}$ which has a uniform distribution with a minimum of 30 and a maximum of 170. The CDF for this parameter is

$$F_X(x) = \frac{x - 30}{140}.$$

(See Appendix B.)

To obtain the first sample in Example 10, the number 0.4114 was drawn at random between zero and one. The CDF is set equal to this number. Then $Intake_{milk} = 87.6$ is the value of x that corresponds to $F_X(x)=0.4114$.

In simple random sampling, each value between zero and one has an equal probability of being selected at random. The 10 values used for the simple random sample of $Intake_{milk}$ in the example in Section 3.4.1 were (0.4114, 0.0645, 0.6007, 0.9937, 0.7541, 0.5983, 0.3315, 0.1224, 0.9684, 0.3184). These values do not cover the CDF uniformly. For example, two values are greater than 0.95, but no values are found between 0.76 and 0.94. However, as the number of samples increase, these gaps tend to fill.

3.4.2.2 Stratified Random Sampling

For stratified random sampling (SRS) a sample of size n , (the interval between zero and one) is divided into n subintervals of equal length. One random sample value is drawn from each of these subintervals. $F_X(x)$ is set equal to each of these values and the n values of parameter are set equal to the x that corresponds with $F_X(x)$.

For example, a stratified random sample of size 10 for *Intake_{milk}* may use the following values drawn between zero and one: (0.4114, 0.0645, 0.6007, 0.9937, 0.7541, 0.5983, 0.3315, 0.1224, 0.8384, 0.2806). Note that only the last two values are different from those used in the simple random sample; however, the large gap in the simple random sample between 0.76 and 0.94 is filled.

3.4.2.3 Latin Hypercube Sampling

Latin hypercube sampling (LHS) takes the stratified random sample, from Section 3.4.2.2, and reorders the samples to achieve, as closely as possible, the desired pairwise rank correlations.

For a Monte Carlo simulation with n stratified random samples and k parameters, there are $(n!)^{k-1}$ different combinations of the parameter values. For example, if n is three and k is two, the six different combinations in Figure 3.1 are possible (Iman and Helton 1985).

In combination (a), the value of X_1 from the first interval is combined with the value of X_2 from the first interval, the value of X_1 from the second interval is combined with the value of X_2 from the second interval, and finally the value of X_1 from the third interval is combined with the value of X_2 from the third interval. This combination has a rank correlation of 1. Combination (b) has a rank correlation of -1. If the two parameters are supposed to be uncorrelated, then combinations (a) and (b) are inappropriate samples because they have perfect rank correlations and LHS would reject those combinations. That is, LHS would choose one of the combinations (c) through (f), if the parameters are supposed to be uncorrelated. Note that the rank correlation for combinations (c) and (d) is 0.5 and for combinations (e) and (f) is -0.5. Although it is impossible to achieve a zero rank correlation, combinations (c) through (f) are more appropriate, if zero correlation is desired.

Figure 3.1. Stratified Random Sampling Combinations

3.4.2.4 Comparison of Simplified Random, Stratified Random, and Latin Hypercube Sampling

In this section, a number of different properties of the three sampling schemes are compared. Then a numerical comparison is performed using the *Dose_{milk ingestion}* example in Section 3.4.1 (Example 10).

- **Computational Ease:** The algorithm is simplest for RS and most complex for LHS, with SRS in the middle. The required computer time is the shortest for RS and longest for LHS, with SRS again in the middle.
- **Estimation of the Mean:** All three sampling schemes give unbiased estimates of the mean (McKay et al. 1979).
- **Precision of Estimated Mean:** The variance of the estimated mean is largest for RS and smallest for LHS, with SRS in the middle (McKay et al. 1979). Therefore, to have the same precision, more samples must be used for RS than for LHS, with the sample size for SRS again in the middle.
- **Aggregation:** Of the three sampling types, only RS allows the results of more than one Monte Carlo run to be aggregated. That is, if two Monte Carlo runs of sample size 100 are performed with RS, the results could be aggregated into one Monte Carlo run of sample size 200. Aggregation cannot be conducted for SRS or LHS.

To examine the magnitude of the differences between SRS and LHS, a Monte Carlo run was performed again using Example 10, using an SRS size of 100. The maximum absolute value of the pairwise rank correlations between *DoseFactor_{ingestion}*, *Intake_{milk}*, *Concentration_{milk}*, and *Holdup* from this SRS was 0.158. Next, these 100 samples were randomly recombined; that is the same values were used, but were randomly assigned to different simulations. This was done 99 times, so that 100 different Monte Carlo runs were performed with the same stratified random sample values, but different combinations for each run. For these 100 Monte Carlo runs, the maximum absolute value of the pairwise rank correlations ranged from 0.084 to 0.325.

Because the parameters are independent, the desired pairwise correlation is zero. LHS creates combinations of the parameter values where the pairwise rank correlations are close to zero. To achieve this effect, 10,000 random combinations were created, and the 100 combinations with the smallest maximum absolute value of the pairwise rank correlations, 0.016 to 0.062, were selected.

Finally, the standard deviations and means of the 100 dose predictions for each of 100 SRS and LHS Monte Carlo runs were calculated and boxplots of these sample statistics are shown in Figure 3.2. These boxplots show the minimum, maximum, and 5th, 25th, 50th (median), 75th, and 95th percentiles. As seen in Figure 3.2, the mean predictions using SRS have more variability than those using LHS. However, the standard deviations of the predicted dose using SRS have less variability than those using LHS. Figure 3.3 shows the envelopes that contain the 100 predicted CDFs from the SRS and LHS Monte Carlo runs. The LHS envelope is generally slightly smaller than the SRS envelope.

Thus, for this example, the LHS strategy has more precision when estimating the mean and SRS has more precision when estimating the standard deviation (uncertainty). The LHS strategy has a slightly more precise estimate of the CDF.

Figure 3.2. Means and Standard Deviations from Two Kinds of Sampling (SRS and LHS)

Figure 3.3. Envelopes Containing Cumulative Distribution Functions

3.5 Description of Uncertainty in Prediction

The fifth step in the uncertainty and sensitivity analysis process is the quantitative description of the uncertainty in the prediction that is a result of the uncertainty in the model parameters. Section 3.5.1 discusses the use of sample statistics as descriptors of uncertainty. Section 3.5.2 discusses the use of the subjective confidence intervals derived from the CDF and Section 3.5.3 describes the use of boxplots.

3.5.1 Variance, Standard Deviation, and Coefficient of Variation

Definitions of the sample statistics, including the variance, standard deviation, and coefficient of variation, are in Appendix A.

When the uncertainty is estimated using mathematical techniques (see Section 3.3), the results are estimates of the mean and variance of the prediction. The standard deviation (SD) and coefficient of variation (CV) can be calculated directly from these results. For Example 2 in Section 3.3.2

$$\begin{aligned}E[Dose_{milk\ ingestion}] &= 1.010\ rad_{thyroid} \\V[Dose_{milk\ ingestion}] &= 2.338\ (rad_{thyroid})^2 \\SD[Dose_{milk\ ingestion}] &= 1.529\ rad_{thyroid} \\CV[Dose_{milk\ ingestion}] &= 1.514\end{aligned}$$

When the uncertainty is estimated using the Monte Carlo technique, the results are a set of predictions. In Example 10, the set contained only 10 predictions. For this section, the number of predictions is increased to 100. The following set of sample statistics are calculated using the formulae in Appendix A

$$\begin{aligned}E[Dose_{milk\ ingestion}] &= 0.896\ rad_{thyroid} \\V[Dose_{milk\ ingestion}] &= 1.426\ (rad_{thyroid})^2 \\SD[Dose_{milk\ ingestion}] &= 1.194\ rad_{thyroid} \\CV[Dose_{milk\ ingestion}] &= 1.334\end{aligned}$$

These descriptors have limited use. Clearly, the uncertainty is large relative to the estimated mean of the dose prediction, as the coefficient of variation for both techniques is greater than 1. However, not much more can be said about the uncertainty with this information.

3.5.2 Subjective Confidence Interval

When the uncertainty is estimated using the Monte Carlo technique, the resulting predictions can be used to estimate the CDF of the prediction. The estimated CDF derived from the 100 predictions used in Section 3.5.1 is shown in Figure 3.4. To create the CDF, the 100 predictions are

Figure 3.4. Estimated Cumulative Probability Distribution Functions and Boxplots of Predicted Dose Using Mathematical and Monte Carlo Techniques

ordered from smallest to largest. Then, the estimated CDF jumps from 0.0 to 0.01 at the smallest prediction, jumps again to 0.02 at the second smallest prediction, and so on. The subjective confidence limits and intervals can be derived from this estimated CDF. As seen in Figure 3.4, the 95 percent subjective confidence limit for the Monte Carlo run is 3.922 rad. The 5 percent subjective confidence limit for the Monte Carlo run is 0.073 rad. Combining these subjective confidence limits, a 90 percent subjective confidence interval of 0.073 rad to 3.922 rad is estimated. Thus, at a subjective confidence level of 90 percent, the true dose is predicted to be in the interval between 0.073 rad and 3.922 rad. [The confidence limits and intervals are called subjective to distinguish them from statistical confidence limits and intervals (see Appendix A)].

When the uncertainty is estimated using the mathematical techniques, subjective confidence intervals can be derived only if the form of CDF is assumed. For Example 2 in Section 3.3.2, the CDF of the predicted dose is assumed to be a lognormal since the Dose Factor is assumed to have a lognormal distribution (see Appendix B), and the concentration of iodine-131 in milk appears to be approximately lognormally distributed. Figure 3.4 shows the estimated CDF and 95 percent subjective confidence limit for the predicted dose using the mathematical technique. The 90 percent subjective confidence interval for this prediction is 0.092 rad to 3.353 rad.

As seen in this section, the subjective confidence intervals provide far more information about the uncertainty in the predicted dose than the sample statistics used in Section 3.5.1. With the subjective confidence interval, the range within which the true dose is predicted to be, with some degree of belief, is presented. This interval now actually displays how large the uncertainty is in terms of the predicted dose.

3.5.3 Boxplot

While the subjective confidence interval presents the uncertainty in the predicted dose, the actual predicted dose is also of interest. The boxplot is a shorthand description of the CDF (see Appendix A). It displays information about both the center of the CDF and the spread of the CDF. The information included in the boxplot changes with the application.

Figure 3.4 shows the boxplots derived from the estimated CDF of predicted dose for both uncertainty estimation techniques. For both techniques, the boxplots include the 5th, 25th, 50th (median), 75th, and 95th percentiles. These percentiles correspond with $p\%$ subjective confidence limits. For the Monte Carlo technique, the minimum and maximum are used in the extremes. For the mathematical technique, the 1st and 99th percentiles are used, as the minimum for the lognormal distribution is zero and the maximum is infinity. Additionally, the estimated mean is included in the boxplots.

As seen in the first set of boxplots in Figure 3.4, the distribution of the predicted doses is highly skewed to the right. Often, when the results are skewed, the boxplots are displayed in logarithmic (base 10) scale, as shown in the second set of boxplots in Figure 3.4. In the logarithmic scale, an order of magnitude (or a factor of 10) is present between each major tick mark on the axis.

From inspecting the boxplots, the following information can be observed simultaneously: the central estimate of the predicted dose, the range of the predictions and the 90 percent subjective confidence interval. The boxplot also shows the estimated mean of the predicted dose (the center of gravity of the probability density function) is not necessarily the most probable predicted dose. In Figure 3.4, the estimated mean is close to the estimated 75th percentile. Thus, it is three times more likely the predicted dose is below the mean than above the mean. Boxplots show 5th, 25th, 50th

(median), 75th, and 95th percentiles. The extremes for the Monte Carlo are the minimum and maximum of 100 predictions, and the 1st and 99th percentile for the mathematical. The dashed line is the estimated mean.

3.6 Comparison of Mathematical and Monte Carlo Techniques

When the model is simple and the parameters are independent, the analytical method of uncertainty estimation is easy to implement. However, to get a good estimate of the variance (the uncertainty in the estimate), the analytical method should be carried out to the third term in the Taylor series.

When the parameters are not independent, the mathematical method is not as reliable because so many terms in the equations must be dropped to make the equation tractable. When the model is complex, the analytical method may be difficult, if not impossible, to implement.

While the uncertainty estimation using the mathematical method is initially easier to implement than the Monte Carlo method for simple models with independent parameters, the information available for uncertainty analysis is very limited for the analytical method. The analytical method estimates only the mean and variance. To discuss these results in terms of subjective confidence intervals about the predicted mean *Dose_{milk ingestion}* or to display a boxplot of the results, the CDF of the result needs to be determined. This determination requires the CDF of the predicted dose be assumed. The CDF can only be assumed for the predicted dose if the CDFs (or PDFs) of the parameters are also assumed. Thus, to effectively analyze the uncertainty in the results from the analytical method, the same amount of effort needs to be extended in determining the PDFs of the parameters as would be extended to use the Monte Carlo method. It appears the amount of effort needed to use the mathematical method is as great as that for the Monte Carlo technique.

In Figure 3.4, the CDF and boxplot from the mathematical method is displayed along with the CDF and boxplot from the Monte Carlo run using stratified random sampling discussed in Section 3.5. As seen in this figure, the differences between the predicted dose from the mathematical and Monte Carlo techniques are quite small, relative to the uncertainty in the predicted dose. The use of the Monte Carlo technique is recommended by Granger and Morgan (1990), IAEA (1989), Iman and Helton (1985), and SENES (1985).

4.0 Sensitivity Analysis Approaches

The sixth step in the uncertainty and sensitivity analysis process is the sensitivity analysis. The purpose of the sensitivity analysis is to identify those parameters whose uncertainties have the largest impact on the uncertainty of the predicted dose. The results of the sensitivity analysis

- contribute to the understanding and interpretation of the predictions from mathematical models (Iman and Helton 1988).
- permit ranking of the parameters with respect to their contribution to the uncertainty in the model prediction (IAEA 1989).

The sensitivity analysis approaches that are appropriate for mathematical and Monte Carlo uncertainty analysis techniques are discussed in Sections 4.1 and 4.2, respectively. The hierarchical approach to uncertainty and sensitivity analysis is discussed in Section 4.3 and demonstrated in Section 5.0. The purpose of Sections 4.1 and 4.2 is to prepare the reader for Section 5.

4.1 Sensitivity Analysis of Mathematically Estimated Uncertainty

This section reviews the standardized partial differential (SPD) method used to perform sensitivity analysis of mathematically estimated uncertainties (Section 3.3). The primary sources of this section are Iman and Helton (1985).

The SPD at basepoint $\underline{x}_0$, for the k th parameter is the function:

$$SPD(\underline{x}_0)_k = \left(\frac{\partial g(\underline{x})}{\partial x_k} \right) \bigg|_{\underline{x}=\underline{x}_0} \times \frac{s_{x_k}}{s_{g(\underline{x})}} \quad (4.1)$$

The SPD measures the effect on the function, in terms of standard deviations, of changing the k th parameter one standard deviation from the basepoint, while all the other parameters are held constant. The SPD has a positive sign when the value of the function increases when the parameter value increases and has a negative sign when the value of the function decreases when the parameter value increases.

Example 12: Sensitivity Analysis of Example 9

The model is:

$$Dose_{milk\ ingestion} = DoseFactor_{ingestion} \times Intake_{milk} \times Concentration_{milk} \times e^{-\lambda_{rad} \times Holdup}$$

Example 12: (contd)

For this example, the SPDs at $\underline{\mu}_x$, the mean values of the parameters, are

$$\begin{aligned}
 SPD(\underline{\mu}_x)_{DoseFactor_{ingestion}} &= E[Intake_{milk}] \times E[Concentration_{milk}] \\
 &\quad \times e^{-0.086 E[Holdup]} \times \frac{SD[DoseFactor_{ingestion}]}{SD[Dose_{milk\ ingestion}]} \\
 &= 100 \times 0.00673 \times e^{-0.086 \times 2} \times \frac{1.37}{1.529} \\
 &= 0.508
 \end{aligned}$$

$$\begin{aligned}
 SPD(\underline{\mu}_x)_{Intake_{milk}} &= E[DoseFactor_{milk}] \times E[Concentration_{milk}] \\
 &\quad \times e^{-0.086 E[Holdup]} \times \frac{SD[Intake_{milk}]}{SD[Dose_{milk\ ingestion}]} \\
 &= 1.78 \times 0.00673 \times e^{-0.086 \times 2} \times \frac{40}{1.529} \\
 &= 0.264
 \end{aligned}$$

$$\begin{aligned}
 SPD(\underline{\mu}_x)_{Concentration_{milk}} &= E[DoseFactor_{milk}] \times E[Intake_{milk}] \\
 &\quad \times e^{-0.086 E[Holdup]} \times \frac{SD[Concentration_{milk}]}{SD[Dose_{milk\ ingestion}]} \\
 &= 1.78 \times 100 \times e^{-0.086 \times 2} \times \frac{0.00611}{1.529} \\
 &= 0.599
 \end{aligned}$$

$$\begin{aligned}
 SPD(\underline{\mu}_x)_{Holdup} &= E[DoseFactor_{milk}] \times E[Intake_{milk}] \times E[Concentration_{milk}] \\
 &\quad \times (-0.086) \times e^{-0.086 E[Holdup]} \times \frac{SD[Holdup]}{SD[Dose_{milk\ ingestion}]} \\
 &= 1.78 \times 100 \times 0.00673 \times (-0.086) \times e^{-0.086 \times 2} \times \frac{0.577}{1.529} \\
 &= -0.033
 \end{aligned}$$

When the parameters are independent of each other, as in this example, the interpretation of the SPDs is straightforward. This sensitivity analysis indicates that uncertainty in $Concentration_{milk}$ and $DoseFactor_{ingestion}$ has the greatest effect on the uncertainty in the estimated $Dose_{milk\ ingestion}$. That is, $Dose_{milk\ ingestion}$ is most sensitive to the lack of knowledge about the true values of $Concentration_{milk}$ and $DoseFactor_{ingestion}$. When the $Concentration_{milk}$ is increased one standard deviation from its mean, the $Dose_{milk\ ingestion}$ increases 0.599 standard deviations. When the $DoseFactor_{ingestion}$ is increased one standard deviation from its mean, the $Dose_{milk\ ingestion}$ increases 0.508 standard deviations. The lack of knowledge about the true value of $Intake_{milk}$ for the reference adult male has about half the effect on the uncertainty of $Dose_{milk\ ingestion}$ as $Concentration_{milk}$ and $DoseFactor_{ingestion}$. Finally, the lack of knowledge about the true value of $Holdup$ has almost no effect on the uncertainty of $Dose_{milk\ ingestion}$.

If the basepoint for the SPDs is changed, the results can be dramatically changed. For example, consider the following basepoints,

$$\underline{x}_1 = \begin{pmatrix} DoseFactor_{ingestion} = 3.0 \\ Intake_{milk} = 100 \\ Concentration_{milk} = 0.002 \\ Holdup = 2.0 \end{pmatrix} \quad \underline{x}_2 = \begin{pmatrix} DoseFactor_{ingestion} = 0.75 \\ Intake_{milk} = 100 \\ Concentration_{milk} = 0.012 \\ Holdup = 2.0 \end{pmatrix}$$

Then the SPDs are

$$\begin{array}{ll} SPD(\underline{x}_1)_{DoseFactor_{ingestion}} = 0.151 & SPD(\underline{x}_2)_{DoseFactor_{ingestion}} = 0.904 \\ SPD(\underline{x}_1)_{Intake_{milk}} = 0.132 & SPD(\underline{x}_2)_{Intake_{milk}} = 0.198 \\ SPD(\underline{x}_1)_{Concentration_{milk}} = 1.008 & SPD(\underline{x}_2)_{Concentration_{milk}} = 0.252 \\ SPD(\underline{x}_1)_{Holdup} = -0.016 & SPD(\underline{x}_2)_{Holdup} = -0.024 \end{array}$$

If the parameters are not independent of each other, the analysis of the SPDs, as done in the above example, may not be meaningful. For example, if two of the parameters have a high positive correlation, then when one of them is increased by one standard deviation, the other parameter must also increase. Thus, the original definition (the SPD measures the effect on the function from changing one parameter while all the other parameters remain constant) is no longer operable.

4.2 Sensitivity Analysis of Monte Carlo Estimated Uncertainty

This section reviews methods available to perform sensitivity analysis of uncertainty estimates derived from the Monte Carlo technique (see Section 3.4). The primary sources of this section are Iman and Helton (1985), IAEA (1989), and Neter et al. (1985).

4.2.1 Measures from Multiple Regression

A multiple linear regression is performed on the results of the Monte Carlo runs on the parameter values used in the runs. For the dose model

$$Dose_{milk\ ingestion} = DoseFactor_{ingestion} \times Intake_{milk} \times Concentration_{milk} \times e^{-\lambda rad} \times Holdup$$

the following multiple linear regression model is used

$$Dose_{milk\ ingestion} = \beta_0 + \beta_1 DoseFactor_{ingestion} + \beta_2 Intake_{milk} + \beta_3 Concentration_{milk} + \beta_4 Holdup + \varepsilon$$

The Monte Carlo simulation with 100 random stratified samples (see Section 3.5) has a coefficient of determination, R^2 , of 0.716. That is, for this multiple regression 71.6 percent of the uncertainty of the 100 dose predictions is explained by a multiple linear regression on the four parameters.

The multiple regression can also be performed on the ranks of the predictions and parameters. For the rank transformed values, R^2 is 0.901. That is, 90.1 percent of the uncertainty of the ranks of the 100 dose predictions is explained by a multiple linear regression on the ranks of the four parameters. This large value of R^2 indicates the relationship between the dose estimates and the parameters is monotonic. Also, with a large R^2 , it is appropriate to use measures derived from the multiple linear regression on the rank transformed values to rank the parameters according to their contribution to the uncertainty in the dose estimate.

Two of the measures that can be used to rank the parameters according to their contribution to the uncertainty in the dose prediction are the SRC and the CPD described in Appendix A. The values of these measures for this example are shown in Table 4.1. For illustration, this table includes the measures for both multiple regressions, although in practice only the results for the multiple regression with the rank transformed values would be used, because the R^2 is so much larger. Additionally, the results of the sensitivity analysis in Section 4.1 are included in Table 4.1.

Table 4.1. Sensitivity Analysis Measures for Model of $Dose_{milk\ ingestion}$ in Examples 12 and 13

Parameter	SRC	CPD	rank SRC	rank CPD	SPD(μ)
$Concentration_{milk}$	0.670	0.603	0.807	0.865	0.599
$DoseFactor_{ingestion}$	0.551	0.511	0.576	0.762	0.508
$Intake_{milk}$	0.330	0.271	0.354	0.552	0.264
$Holdup$	-0.008	0.000	-0.040	0.015	-0.033
R^2	0.716		0.901		

First, a review of what each measure means is discussed. For the *Concentration_{milk}*,

- the SRC of 0.670 means that when *Concentration_{milk}* is changed by one standard deviation, the predicted *Dose_{milk ingestion}* changes by 0.670 standard deviations with the remaining parameters held constant.
- the CPD of 0.603 means that *Concentration_{milk}* explains 60.3 percent of the remaining uncertainty of the predicted *Dose_{milk ingestion}*, after the contributions of *DoseFactor_{ingestion}*, *Intake_{milk}*, and *Holdup* have been accounted for. That is, a multiple regression with the three parameters, *DoseFactor_{ingestion}*, *Intake_{milk}*, and *Holdup*, explains 28.5 percent of the uncertainty of predicted *Dose_{milk ingestion}*. This leaves 71.5 percent of the uncertainty of the predicted *Dose_{milk ingestion}* unexplained. When the *Concentration_{milk}* is added to the multiple regression, 60.3 percent of the unexplained uncertainty in *Dose_{milk ingestion}*, or an additional 43.1 percent (60.3 of 71.5 percent) of the total uncertainty of the uncertainty *Dose_{milk ingestion}*, is explained. This brings total percent of explained uncertainty to 71.6 percent (28.5 percent plus 43.1 percent).
- the rank SRC of 0.807 means that when the rank of *Concentration_{milk}* is changed by one place, the rank of the estimated *Dose_{milk ingestion}* changes by 0.807 places, with the remaining parameters held constant. The SRC and regression coefficient are the same when a regression is performed with the rank transformed values.
- the rank CPD of 0.865 means that rank of *Concentration_{milk}* explains 86.5 percent of the remaining uncertainty of the ranks of the estimated *Dose_{milk ingestion}*, after the contributions of the ranks of *DoseFactor_{ingestion}*, *Intake_{milk}*, and *Holdup* have been accounted for.
- the SPD(μ) of 0.599 means that when *Concentration_{milk}* is changed by one standard deviation from its mean, the estimated *Dose_{milk ingestion}*, changes by 0.599 standard deviations, with the remaining parameters held constant at their mean value.

Because the example is relatively simple and the four parameters are independent, it is not surprising the ranking of the four parameters remains the same for all of the measures. However, for the rank SRC, a markedly greater difference is shown in the values for *Concentration_{milk}* and *DoseFactor_{ingestion}*, which may lead to assigning more importance to the uncertainty in *Concentration_{milk}* than in *DoseFactor_{ingestion}*. Remember that when interpreting these measures:

- The SRC and CPD should be used only when the R^2 for the multiple regression, either for the actual values or the rank transformed values is large. If the R^2 is small for both cases, then the relationship between the estimated values and the parameters is not linear or monotonic, in which case this method is inappropriate for a sensitivity analysis.
- If the parameters are independent, the SRC is a good measure of sensitivity and is much easier to interpret. However, if the parameters are correlated, the SRC must be used with caution, because it is no longer meaningful to change only one parameter at a time.
- If the parameters are correlated, the CPD is a good measure of sensitivity, as it takes the correlations between parameters into account. However, the CPD tends to give the impression that a parameter is more important than it actually is. In the example previously given, *Concentration_{milk}* had a CPD of 60.3 percent, but actually it only

explained 43.1 percent of the uncertainty of $Dose_{milk\ ingestion}$. Indeed, $DoseFactor_{ingestion}$ and $Intake_{milk}$ had CPDs of 0.511 and 0.271, respectively, but combined with $Holdup$ explained only 28.5 percent of the uncertainty of $Dose_{milk\ ingestion}$.

4.2.2 Alternative Measures When Multiple Regression is Inappropriate

If the coefficient of determination from a multiple regression is small (less than 50 percent), more than 50 percent of the uncertainty in the predicted dose is not explained by either a linear or monotonic relationship with one or more of the model parameters. Thus, the measures from multiple regression are no longer appropriate for the sensitivity analysis, because so much of the uncertainty is unexplained. When this occurs, the two recommended options are

- transform the parameters so the new values have a linear or monotonic relationship with the predicted dose and return to the measures from multiple regression (demonstrated in Section 5.5)
- partition the uncertainty by holding subsets of the parameters constant and measuring the reduction in the uncertainty for each subset (demonstrated in Section 5.6).

4.3 Hierarchical Uncertainty and Sensitivity Analyses

When the computational model is complex and contains many model parameters, the uncertainty and sensitivity analyses should be performed hierarchically. For example, consider the model in Example 8

$$Dose_{total} = DoseFactor_{ingestion} (Intake_{milk} \times Concentration_{milk} + Intake_{lvg} \times Concentration_{lvg}) \\ + DoseFactor_{inhalation} (Intake_{air} \times Concentration_{air})$$

With only eight parameters, this model appears simple. However, the three concentration parameters (milk, leafy vegetables, and air) are actually the predictions of other computational models, each with another set of parameters, some of which are, in turn, predictions of other computational models (see Figures 2.1 through 2.6).

The first tier of the uncertainty and sensitivity analyses of the dose would consider the eight parameters in the model. If the uncertainty in the dose is sensitive to one or more of the concentration parameters, then the uncertainty and sensitivity analyses of the dose would proceed to the second tier. That is, uncertainty and sensitivity analyses on the models for the sensitive concentrations would be performed as part of the uncertainty and sensitivity analyses for the dose.

The hierarchical approach to uncertainty and sensitivity analyses is demonstrated in Section 5.0. This approach leads to uncertainty and sensitivity analyses on all the major components of HEDRIC.

5.0 Demonstration of Uncertainty and Sensitivity Analysis Methods

This section demonstrates the uncertainty and sensitivity analysis methods discussed in preceding sections. The demonstration of the uncertainty and sensitivity analysis methods has the following three purposes

- 1) Education

Sections 3 and 4 of this document describe the methods and show some limited examples of their use. In this section, the methods are used on the actual results of HEDRIC to illustrate the entire uncertainty and sensitivity analysis process.

- 2) Proof of Principle

The demonstration is for the most complex set of HEDR models, HEDRIC. If the methods work for these models, they will also work on simpler models, in particular, the models for the water pathway, which are much simpler. Refer to Section 3.0.

- 3) Report of Actual Results to Date

Step 3 of the HEDR uncertainty and sensitivity analysis process, which is the strategy for specifying parameter uncertainties (see Snyder et al. 1992 and Section 3.2 of this document), calls for reexamining those parameters whose uncertainties are most important to the uncertainty in dose. The demonstration in this section produces the necessary ranking to proceed with that reexamination.

An overview of the demonstration is given in Section 5.1. In Sections 5.2 through 5.6, the hierarchical uncertainty and sensitivity analyses are performed. The hierarchy analysis starts at the dose model (Section 5.2). The second tier considers the model for iodine-131 concentration in milk (Section 5.3). The third tier considers the model for iodine-131 concentration in pasture (Section 5.4) and harvested plant products (Section 5.5). Finally, the fourth tier considers the model for surface deposition of iodine-131 (Section 5.6).

Each section has four subsections:

- 1) description of the model and parameters
- 2) description of the uncertainty analysis
- 3) description of the sensitivity analysis
- 4) summarization of the results.

The summaries in Section 5.2 through 5.6 consist of three elements: first, the key conclusions of the uncertainty and sensitivity analyses; second, the key implications the uncertainty and sensitivity analyses have on the model and its parameter (that is, what parameters need to be reexamined); and third,

the key implications the uncertainty and sensitivity analyses have on the development of uncertainty and sensitivity analyses plans. Section 5.7 states the conclusions that can be drawn from the demonstration.

5.1 Demonstration Overview

The uncertainty and sensitivity analysis process begins with determining the assessment question. For this demonstration, the assessment question is

What is the annual dose to the thyroid of a reference adult male from ingestion of cow's milk containing iodine-131 for the year 1945?

The adult male is used because no age dependent-parameters are in this age category (for instance, dose factors change with age of infants). The milk ingestion pathway is considered because it is the major dose pathway, if milk is consumed.

The assessment question is asked about cow's milk for four different combinations of cow type (Snyder et al. 1992) and feeding regime (Beck et al. 1992):

<u>Cow Type</u>	<u>Feeding Regime</u>
Individual	1 - irrigated pasture
Average	3 - stored feed

The individual cow is a single cow from which milk is obtained; the same cow is the source of milk for the entire year. The average cow's milk is a mixture of milk from 25 different cows and the same 25 cows are the source of milk for the entire year. The average cow is a surrogate for a herd of cows, where the milk is mixed, or averaged across cows. The major component of feeding Regime 1 is fresh pasture. Additionally feeding Regime 1 includes grain, alfalfa, silage, and soil. Feeding Regime 3 has no fresh pasture, and includes only grain, alfalfa, and soil.

The assessment question is asked about the reference adult male residing at the following eight locations, shown in Figure 5.1. Each circle in the figure indicates the distance from the release point of the iodine-131.

- Eltopia, Washington
- Richland, Washington
- Ritzville, Washington
- Benton City, Washington
- Spokane, Washington
- Yakima, Washington
- Pendleton, Oregon

★ B & T Plant Location

Figure 5.1. HEDR Atmospheric Model Domain

- Baker, Oregon

These eight locations are selected because they represent a variety of different spatial attributes

- Location within the major iodine-131 plume
- Total iodine-131 deposited in 1945
- Pattern of iodine-131 deposition.

Table 5.1 shows the average predicted total surface contamination for 1945 and coefficient of variation (CV) of the 100 predictions from RATCHET for each location along with the average predicted source term for 1945 and CV of the 100 predictions from STRM. The distance from release point does not accurately portray disposition or dose. Figure 5.2 shows the average predicted, weekly surface contamination for each location. As seen in this figure, a large variation from week to week is present in the surface contamination of iodine-131, even within the major plume. The farther the location is from the major plume, the more episodic the surface deposition becomes. (Surface deposition is a rate, curies per squared meter per day. Surface contamination is a cumulative total of the amount of iodine-131 that has deposited.) For example, Baker and Yakima receive, on an average, 95 percent of their annual surface deposition in only 12 noncontiguous days, that is, days not touching the immediate preceding or following day. Eltopia takes 109 noncontiguous days to accumulate 95 percent of its annual surface deposition. Thus, Eltopia, Richland, Ritzville, and Spokane are within the major plume of iodine-131 fallout from Hanford. Benton City and Pendleton

Table 5.1. Average Predicted Annual Source Term and Surface Contamination of Iodine-131 - 1945

Location	Average Predicted Annual Iodine-131 ^(a)	Coefficient of Variation ^(b)	Average Number of Noncontiguous Days To Receive P% of Annual Iodine-131				
			50%	75%	90%	95%	99%
SOURCE	555,909	0.054	25	66	179	206	245
Richland	20690	0.200	15	37	69	91	132
Eltopia	20526	0.187	19	46	83	109	158
Benton City	5783	0.253	8	21	41	55	86
Ritzville	5251	0.188	14	39	73	97	143
Spokane	2165	0.146	16	38	65	84	123
Pendleton	2149	0.154	5	12	24	34	54
Yakima	181	0.351	3	5	9	12	20
Baker	73	0.350	3	5	9	12	20

(a) For SOURCE, Annual Iodine-131 is in Ci. At each location Annual Iodine is the Surface Deposition in nCi per m².

(b) The average and CV for SOURCE are for 100 predictions from STRM. The average and CV of surface contamination at each location are for 100 predictions from RATCHET.

Figure 5.2. Average of 100 Predicted Weekly Total Surface Depositions and the Harvest/Pasture Seasons - 1945

are on the edge of the major plume because of the more episodic nature of their surface deposition, as compared to locations within the plume with the same total surface deposition. Because of the small amount of iodine-131 they received, Yakima and Baker are outside the major plume.

5.2 Uncertainty and Sensitivity Analyses of Dose from Milk Ingestion

The hierarchical uncertainty and sensitivity analysis process begins with the analysis of the model and its parameters for dose to the thyroid from ingestion of milk containing iodine-131. This model and its parameters are described in Section 5.2.1. The uncertainty analysis is presented in Section 5.2.2, followed by the sensitivity analysis in Section 5.2.3. The key conclusions of the uncertainty and sensitivity analyses are summarized in Section 5.2.4.

5.2.1 Model and Parameters

The model is

$$Dose_{milk\ ingestion} = DoseFactor_{ingestion} \times Intake_{milk} \times Concentration_{milk} \times e^{-\lambda_{rad} \times Holdup}$$

where $Dose_{milk\ ingestion}$ = annual dose to the thyroid for a reference adult male from ingestion of cow's milk containing iodine-131 for the year 1945, $rad_{thyroid}$.

$DoseFactor_{ingestion}$ = dose factor for a reference adult male, from Snyder et al. (1992), $rad_{thyroid} / Ci_{ingested}$.

$Intake_{milk}$ = annual amount of cow's milk ingested by a reference adult male, set arbitrarily for demonstration, $liters_{ingested}$.

$Concentration_{milk}$ = average daily concentration of iodine-131 in cow's milk in 1945, 100 predictions generated by DESCARTES (as implemented in PILOT), $Ci_{ingested} / liter_{ingested}$.

$Holdup$ = time between collection of milk and ingestion of milk, from Snyder et al. (1992), days.

λ_{rad} = radiological decay rate for iodine-131, from Snyder et al. (1992), 0.086/day.

The relative likelihoods of the alternative values for $DoseFactor_{ingestion}$, $Intake_{milk}$, and $Holdup$ are expressed with the following PDFs (from Appendix B):

$DoseFactor_{ingestion}$: lognormal distribution with $\mu_x = 0.344$ and $\sigma_x^2 = (0.682)^2$, so that

$$E[DoseFactor_{ingestion}] = 1.78 \text{ rad}_{thyroid} / \mu\text{Ci}$$

$$V[DoseFactor_{ingestion}] = (1.37 \text{ rad}_{thyroid} / \mu\text{Ci})^2$$

$Intake_{milk}$: uniform distribution with $a(\text{minimum}) = 30 \text{ liters}$ and $b(\text{maximum}) = 170 \text{ liters}$, so that

$$E[Intake_{milk}] = 100 \text{ liters}$$

$$V[Intake_{milk}] = (40 \text{ liters})^2$$

$Holdup$: uniform distribution with $a(\text{minimum}) = 1 \text{ day}$ and $b(\text{maximum}) = 3 \text{ days}$, so that

$$E[Intake_{milk}] = 2 \text{ days}$$

$$V[Intake_{milk}] = \frac{1}{3}(\text{day})^2$$

Stratified random sampling is used to obtain 100 samples of each of these parameters.

Table 5.2 shows the mean and CV of the 100 predictions of $Concentration_{milk}$ for each of the different combinations of cow type (individual and average) and feeding regime at each location. These concentrations in cow's milk are measured in $\mu\text{Ci/L}$ and are taken from the PILOT database.

5.2.2 Uncertainty Analysis

The mean and CV for the 100 predicted annual doses to the thyroid of a reference adult male from ingestion of cow's milk containing iodine-131 for the year 1945, for each cow type and feeding regime combination, at each location are shown in Table 5.3. Figure 5.3 shows the boxplots for the predicted annual doses. The boxplots show the key percentiles from the subjective CDF of the predicted annual doses, 5th, 25th, 50th (median), 75th, and 95th. Additionally, the mean of the predicted annual doses is included in the boxplot, as are the minimum and maximum predictions.

Table 5.4 shows the 5th and 95th subjective confidence limits (the 5th and 95th percentiles from the subjective CDF) which give a 90 percent subjective confidence interval. For example, at a 90 percent subjective confidence level, the true 1945 annual dose to thyroid for a reference adult male in Eltopia from ingestion of an individual cow's milk (on feeding Regime 1) containing iodine-131 is predicted to be between 0.331 and 17.75 rad. Alternatively, at a 95 percent subjective confidence level, the true dose is predicted to be less than 17.75 rad.

Table 5.2. Mean and Coefficient of Variation of 100 Predicted Concentrations of Iodine-131 in Cow's Milk - 1945

<u>Location</u>	<u>Regime 1</u>				<u>Regime 3</u>			
	<u>Individual Cow</u>		<u>Average Cow</u>		<u>Individual Cow</u>		<u>Average Cow</u>	
	<u>Mean</u>	<u>CV</u>	<u>Mean</u>	<u>CV</u>	<u>Mean</u>	<u>CV</u>	<u>Mean</u>	<u>CV</u>
Eltopia	31.35	0.87	31.70	0.40	6.73	0.91	6.73	0.49
Richland	21.09	0.95	20.82	0.43	4.12	1.10	3.82	0.54
Ritzville	6.72	0.94	6.63	0.43	1.19	1.01	1.15	0.53
Benton City	6.13	1.11	5.67	0.50	1.41	1.31	1.33	0.67
Spokane	2.27	0.94	2.27	0.42	0.53	1.19	0.52	0.63
Pendleton	1.70	1.07	1.71	0.58	0.20	1.21	0.18	0.75
Baker	0.13	0.91	0.13	0.55	0.05	1.44	0.04	0.97
Yakima	0.05	1.29	0.05	0.74	0.02	1.48	0.02	0.96

Table 5.3. Mean and Coefficient of Variation of 100 Predicted Doses from Cow's Milk Containing Iodine-131 - 1945

<u>Location</u>	<u>Regime 1</u>				<u>Regime 3</u>			
	<u>Individual Cow</u>		<u>Average Cow</u>		<u>Individual Cow</u>		<u>Average Cow</u>	
	<u>Mean</u>	<u>CV</u>	<u>Mean</u>	<u>CV</u>	<u>Mean</u>	<u>CV</u>	<u>Mean</u>	<u>CV</u>
Eltopia	4.566	1.327	4.712	1.148	0.993	1.383	1.042	1.528
Richland	3.002	1.390	3.069	1.146	0.560	1.385	0.544	1.109
Ritzville	0.940	1.378	0.936	0.942	0.163	1.348	0.162	1.080
Benton City	0.863	1.405	0.814	1.037	0.205	1.639	0.202	1.312
Spokane	0.320	1.341	0.332	1.114	0.072	1.230	0.075	1.132
Pendleton	0.224	1.587	0.227	1.003	0.023	1.120	0.023	1.195
Baker	0.020	1.454	0.021	1.253	0.007	2.148	0.006	1.466
Yakima	0.008	1.780	0.007	1.214	0.002	1.543	0.002	1.267

Figure 5.3. Boxplots of the Annual Dose to a Reference Adult Male from Ingestion of Cow's Milk Containing Iodine-131 - 1945

Table 5.4. 5th and 95th Subjective Confidence Limits Giving 90 Percent Subjective Confidence Interval for Dose - 1945

Location	Regime 1				Regime 3			
	Individual Cow		Average Cow		Individual Cow		Average Cow	
	5th	95th	5th	95th	5th	95th	5th	95th
Eltopia	0.331	17.75	0.680	12.38	0.0553	4.407	0.1250	2.517
Richland	0.190	10.32	0.489	8.68	0.0310	2.087	0.0919	1.590
Ritzville	0.069	3.38	0.152	2.64	0.0089	0.563	0.0249	0.472
Benton City	0.048	3.52	0.110	2.56	0.0107	0.894	0.0225	0.666
Spokane	0.024	1.16	0.052	0.94	0.0038	0.276	0.0073	0.270
Pendleton	0.016	0.69	0.036	0.59	0.0017	0.084	0.0032	0.067
Baker	0.0011	0.078	0.0023	0.056	0.00017	0.0254	0.00039	0.029
Yakima	0.0003	0.026	0.0007	0.022	0.00011	0.0079	0.00020	0.005

Tables 5.3 and 5.4 and Figure 5.3 indicate that

- predicted dose is more than four times greater for feeding Regime 1 as compared to feeding Regime 3 (see Table 5.3)
- uncertainty in the predicted dose from the individual cow's milk is much larger than for the average cow's milk (see Table 5.4 and Figure 5.3)
- uncertainty increases as the location is further removed from the source and the major plume (see Figure 5.3)
- It can be misleading to use CV as a measure of the uncertainty in the predicted dose using the Monte Carlo technique with only 100 samples. This is because of the highly skewed nature of the predictions and the sensitivity of the variance (or standard deviation) estimate to unusually large predictions. For example, in Table 5.3, at Eltopia for Regime 3, the CV for the predicted dose from the individual cow (1.383) is smaller than the CV for the average cow (1.528). This is due primarily to the maximum prediction. However, as seen in Table 5.4, the 90 percent subjective confidence interval at Eltopia for Regime 3 is 0.0553 to 4.407 rad for the individual cow and 0.1250 to 2.517 rad for the average cow. Using the subjective confidence interval, the measure of uncertainty is now consistent with the fact that average cow's milk has less uncertainty in the concentration of iodine-131 than the individual cow's milk.

5.2.3 Sensitivity Analysis

To examine the sensitivity of the uncertainty in the predicted dose to the uncertainties in the parameters, the Coefficient of Partial Determinations (CPDs) and Standardized Regression Coefficients (SRCs) are used (see Section 4.2). The multiple regression models used the annual doses as the dependent variables and the following parameters as independent variables: average concentration of iodine-131 in milk during 1945, dose factor, intake, and holdup. The multiple regression was

performed on both the actual variables and on the ranks of the variables for each location. For all multiple regressions, the ranks had a larger coefficient of determination. Thus, the results for the ranks will be used in the sensitivity analysis.

Figures 5.4 through 5.7 show the rank CPDs and the rank SRCs for annual dose for each of the four combinations of cow type and feeding regimes, respectively, at the eight locations. Figures 5.4 and 5.5 show the sensitivity measures for annual dose from individual cow's milk, feeding Regimes 1 and 3, respectively. Figures 5.6 and 5.7 show the sensitivity measures for annual dose from average cow's milk, feeding Regimes 1 and 3, respectively. As seen in these figures:

- For milk from the individual cow, the uncertainty in the predicted dose is most sensitive to the uncertainty in concentration of iodine-131 in milk. The dose factor is the second most sensitive parameter (see Figures 5.4 and 5.5).
- For milk from the average cow, the uncertainty in the predicted dose is most sensitive to the uncertainty in the dose factor for all the locations except those outside the major plume (Baker, Regime 3 only, and Yakima) (see Figures 5.6 and 5.7).
- For milk from the average cow, the sensitivity of the uncertainty in dose to the uncertainty in the concentration of iodine-131 becomes larger as the location is further from the major iodine-131 plume.
- The uncertainty in the predicted dose is not sensitive to the uncertainty in the holdup time. The negative sign for the SRCs indicates the predicted dose decreases as the holdup time increases.
- For the milk from the individual cow, intake of milk is the third most sensitive parameter. However, for the average cow in the major plume, the uncertainty in the intake of milk is the second most sensitive parameter, ahead of concentration of iodine-131 in milk. (The uncertainty in intake is an arbitrary choice for this demonstration.) The uniform distribution with a minimum of 30 liters and a maximum of 170 liters corresponds to an intake of between 50 and 500 milliliters (about one-half quart) of milk per day. The uncertainty, due to an almost complete lack of knowledge about the total consumption of milk, is important to the uncertainty in dose to reference adult males consuming milk from an average cow.
- The order of the most sensitive parameters depends on the combination of cow type and feeding regime that produced the milk. As the uncertainty of the concentration of iodine-131 decreases, the importance of the uncertainty in the dose factor and milk intake increases.

5.2.4 Summary of Results from the Uncertainty and Sensitivity Analyses

The key conclusions from the uncertainty and sensitivity analyses of the dose for the demonstration are as follows:

- Uncertainty in the dose factor and the concentration of iodine-131 in milk are the major contributors to the uncertainty in dose.
- Uncertainty in the predicted doses is large. The upper limit of the shortest 90 percent subjective confidence interval is more than 17 times larger than the lower limit.

Figure 5.4. Sensitivity Measures for Annual Dose from Ingestion of Milk from Individual Cow, Feeding Regime 1

Figure 5.5. Sensitivity Measures for Annual Dose from Ingestion of Milk from Individual Cow, Feeding Regime 3

Figure 5.6. Sensitivity Measures for Annual Dose from Ingestion of Milk from Average Cow, Feeding Regime 1

Figure 5.7. Sensitivity Measures for Annual Dose from Ingestion of Milk from Average Cow, Feeding Regime 3

- Uncertainty in the quantity of milk intake can also be a major contributor to the uncertainty in dose.
- Uncertainty in the dose is insensitive to the uncertainty in the holdup time for the backyard cow pathway, which ranges only from one to three days.

The key implications of these uncertainty and sensitivity analyses of the dose for the demonstration are as follows:

- Choice of the PDF which describes the uncertainty in the dose factor should be thoroughly examined.
- Choice of the PDF for the intake of milk by the specified reference individual, in this case the reference adult male, should be thoroughly examined.

The key implications this demonstration has on the Uncertainty and Sensitivity Analyses Plan for the HEDR Project are:

- Multiple regression technique used in the sensitivity analysis works very well for this model, as demonstrated by the large coefficients of determination for each regression at each location.
- Subjective confidence interval should be the primary measure of the uncertainty in the dose predictions.
- Uncertainty and sensitivity analyses of the annual dose to the thyroid of the reference adult male from ingestion of milk containing iodine-131 need to continue to the second tier, that is, to uncertainty and sensitivity analyses of the concentration of iodine-131 in cow's milk (see Section 5.3).

5.3 Uncertainty and Sensitivity Analyses of Iodine-131 Concentration in Milk

The uncertainty and sensitivity analyses of iodine-131 concentration in milk is the second tier in the hierarchical uncertainty and sensitivity analyses of the 1945 annual dose to the thyroid of the reference adult male from ingestion of cow's milk containing iodine-131. The model and its parameters are described in Section 5.3.1. The uncertainty analysis is given in Section 5.3.2, followed by the sensitivity analysis in Section 5.3.3. The key conclusions of the uncertainty and sensitivity analyses are summarized in Section 5.3.4.

5.3.1 Model and Parameters

The model is

$$Concentration_{milk} = \frac{TF_{cowType} \times \sum_{day=1}^{365} \sum_{foodType=1}^f (Concentration_{day, foodType} \times Intake_{day, foodType})}{365}$$

where $Concentration_{milk}$ = the average concentration of iodine-131 in milk for a given type of cow and feeding regime in 1945, Ci / L_{milk}

$TF_{cowType}$ = Transfer factor, the fraction of the cow's daily intake of activity that can be found in a liter of milk, $(Ci / L_{milk}) / (Ci_{ingested})$

$Concentration_{day, foodType}$ = the concentration of iodine-131 in food type f on a given day in 1945, $Ci / Kg_{foodType}$

$Intake_{day, foodType}$ = amount of food type f consumed on a given day, $Kg_{foodType}$.

The concentration of iodine-131 in milk is predicted for the four different combinations of cow type and feeding regime at the eight different locations described in Section 5.1. A total of 100 predictions are calculated using the Monte Carlo technique by DESCARTES (PILOT). Each prediction is actually a vector of 365 daily concentrations. The concentration is calculated by averaging the 365 daily concentrations within each prediction vector.

The uncertainties in the feed-to-milk transfer factors for each cow type are described by a PDF, which is documented in Snyder et al. (1992). For the individual cow, a truncated lognormal distribution is used with a median of 0.0092 (mean of 0.012), standard deviation of 0.01, minimum (0.1 percentile) of 0.00093, and a maximum (99.9 percentile) of 0.091. For the average cow, a truncated normal distribution is used with a mean of 0.012, standard deviation of 0.002, minimum (0.1 percentile) of 0.006, and a maximum (99.9 percentile) of 0.018. The mean and standard deviation are for the non-truncated distributions. Stratified random sampling is used to obtain the 100 samples.

The uncertainties in the daily intake of the different food types for each of the feeding regimes are described by constrained triangular PDFs, which are documented in Beck et al. (1992). The soil consumption is documented in Snyder et al. (1992). Table 5.5 shows the minimum, mode, and maximum for the triangular distribution for each food type and the minimum and maximum within which the total daily consumption of plant products is constrained. Table 5.6 shows the mean and CV of the annual consumption of each food type at Eltopia for the 100 samples used in the Monte Carlo run. The total annual consumption of each food type for Regime 3 was the same for each location. However, for Regime 1 the total annual consumption of each food type is different for each location. This is because the food consumption changes by seasons, which are different for each location.

The iodine-131 concentration in each of the food types is generated by DESCARTES (PILOT). Alfalfa is harvested three times a year. Grain and silage is harvested once. Table 5.7 shows the mean and CV for each of the harvested food types, at harvest, at each location. After harvest, the concentration of iodine-131 decays daily. As a result, the quantity of iodine-131 present on a given day is 91.76 percent of that present on the previous day. For alfalfa, at the first harvest, the cow begins eating only from the first harvest. At the second harvest, the cow begins eating half from the second harvest and half from the first harvest. At the third harvest, the cow begins eating one-third from each harvest. Newly harvested grain and silage is not mixed with the previous harvest (from the last year). The mean and CV of the average daily concentration of iodine-131 in pasture, during pasture season, and soil at each location are shown in Table 5.8.

Table 5.5. Minimum, Mode, and Maximum of Triangular PDFs for Daily Consumption (kg/day)
Feeding Regime

REGIME 1

<u>Season</u>	<u>Food Type</u>	<u>Minimum</u>	<u>Mode</u>	<u>Maximum</u>
winter	grain	0.00	1.50	10.0
	alfalfa	5.00	9.50	15.0
	soil	1.00	2.00	4.0
spring	grain	0.00	1.50	10.0
	alfalfa	2.00	5.25	10.0
	pasture	1.00	4.25	10.0
	soil	0.50	1.00	2.0
summer	grain	0.00	1.50	10.0
	alfalfa	0.00	1.00	5.0
	pasture	2.00	8.50	15.0
	soil	0.25	0.50	1.0
early fall	grain	0.00	1.50	10.0
	alfalfa	2.00	5.25	10.0
	pasture	1.00	4.25	10.0
	soil	0.50	1.00	2.0
late fall	grain	0.00	1.50	10.0
	alfalfa	2.00	5.25	10.0
	silage	1.00	4.25	10.0
	soil	1.00	2.00	4.0
TOTAL		8.00		20.0

REGIME 3

<u>Food Type</u>	<u>Minimum</u>	<u>Mode</u>	<u>Maximum</u>
grain	0.00	1.00	10.0
alfalfa	5.00	9.00	15.0
soil	1.00	2.00	4.0
TOTAL	7.0		15.0

Winter - January 1 to start of pasture season

Spring - first two weeks of pasture season

Early Fall - last two weeks of pasture season

Late Fall - end of pasture season to December 31

Table 5.6. Mean and Coefficient of Variation of the 100 Annual Consumption Values at Eltopia (kg/year)

<u>Food Type</u>	<u>Regime 1</u>		<u>Regime 3</u>	
	<u>Mean</u>	<u>CV</u>	<u>Mean</u>	<u>CV</u>
Pasture	1181	0.056		
Alfalfa	864	0.032	2211	0.014
Grain	497	0.028	453	0.024
Silage	377	0.101		
Soil	615	0.022	877	0.004

Table 5.7. Mean and Coefficient of Variation of 100 Predicted Concentrations of Iodine-131 in Harvested Food (nCi/kg)^(a) - 1945

<u>Alfalfa Location</u>	<u>1st Harvest</u>		<u>2nd Harvest</u>		<u>3rd Harvest</u>	
	<u>Mean</u>	<u>CV</u>	<u>Mean</u>	<u>CV</u>	<u>Mean</u>	<u>CV</u>
Eltopia	1030.0	0.594	609	0.919	309.0	0.805
Richland	451.0	0.536	388	1.044	368.0	1.246
Ritzville	100.0	0.578	58	0.978	331.0	0.783
Benton City	278.0	0.762	15	2.468	55.1	1.267
Spokane	42.0	1.090	67	1.015	50.0	1.507
Pendleton	21.0	1.130	11	2.087	17.9	1.144
Baker	0.3	1.912	10	1.276	5.8	1.689
Yakima	3.0	1.177	0.00006	1.292	0.3	1.041

<u>Location</u>	<u>Grain</u>		<u>Silage</u>	
	<u>Mean</u>	<u>CV</u>	<u>Mean</u>	<u>CV</u>
Eltopia	652	0.877	765.0	0.945
Richland	359	0.835	411.0	0.838
Ritzville	86	1.014	428.0	1.120
Benton City	21	1.321	53.7	1.595
Spokane	95	0.598	89.8	0.758
Pendleton	17	1.176	42.2	0.854
Baker	13	1.026	1.9	1.125
Yakima	0.001	1.465	0.4	0.950

(a) The sum of the inner and outer vegetation compartments.

Table 5.8. Mean and Coefficient of Variation of 100 Predicted Average Daily Concentrations of Iodine-131 in Pasture and Soil (nCi/kg)^(a) - 1945

	Pasture		Soil	
	Mean	CV	Mean	CV
Eltopia	753.7	0.392	10.71	0.372
Richland	501.4	0.389	10.68	0.371
Ritzville	189.7	0.446	2.74	0.359
Benton City	148.0	0.462	2.93	0.424
Spokane	84.5	0.391	1.15	0.359
Pendleton	49.9	0.560	1.13	0.348
Baker	3.5	0.541	0.04	0.436
Yakima	1.4	0.813	0.09	0.478

(a) Pasture is the sum of the inner and outer vegetation compartments and soil is a mixture of 10 percent upper soil layer and 90 percent root zone.

The range in the harvest dates, and the start and end dates of the pasture seasons are shown in Figure 5.2. The uncertainties in these parameters are described by triangular PDFs. Except for Spokane and Baker, these dates are documented in Beck et al. 1992. The dates for Spokane and Baker were extrapolated from the information about the locations closest to them in Beck et al. (1992). Although these dates are not explicitly included in the above model, they do trigger the change of food types for feeding Regime 1 and the concentration of iodine-131 in the harvested food types are dependent on these dates.

5.3.2 Uncertainty Analysis

The mean and CV for the 100 predicted average daily concentrations of iodine-131 in cow's milk for the four different combinations of cow type and feeding regime are shown in Table 5.2. Figure 5.8 shows the boxplots for the predicted concentrations. The boxplots show the key percentiles from the subjective CDF of the predicted concentration, 5th, 25th, 50th (median), 75th, and 95th. Additionally, the estimated mean of the predicted concentration is included in the boxplot, as are the minimum and maximum predictions.

Table 5.9 shows the 5 percent and 95 percent subjective confidence limits (the 5th and 95th percentiles from the subjective CDF) which give a 90 percent subjective confidence interval. For example, at a 90 percent subjective confidence level, the true average daily concentration of iodine-131 in an individual cow's milk at Eltopia, on feeding Regime 1, is predicted as between the 6.1 and 97.3 nCi/L. Alternatively, a 95 percent subjective confidence level, the true average daily concentration of iodine-131 in the milk from an individual cow at Eltopia, on feeding Regime 1, is predicted as less than 97.3 nCi/L.

Table 5.9. 5th and 95th Subjective Confidence Limits That Give a 90 Percent Subjective Confidence Interval for Average Daily Concentration of Iodine-131 in Milk (nCi/L) - 1945

Location	Regime 1				Regime 3			
	Individual Cow		Average Cow		Individual Cow		Average Cow	
	5th	95th	5th	95th	5th	95th	5th	95th
Eltopia	6.10	97.3	14.47	52.3	1.077	21.71	2.336	14.36
Richland	3.84	65.6	8.94	38.8	0.578	15.29	1.377	8.28
Ritzville	1.11	20.6	2.68	12.2	0.194	4.04	0.413	2.48
Benton City	0.74	19.6	2.40	10.7	0.196	5.24	0.444	3.04
Spokane	0.40	7.2	0.98	4.1	0.085	1.73	0.187	1.24
Pendleton	0.31	4.5	0.60	3.6	0.029	0.74	0.064	0.48
Baker	0.02	0.3	0.04	0.3	0.004	0.19	0.005	0.13
Yakima	0.01	0.2	0.01	0.1	0.002	0.07	0.003	0.04

Tables 5.2 and 5.9 and Figure 5.8 indicate the

- predicted average daily concentration of iodine-131 in cow's milk is approximately four to five times greater for Regime 1 as for Regime 3 (see Table 5.2).
- uncertainty, as measured by the 90 percent subjective confidence intervals shown in Table 5.9, in the predicted average daily concentration of iodine-131 in cow's milk is smaller for Regime 1 than for Regime 3.
- difference in the uncertainty in the predicted average daily concentration of iodine-131 in cow's milk between Regime 1 and 3 increases with the magnitude of the uncertainty, which increases as the location is further removed from the major plume. The boxplots show the 5th, 25th, 50th (median), 75th, and 95th percentiles, as well as the minimum, maximum, and mean of 100 predictors. Figure 5.8 shows four boxplots for the four different combinations:
 - Individual cow on feeding Regime 1
 - Individual cow on feeding Regime 3
 - Average cow on feeding Regime 1
 - Average cow on feeding Regime 3.
- uncertainty in the predicted average daily concentration of iodine-131 in individual cow's milk is much larger than the uncertainty for the average cow's milk (see Figure 5.8).

5.3.3 Sensitivity Analysis

To examine the sensitivity of the uncertainty in the predicted iodine-131 concentration in cow's milk to the uncertainties in the parameters, the CPDs and SRCs are used, see Section 4.3. The

Figure 5.8. Boxplots of Average Daily Iodine-131 Concentration in Cow's Milk - 1945

multiple regression models used the average concentration of iodine-131 in milk as the dependent variable and the following parameters as independent variables: feed-to-milk transfer factor, iodine-131 concentration in pasture (for Regime 1 only) averaged over the pasture season, alfalfa concentration of iodine-131 at each of the three harvests, grain concentration of iodine-131 at harvest, silage concentration of iodine-131 at harvest (for Regime 1 only), annual integrated soil concentration of iodine-131, annual intake of pasture (for Regime 1 only), annual intake of alfalfa, annual intake of grain, annual intake of silage (for Regime 1 only), and annual intake of soil. The multiple regression was performed on both the actual variables and on the ranks of the variables. For the individual cow, the multiple regressions using the ranks generally had a slightly higher coefficient of determination. Thus, the results for the ranks will be used for the individual cow. The opposite was true for the average cows.

Figures 5.9 through 5.12 show the CPDs and SRCs for each of the four combinations of cow type and feeding regimes, respectively, at the eight locations. The figures include only those parameters that had a CPD of 0.1 or greater for at least one of the eight sites. As seen in these figures

- For the individual cow, the uncertainty in the concentration of iodine-131 in milk is most sensitive to the uncertainty in the feed-to-milk transfer factor (see Figures 5.9 and 5.10). However, this is not true for the average cow. For the average cow, the uncertainty in the concentration of iodine-131 in milk is most sensitive to the uncertainty in the concentration of iodine-131 in the plant products they consume. (See Figures 5.11 and 5.12.)
- The sensitivity of concentration of iodine-131 in milk to the uncertainty in the transfer factor is greatest for the locations within the major plume, where the uncertainties in the concentration of iodine-131 in the different food types the cow consumed are the smallest (see Figures 5.9 through 5.12).
- For the cow on feeding Regime 1, either individual or average, the uncertainty in the concentration of iodine-131 in milk is insensitive to the uncertainties in the concentration of iodine-131 in the harvested food types, alfalfa, grain, and silage (see Figures 5.9 and 5.11). That is, the uncertainty in milk is sensitive only to the pasture.
- For both individual and average cows on feeding Regime 3, the uncertainty in iodine-131 concentration in milk is generally most sensitive to the iodine-131 concentration in the first harvest alfalfa. (See Figures 5.10 and 5.12.) However, this is not true for all locations. For example at Ritzville, the third harvest alfalfa is the most sensitive food type. This is due to the large surface deposition of iodine-131 that occurred during the two-week period when the third harvest of alfalfa also occurred, see Figure 5.2.
- For the cow on feeding Regime 3, especially the average cow (see Figure 5.12), the number of food types the uncertainty in milk concentration is sensitive to depends on the location. As seen in Figure 5.2, each location has a different surface deposition pattern over time, and the relationship between surface deposition and harvest dates are different at each location.

Figure 5.9. Sensitivity Measures for Average Daily Iodine-131 Concentration in Milk from Individual Cow, Feeding Regime 1 - 1945

Figure 5.10. Sensitivity Measures for Average Daily Iodine-131 Concentration in Milk from Individual Cow, Feeding Regime 3 - 1945

Figure 5.11. Sensitivity Measures for Average Daily Iodine-131 Concentration in Milk from Average Cow, Feeding Regime 1 - 1945

Figure 5.12. Sensitivity Measures for Average Daily Iodine-131 Concentration in Milk from Average Cow, Feeding Regime 3 - 1945

5.3.4 Summary of Results from the Uncertainty and Sensitivity Analyses

The key conclusions from the uncertainty and sensitivity analyses of the dose for the demonstration are:

- The uncertainty in the predicted concentration of iodine-131 in the milk of an individual cow is large, the upper limit of the shortest 90 percent subjective confidence interval is more than 17 times larger than the lower limit.
- The uncertainty in the predicted concentration of iodine-131 in milk of the average cow is much smaller; the upper limit for many of the 90 percent subjective confidence intervals is less than five times larger than the lower limit.
- The uncertainty in the feed-to-milk transfer factor is the major contributor to the uncertainty in the concentration of iodine-131 in milk from an individual cow.
- The importance of the uncertainty in the harvested food types in feeding Regime 3 depends on the location.
- The uncertainty in the predicted concentration of iodine-131 in milk is insensitive to the uncertainty in the harvested food types in feeding Regime 1.
- The uncertainty in the concentration of iodine-131 in cow's milk is insensitive to the uncertainty in the amount of each food type the cow consumes, the amount of soil the cow consumes, and the concentration of iodine-131 in soil.

The key implications of this uncertainty and sensitivity analyses of the dose for the demonstration are:

- The choice of the PDF which describes the uncertainty in the feed-to-milk transfer factor for the individual cow should be thoroughly examined. Additionally, the transfer factor for the average cow depends on the assumption that transfer factors for individual cows are independent. Because such a large reduction is found in the uncertainty for the average cow versus the individual cow, the independence assumption should also be thoroughly examined.
- If the feed-to-milk transfer factor for the average cow is indeed valid, the number of cows an individual has is a very important factor. That is, the backyard cow pathway needs to take into account whether the backyard cow is a single cow or a small herd of cows.

The key implications this uncertainty and sensitivity analyses of the dose for the demonstration has on the Uncertainty and Sensitivity Analyses Plan are:

- The multiple regression technique used in the sensitivity analysis works very well for this model, as demonstrated by the large coefficients of determination for each regression at each location.
- The uncertainty and sensitivity analyses of the annual dose to the thyroid of the reference adult male from ingestion of milk containing iodine-131 needs to continue to the second tier: uncertainty and sensitivity analyses of the concentration of iodine-131 in pasture (see Section 5.4) and harvested food types (see Section 5.5).

5.4 Uncertainty and Sensitivity Analyses of Iodine-131 Concentration in Pasture

The uncertainty and sensitivity analyses of iodine-131 concentration in pasture is part of the third tier in the hierarchical uncertainty and sensitivity analyses of the 1945 annual dose to the thyroid of the reference adult male from ingestion of cow's milk containing iodine-131. In this section, first the model and its parameters are described in Section 5.4.1. Then the uncertainty analysis is given in Section 5.4.2, followed by the sensitivity analysis in Section 5.4.3. The key conclusions of the uncertainty and sensitivity analyses are summarized in Section 5.4.4.

5.4.1 Model and Parameters

The conceptual model for the concentration of iodine in pasture is shown in Figures 2.2 and 2.3. The coupled differential equations used in the computational model are shown in Snyder et al. (1992).

The parameters used in these coupled differential equations and the PDFs which describe the uncertainty in each parameter are documented in Snyder et al. (1992), and are shown in Table 5.10. The parameters in Table 5.10 were sampled using stratified random sampling.

The start and end dates of the pasture season are documented in Beck et al. (1992) for all the locations, except Spokane and Baker which were extrapolated from the closest locations. The start and end dates are shown in Figure 5.2 and have triangular PDFs. They are also sampled using stratified random sampling.

The estimated daily surface deposition of iodine-131, at each location, is generated by RATCHET. The average weekly surface depositions at each location are shown in Figure 5.2. The average and CV of the 100 predictions of total annual surface deposition of iodine-131 are shown in Table 5.1. The average and CV of the predicted total surface deposition and daily surface deposition during the pasture season are shown in Table 5.11.

A total of 100 predictions of concentration of iodine-131 in pasture are calculated using the Monte Carlo technique by DESCARTES (PILOT). Each prediction is actually a vector of 365 daily concentrations. Because the concentration of iodine-131 in pasture only during the pasture season are used in predicting the concentration of iodine-131 in milk, the average daily concentration of iodine-131 in pasture during the pasture season is used in the uncertainty and sensitivity analyses. This is calculated by averaging the concentration of pasture during the pasture season for each of the 100 prediction vectors (for each prediction, the start and end dates of the pasture season are different).

5.4.2 Uncertainty Analysis

The mean and CV for the 100 predicted average daily concentrations of iodine-131 in pasture are shown in Table 5.8. Figure 5.13 shows the boxplots for the predicted concentrations. The boxplots show the key percentiles from the subjective CDF of the predicted concentration, 5th, 25th, 50th (median), 75th, and 95th. Additionally, the mean of the predicted concentration is included in the boxplot and the minimum and maximum predictions.

Table 5.12 shows the 5 percent and 95 percent subjective confidence limits (the 5th and 95th percentiles from the subjective CDF) which give a 90 percent subjective confidence interval. For

Table 5.10. PDFs of Parameters Used in Coupled Differential Equations for the Concentration of Iodine-131 in Pasture^(a)

<u>Parameter</u>	<u>Units</u>	<u>PDF</u>	<u>Minimum</u>	<u>Maximum</u>	
Alpha, for interception	m ² /kg(dry)	Uniform	1.0	4.0	
Leaching rate	day ⁻¹	Uniform	4.0E-6	5.0E-3	
Percolation rate	day ⁻¹	Uniform	0.14	8.2	
Soil density, root zone	kg(wet)/m ²	Uniform	186	230	
Soil density, upper soil	kg(wet)/m ²	Uniform	1.10	1.45	
Deposition velocity	m/d	Uniform	86.4	2590	
Concentration ratio	Ci/kg(dry) of veg per Ci/kg(wet) of soil	Loguniform	0.01	0.25	
			<u>Minimum</u>	<u>Mode</u>	<u>Maximum</u>
Weathering rate	day ⁻¹	Triangular	0.0347	0.0495	0.0866
Maximum biomass	kg(dry)/m ²	Triangular	0.1	0.3	0.7
			<u>Minimum</u>	<u>Median</u>	<u>Maximum</u>
Mass loading	kg(dry)/m ³	Lognormal	8.0E-9	7.0E-08	6.0E-7
			<u>Value</u>		
Radiological decay rate	day ⁻¹	Constant	0.086		
Rainsplash rate	day ⁻¹	Constant	0.0		
Translocation parameter	day ⁻¹	Constant	0.0		
Minimum biomass	kg(dry)/m ²	Constant	0.04		
Growth rate	day ⁻¹	Constant	0.12		
Senescence rate	day ⁻¹	Constant	0.09		

(a) Snyder et al. (1992).

Table 5.11. Average and Coefficient of Variation of 100 Predicted Total and Daily Surface Deposition of Iodine-131 During the Pasture Season (nCkg/m²)

<u>Location</u>	<u>Total</u>		<u>Daily</u>	
	<u>Mean</u>	<u>CV</u>	<u>Mean</u>	<u>CV</u>
Eltopia	14700	0.232	81.94	0.216
Richland	10350	0.237	57.75	0.223
Ritzville	3090	0.268	20.07	0.271
Benton City	2887	0.369	16.5	0.343
Spokane	1034	0.225	9.04	0.193
Pendleton	1034	0.448	6.23	0.425
Baker	61	0.413	0.42	0.397
Yakima	25	0.782	0.15	0.736

Figure 5.13. Boxplots of the Average Daily Concentration of Iodine-131 in Pasture During Pasture Season - 1945

Table 5.12. 5th and 95th Subjective Confidence Limits Which Give a 90 Percent Subjective Confidence Interval for Predicted Average Daily Concentration of Iodine-131 in Pasture (nCi/kg(dry))

<u>Location</u>	<u>5th</u>	<u>95th</u>
Eltopia	330.0	1210.0
Richland	190.0	837.5
Ritzville	78.9	351.0
Benton City	62.8	260.2
Spokane	38.4	137.4
Pendleton	16.8	98.1
Baker	1.3	7.3
Yakima	0.3	3.4

example, at a 90 percent subjective confidence level, the true average daily concentration of iodine-131 in pasture at Eltopia is predicted as between the 330 and 1210 nCi/kg. Alternatively, at a 95 percent subjective confidence level, the true average daily concentration of iodine-131 in pasture grass at Eltopia is predicted as less than 1210 nCi/kg.

As seen in Figure 5.13, the uncertainty in the predicted average daily concentration of iodine-131 in pasture increases as the location moves further from the major plume.

5.4.3 Sensitivity Analysis

To examine the sensitivity of the uncertainty in the predicted average daily concentration of iodine-131 in pasture to the uncertainties in the parameters, the CPDs and SRCs are used, see Section 4.2. The multiple regression models used the non-constant parameters listed in Table 5.10, the start and end days of the pasture season, and the predicted average daily surface deposition of iodine-131 during the growing season. The multiple regressions are performed on both the actual variables and on the ranks of the variables. The coefficients of determination were about the same for both the actual variables and the ranks. Thus, the results using the actual variables are presented.

A second set of multiple regressions is done using the predicted total annual deposition instead of the average daily deposition during the pasture season. The average daily deposition during the pasture season is a function of the start and end days of the pasture season. Therefore, to examine if the start and end of the pasture season are important parameters, the deposition parameter must be independent of the start and end days. These multiple regressions are also performed on both the actual variables and on the ranks of the variables. The coefficients of determination were about the same for both the actual variables and the ranks. Thus, the results using the actual variables are presented.

Figures 5.14 and 5.15 show the CPDs and SRCs at the eight locations from the multiple regressions using average daily deposition during pasture season and total annual deposition, respectively. The figures include only those parameters that had a CPD of 0.1 or greater for at least one of the eight sites. As seen in these figures:

- When average daily deposition during pasture season is one of the independent variables, the uncertainty in the concentration of iodine-131 in pasture is most sensitive to the uncertainties in daily surface deposition of iodine-131 and the alpha parameter from the interception function. Which of these parameters is more important depends on the location. (See Figure 5.14. Note this figure is from multiple regression using average deposition during pasture season as an independent variable.)
- When average daily deposition during pasture season is one of the independent variables, the uncertainty in the concentration of iodine-131 in pasture is only marginally sensitive to the uncertainty in the end date of the pasture season at Pendleton.
- When total annual deposition is one of the independent variables, the uncertainty in the concentration of iodine-131 in pasture is sensitive to the uncertainties in total surface deposition of iodine-131 and the alpha parameter from the interception function. However, the uncertainty in the concentration of iodine-131 in pasture is also sensitive to the uncertainty in the end of the pasture season, at all the locations except Ritzville, Spokane, and Baker. (See Figure 5.15. In contrast to Figure 5.14, this figure is from multiple regression using total deposition as an independent variable.)
- The uncertainty in the concentration of iodine-131 in pasture is also sensitive to the uncertainty in the maximum biomass parameter of pasture. The daily biomass has two functions in the model. First, it is part of the interception function. If the concentration of iodine-131 in pasture was sensitive to the biomass as part of the interception function, the SRC would have a positive sign, because as the biomass increases so does the amount of deposition intercepted by the pasture. The concentration of iodine-131 is the ratio of the quantity in the inner and outer vegetation divided by the biomass. In this case, as the biomass increases, the concentration decreases, and the SRC would have a negative sign. As seen in Figures 5.14 and 5.15, biomass has a negative sign, thus the uncertainty in concentration of iodine-131 in pasture is sensitive to the use of biomass in calculating the concentration from the quantity.
- The uncertainty in the concentration of iodine-131 in pasture is slightly sensitive to the uncertainty in weathering rate.

5.4.4 Summary of Results from the Uncertainty and Sensitivity Analyses

The key conclusions from the uncertainty and sensitivity analyses of the dose for the demonstration are:

- The uncertainty in the predicted average daily concentration of iodine-131 in pasture during pasture season is relatively small compared to the uncertainty in the dose to the thyroid due to ingestion of milk from cows eating the pasture. The upper limit for most of the 90 percent subjective confidence intervals of the predicted average daily concentration is less than five times larger than the lower limit. The shortest 90 percent subjective confidence intervals for predicted dose had a factor of 17 difference between the upper and lower limits.

Figure 5.14. Sensitivity Measures for Average Daily Concentration of Iodine-131 in Pasture - 1945

Figure 5.15. Sensitivity Measures for Average Daily Concentration of Iodine-131 in Pasture - 1945

- The uncertainty in the predicted concentration of iodine-131 in pasture is most sensitive to the uncertainties in the predicted surface deposition, the alpha parameter, end dates of the pasture season, the maximum biomass of pasture, and the weathering rate. The order of importance of these parameters changes with location.

The key implications of this uncertainty and sensitivity analyses of the dose for the demonstration are:

- The choice of the CDF which describes the uncertainty in the alpha parameter for interception should be examined.
- The choice of the CDF for maximum biomass and the function for calculating the daily biomass should be examined.
- The uncertainty in the end date of the pasture season is important. The uncertainty in the start date of the pasture season is not important because relatively little surface deposition of iodine-131 occurred at that time in 1945. This may not be the case in 1946 and later years. The choice of the CDF for the start and end dates of the pasture seasons should be examined.

The key implications that this uncertainty and sensitivity analyses of the dose for the demonstration have on the Uncertainty and Sensitivity Analyses Work Plan are:

- The multiple regression technique used in the sensitivity analyses works very well for this model, as demonstrated by the large coefficients of determination for almost all of the regressions at each location. The low coefficient of determination for the regression using total deposition at Yakima is due to the fact that most of the surface deposition occurred outside the pasture season. Thus, little relationship exists between the uncertainty in the total deposition and the uncertainty of the deposition that occurred during the pasture season.
- The multiple regression technique is limited in its ability to measure the sensitivity of the uncertainty of dose to parameters related to time, because the relationships are not necessarily linear or monotonic. For example, a Monte Carlo prediction with an early end date for pasture season could have as great a pasture concentration as another Monte Carlo prediction with a late date for the pasture season. This similarity in predictions is dependent on how episodic the deposition is during the period of time and how the dates of the episodes match with the dates of the end of the pasture season.
- The uncertainty and sensitivity analyses of the annual dose to the thyroid of the reference adult male from ingestion of milk containing iodine-131 needs to be continued to the surface deposition of iodine-131 (see Section 5.6).

5.5 Uncertainty and Sensitivity Analyses of Iodine-131 Concentration in Harvested Plant Products

The uncertainty and sensitivity analyses of iodine-131 concentration in harvested plant products is part of the third tier in the hierarchical uncertainty and sensitivity analyses of the 1945 annual dose to the thyroid of the reference adult male from ingestion of cow's milk containing iodine-131.

For this demonstration, only the uncertainty and sensitivity analyses of grain will be shown. In the actual uncertainty and sensitivity analyses, all the sensitive harvest food products would be analyzed. In this section, first the model and its parameters are described in Section 5.5.1. The uncertainty analyses is given in Section 5.4.2, followed by the sensitivity analysis in Section 5.5.3. The key conclusions of the uncertainty and sensitivity analyses are summarized in Section 5.5.4.

5.5.1 Model and Parameters

The conceptual model for the concentration of iodine in harvested plant products is shown in Figures 2.2 and 2.3. The coupled differential equations used in the computational model are shown in Snyder et al. (1992). This is the same model as the one used for pasture that was analyzed in Section 5.4. The difference between pasture and harvested plant products is the pasture is eaten fresh every day, while the harvested product contains the iodine-131 present on the harvest day, which decays away as time passes.

The parameters used in these coupled differential equations and the CDFs that describe the uncertainty in each parameter are documented in Snyder et al. (1992), and are shown in Tables 5.10 and 5.13. The parameters in Table 5.13 are specific for grain. The remaining parameters are the same as those used for pasture and shown in Table 5.10. The parameters were sampled using stratified random sampling.

The harvest dates are documented in Beck et al. (1992) for all the locations, except Spokane and Baker, which were extrapolated from the closest locations. The range of harvest dates are shown in Figure 5.2 and have triangular CDFs. They are also sampled using stratified random sampling.

The estimated daily surface deposition of iodine-131, at each location, is generated by RATCHET. The average weekly surface depositions at each location are shown in Figure 5.2. The average and CV of the 100 predictions of total annual surface deposition of iodine-131 are shown in Table 5.1.

A total of 100 predictions of concentration of iodine-131 in grain at harvest are calculated using the Monte Carlo technique by DESCARTES (PILOT).

Table 5.13. PDFs of Parameters Used in Coupled Differential Equations for the Concentration of Iodine-131 Which are Specific to Grain^(a)

<u>Parameter</u>	<u>Units</u>	<u>PDF</u>	<u>Minimum</u>	<u>Maximum</u>	
Translocation parameter	None	Loguniform	0.01	0.2	
			<u>Minimum</u>	<u>Mode</u>	<u>Maximum</u>
Maximum biomass	kg(dry)/m ²	Triangular	0.09	0.14	0.3
				<u>Value</u>	
Minimum biomass	kg(dry)/m ²	Constant		0.01	
Growth rate	day ⁻¹	Constant		0.12	
Senescence rate	day ⁻¹	Constant		0.08	

(a) Snyder et al. (1992).

5.5.2 Uncertainty Analysis

The mean and CV for the 100 predicted average daily concentrations of iodine-131 in grain are shown in Table 5.7. Figure 5.16 shows the boxplots for the predicted concentrations. The boxplots show the key percentiles from the subjective CDF of the predicted concentration, 5th, 25th, 50th (median), 75th, and 95th. Additionally, the mean of the predicted concentration is included in the boxplot and the minimum and maximum predictions.

Table 5.14 shows the 5th and 95th subjective confidence limits (the 5th and 95th percentiles from the subjective CDF) which give a 90 percent subjective confidence interval. For example, at a 90 percent subjective confidence level, the true concentration of iodine-131 in grain at Eltopia is predicted as between the 127 and 1942 nCi/kg. Alternatively, at a 95 percent subjective confidence level, the true concentration of iodine-131 in grain at Eltopia is predicted as less than 1942 nCi/kg.

As seen in Figure 5.16, the uncertainty in the predicted concentration of iodine-131 in grain increases when the location is at the edge of or outside the major plume. Note the concentration at Yakima is too small to show in this figure.

5.5.3 Sensitivity Analysis

To examine the sensitivity of the uncertainty in the predicted concentration of iodine-131 in grain at harvest to the uncertainties in the parameters, the CPDs and SRCs are used (see Section 4.2). The multiple regression models used the non-constant parameters listed in Tables 5.10 and 5.13, the harvest day, and the predicted weighted sum of the daily surface deposition of iodine-131 for the week prior to harvest.

The predicted weighted sum of the daily surface deposition of iodine-131 for the week prior to harvest is used as an independent variable because it is linearly related to the concentration of iodine-131 in grain at harvest, where the deposition on any given day prior to harvest is not. If surface deposition is constant, the most important deposition for grain would be the deposition on the day before harvest. However, when the deposition is episodic, for some realizations the deposition on the day before harvest would be important, but for other realizations other days would have a much greater deposition, and thus contribute far more to the concentration of iodine-131 in grain at harvest. The weighting was based on the radioactive decay rate for iodine-131. The weighted sum was the deposition the day before harvest, plus 91.6 percent of the deposition two days before harvest, plus 84.2 percent of the deposition three days before harvest, and so forth.

The multiple regressions are performed on both the actual variables and on the ranks of the variables. The coefficients of determination were about the same for both the actual variables and the ranks. Thus, the results using the actual variables are presented.

Figure 5.17 shows the CPDs and SRCs at the eight locations from the multiple regressions. The figure includes only those parameters that had a CPD of 0.1 or greater for at least one of the eight sites. As seen in these figures:

- The uncertainty in the concentration of iodine-131 in grain at harvest is most sensitive to the uncertainty in the weighted sum of surface deposition of iodine-131 the week before harvest at all the locations, except Yakima. As seen in Figure 5.2 very little surface

Figure 5.16. Boxplots of the Concentration of Iodine-131 in Grain at Harvest - 1945

Table 5.14. 5th and 95th Subjective Confidence Limits Which Give a 90 Percent Subjective Confidence Interval for Predicted Concentration of Iodine-131 in Grain at Harvest (nCi/kg(dry))

<u>Location</u>	<u>5th</u>	<u>95th</u>
Eltopia	127.2	1942
Richland	77.4	960
Ritzville	14.5	220
Benton City	1.7	87
Spokane	26.0	210
Pendleton	1.1	66
Baker	0.2	38
Yakima	5.1E-5	3.8E-3

deposition happened in Yakima for several weeks prior to grain harvest, so the weighted sum would have to extend that far to pick up the surface deposition of iodine-131 contributing to the concentration of iodine-131 in grain.

- The uncertainty in the concentration of iodine-131 in grain at harvest is also sensitive to the uncertainty in the alpha parameter for interception. However, grain is not as sensitive to this parameter as is pasture.
- The uncertainty in the concentration of iodine-131 in grain at harvest is not sensitive to the uncertainty in the other parameters.

5.5.4 Summary of Results from the Uncertainty and Sensitivity Analyses

The key conclusions from the uncertainty and sensitivity analyses of the dose for the demonstration are:

- The uncertainty in the predicted concentration of iodine-131 in grain at harvest is relatively large compared to the uncertainty in the predicted average daily concentration of iodine-131 in pasture (see Figures 5.13 and 5.16).
- The uncertainty in the predicted concentration of iodine-131 in grain is sensitive to the uncertainties in the predicted surface deposition a week prior to harvest and in the alpha parameter. The importance of harvest date is implied by the importance of the surface deposition a week prior to harvest.

The key implications of this uncertainty and sensitivity analyses of the dose for the demonstration are:

- The choice of the CDF which describes the uncertainty in the alpha parameter for interception should be examined. (This was also concluded in Section 5.4).

Figure 5.17. Sensitivity Measures for Concentration of Iodine-131 in Grain at Harvest - 1945

- Harvest dates of grain should be examined within the region of the HEDR study area where the dose to thyroid from ingestion of milk from cows on stored feed only feeding regimes is above the dose decision level. (The same conclusion will probably be made for alfalfa after uncertainty and sensitivity analyses are performed on the concentration of iodine-131 in alfalfa.)

The key implications that these uncertainty and sensitivity analyses of the dose for the demonstration has on the Uncertainty and Sensitivity Analyses Plan are:

- The multiple regression technique used in the sensitivity analysis works very well for this model, as demonstrated by the large coefficients of determination for almost all of the regressions at each location.
- The use of the weighted sum of the surface deposition prior to harvest, does linearize the relationship of deposition of iodine-131 with the concentration of iodine-131 in a harvested plant product. However, the importance of the harvest date is only implied by this method, not measured.
- The uncertainty and sensitivity analyses of the annual dose to the thyroid of the reference adult male from ingestion of milk containing iodine-131 needs to be continued on to the surface deposition of iodine-131 (see Section 5.6).

5.6 Uncertainty and Sensitivity Analyses of Surface Deposition Iodine-131

The uncertainty and sensitivity analyses of surface deposition of iodine-131 is part of the fourth tier in the hierarchical uncertainty and sensitivity analyses of the 1945 annual dose to the thyroid of the reference adult male from ingestion of cow's milk containing iodine-131. In this section, first the model and its parameters are described in Section 5.6.1. The uncertainty analysis is given in Section 5.6.2, followed by the sensitivity analysis in Section 5.6.3. The key conclusions of the uncertainty and sensitivity analyses are summarized in Section 5.6.4.

5.6.1 Model and Parameters

The RATCHET code (Ramsdell and Burk 1992) calculates the movement, diffusion, and deposition of iodine-131 released to the atmosphere by operations at the Hanford Site. Figure 5.18 shows the boxplots for the predicted depositions. The boxplots show the key percentiles from the subjective CDF of the predicted concentration, 5th, 25th, 50th (median), 75th, and 95th. Additionally, the mean of the predicted deposition is included in the boxplot and the minimum and maximum predictions.

Transport calculations are based on the effective release height, and the sequence of wind fields and mixing depths following release. The effective release height is determined by the meteorological conditions at the time of release. Thus, uncertainties in meteorological conditions and release times contribute to the uncertainty in release height. Uncertainties are also associated with the plume rise model used to determine effective release height. After material is released, its movement is controlled by the sequence of wind fields and mixing heights between the time of release and the time that the material leaves the model domain. Again, the uncertainties in meteorological conditions and release times contribute significantly to the uncertainty in movement. The factors are also

Figure 5.18. Boxplots of the Annual Surface Deposition - 1945

associated with uncertainty in diffusion and deposition. However, an additional factor contributes to the uncertainty in deposition. That factor is the partitioning of iodine among various forms (species). The uncertainties in release time are accounted for in RATCHET by using the source term realizations generated by the STRM code.

RATCHET treats the uncertainties in the meteorological conditions by adding random values to the wind directions and speeds reported at the meteorological stations. The random values are drawn from uniform distributions determined by the imprecision of recorded values. RATCHET also treats uncertainties in the determination of stability class, precipitation rate, and mixing height. Piece-wise uniform distributions, based on data, are used to account for uncertainties in stability class and

precipitation rate, and uniform distributions used to determine parameter values in the calculation of mixing layer depths. In each case, simple random sampling is used because the sampling is performed every hour for data entered for each meteorological station.

Random sampling was limited to the input meteorological data to minimize the potential for inducing improper correlations within RATCHET, based on an examination of the algorithms used to compute transport, diffusion, and deposition. The variability introduced by random sampling associated with the input meteorological data is estimated throughout RATCHET by use of the meteorological data.

The deposition characteristics of iodine depend on its chemical and physical form, and iodine exists in the atmosphere in many forms. RATCHET combines all of the atmospheric iodine forms into three groups. These groups are highly reactive gases (such as I_2), slightly reactive gases, and particulates. The true partitioning of iodine-131 among the species is unknown and may vary from day to day. However, information on iodine partitioning in the literature and in the HEDR Project meteorological database is not adequate to support modeling the temporal variation of iodine partitioning. As a result, the fraction of iodine in each partition is selected by constraining simple random sampling, and is held constant throughout each realization. The fraction ranges are based on data in the literature. Particulate and highly reactive iodine fractions are selected by sampling from uniform distributions between 0.1 and 0.5. The slightly reactive iodine fraction is determined by subtraction. If this fraction is less than 0.1, the slightly reactive fraction is determined by simple random sampling from a uniform distribution between 0.1 and 0.2. The particulate and highly reactive fractions are adjusted so the sum of the fractions equals 1.0.

The average weekly surface contamination at each location is shown in Figure 5.2. The average and CVs of the 100 predictions of total surface deposition of iodine-131 in 1945 are shown in Table 5.1.

5.6.2 Uncertainty Analysis

The mean and CV for the 100 predicted annual surface depositions of iodine-131 are shown in Table 5.1. Figure 5.18 shows the boxplots for the predicted depositions. The boxplots show the key percentiles from the subjective CDF of the predicted concentration, 5th, 25th, 50th (median), 75th, and 95th. Additionally, the mean of the predicted deposition is included in the boxplot and the minimum and maximum predictions.

Table 5.15 shows the 5th and 95th subjective confidence limits (the 5th and 95th percentiles from the subjective CDF) which give a 90 percent subjective confidence interval. For example, at a 90 percent subjective confidence level, the true annual surface deposition of iodine-131 at Eltopia is predicted as between the 14,650 and 27,340 nCi/m².

As seen in Figure 5.18, the uncertainty in the predicted total surface deposition of iodine-131 increases when the location is outside the major plume.

5.6.3 Sensitivity Analysis

To examine the sensitivity of the uncertainty in the annual surface deposition of iodine-131 to the uncertainties in the annual release of iodine-131 and the partitioning of iodine-131, the CPDs and SRCs are used, see Section 4.2.

Table 5.15. 5th and 95th Subjective Confidence Limits Which Give a 90 Percent Subjective Confidence Interval for Predicted Annual Surface Deposition of Iodine-131 (nCi/m²)

<u>Location</u>	<u>5th</u>	<u>95th</u>
Eltopia	14650	27340
Richland	14930	28080
Ritzville	3670	7067
Benton City	3983	8720
Spokane	1630	2689
Pendleton	1629	2641
Baker	35	120
Yakima	85	284

The multiple regressions are performed on both the actual variables and on the ranks of the variables. The coefficients of determination were about the same for both the actual variables and the ranks. Thus, the results using the actual variables are presented.

Figure 5.19 shows the CPDs and SRCs at the eight locations from the multiple regressions. As seen in these figures:

- The multiple regressions only explain 50 percent or more of the uncertainty in the predicted annual surface deposition of iodine-131 at Eltopia, Richland, and Spokane. Parameters, other than partitioning of iodine and the amount of iodine released, are responsible for a majority of the uncertainty in the predicted annual surface deposition at the other locations.
- In spite of the low coefficient of determination, a consistent pattern is found in the CPDs and SRCs across the locations. The variability in iodine partitioning directly affects doses by changing the amount of iodine available for deposition. However, the effects of changing the partitioning are not simple. For example, increasing the slightly reactive fraction reduces the deposition and ingestion pathway doses close to the Hanford Site. However, by reducing the deposition near the source, increasing the slightly reactive fraction increases the amount of iodine available for deposition at greater distances. This effect can be seen in the systematic change in SRC from negative to positive in Figure 5.19.
- The uncertainty in the annual surface deposition of iodine-131 is not very sensitive to the uncertainty in the total release of iodine-131 in that year. At Spokane and Pendleton, the source term has the same relative sensitivity as the partitioning parameters, but considering the coefficient of determination this sensitivity is not very strong. The results, or lack of results, should not be surprising after looking at Figure 5.2. As seen in this figure and in Table 5.1, for most of the locations, the total surface deposition is dominated by the deposition over a shorter time period. Thus, the uncertainty in the annual amount of iodine-131 released would not be related to the uncertainty in the surface deposition over a short period of time.

Figure 5.19. Sensitivity Measures for Annual Surface Deposition - 1945

The multiple regression technique is not very appropriate for the sensitivity analysis of RATCHET because it is suspected that a major contributor to the uncertainty in the surface deposition and air concentration predictions from RATCHET is the relationship between release time of the iodine-131 at the stacks (from STRM) and the meteorological conditions at that time. Thus, a sensitivity analysis based on the partitioning of the uncertainties by holding subsets of the parameters constant and measuring the reduction in the uncertainty for each subset should be used.

The uncertainty in the concentration of iodine-131 in the air integrated over the month of April 1945 is used to demonstrate the partitioning of the uncertainties. RATCHET was run for the following four different cases (50 realizations each):

- Normal, all the parameter uncertainties are included.
- No parameter uncertainties for the meteorological data included, only the uncertainty in the hourly amount of iodine-131 released (from STRM) is included.
- Uniform release of iodine-131 for the month, no hourly variation in the amount released, but the total amount released in April 1945 did change with each realization. The parameter uncertainties for the meteorological data are included.
- Uniform release of iodine-131 is used with no parameter uncertainties for the meteorological data.

Tables 5.16 and 5.17 show the average and CVs for the 50 predictions from these four RATCHET runs. The CVs indicate the effects of the uncertainty in the hourly releases of iodine-131 and the meteorological parameters on the uncertainty in the concentration of iodine-131 change with location.

Table 5.16. Average of 50 Predictions of Integrated Air Concentration of Iodine-131 ($\text{Ci}\cdot\text{sec}/\text{m}^2$) for April 1945 from Four RATCHET Runs with Subsets of the Parameters Held Constant

	<u>No Met. Uncertainty and Uniform Release</u>	<u>Met. Uncertainty and Uniform Release</u>	<u>No Met. Uncertainty and Hourly Release</u>	<u>Met. Uncertainty and Hourly Release</u>
Eltopia	4.560E-5	5.599E-5	5.631E-5	6.167E-5
Richland	2.306E-5	3.130E-5	2.191E-5	2.492E-5
Ritzville	1.313E-5	1.430E-5	1.209E-5	1.282E-5
Benton City	1.642E-5	2.220E-5	7.894E-7	2.206E-6
Spokane	1.154E-5	1.093E-5	1.109E-5	1.055E-5
Pendleton	8.474E-7	6.323E-7	8.807E-7	8.013E-8
Baker	9.400E-9	1.991E-8	9.862E-10	2.257E-9
Yakima	7.403E-11	1.003E-10	5.105E-11	8.961E-11

Table 5.17. Coefficients of Variations for 50 Predictions of Integrated Air Concentration of Iodine-131 for April 1945 from Four RATCHET Runs with Subsets of the Parameters Held Constant

	<u>No Met. Uncertainty and Uniform Release</u>	<u>Met. Uncertainty and Uniform Release</u>	<u>No Met. Uncertainty and Hourly Release</u>	<u>Met Uncertainty and Hourly Release</u>
Eltopia	0.056	0.170	0.234	0.406
Richland	0.056	0.479	0.485	1.155
Ritzville	0.057	0.094	0.263	0.246
Benton City	0.058	0.313	0.269	0.423
Spokane	0.058	0.116	0.226	0.231
Pendleton	0.060	0.327	0.770	0.751
Baker	0.060	0.754	0.208	0.858
Yakima	0.068	1.727	1.817	3.133

5.6.4 Summary of Results from the Uncertainty and Sensitivity Analyses

The key conclusions from the uncertainty and sensitivity analyses of the dose for the demonstration are:

- The uncertainty in the predicted total surface deposition of iodine-131 in 1945 is small compared to the uncertainty in the predicted annual dose to thyroid from ingestion of cow's milk containing iodine-131 (see Figures 5.3 and 5.18). The length of the 90 percent subjective confidence intervals have decreased from being more than a factor of 17 to being less than a factor of 2.
- Within the major plume of iodine-131, the uncertainty in the total surface deposition of iodine-131 is sensitive to the uncertainty in the partitioning of iodine.

The key implications of these uncertainty and sensitivity analyses of the dose for the demonstration are:

- The choice of the CDFs which describes the uncertainty in the partitioning of iodine should be examined.

The key implications that these uncertainty and sensitivity analyses of the dose for the case study has on the Uncertainty and Sensitivity Analyses Plan are

- The multiple regression technique used in the sensitivity analysis did not work very well for this model, as demonstrated by the small coefficients of determination for almost all of the regressions at each location.
- The sensitivity analysis for the surface deposition should be conducted using shorter time periods which are relevant to the important plant products, not the entire year.

- The multiple regression technique will not work, and was not attempted for examining the sensitivity of surface deposition to the uncertainties in wind direction, wind speed, and so forth. These parameters change hourly, at a number of different meteorology stations and form a wind field over the entire grid. For these parameters, the sensitivity analysis using the partitioning of the uncertainty by holding subsets of the parameters constant will be used.
- The uncertainty and sensitivity analyses of the annual dose to the thyroid of the reference adult male from ingestion of milk containing iodine-131 need not continue to another model. (However, the remaining models, STRM and RM should have uncertainty and sensitivity analyses performed on them as part of the model reliability process.)

5.7 Conclusions

The key conclusions from this demonstration are divided into three categories: 1) results of the uncertainty and sensitivity analyses on dose to the thyroid from ingestion of cow's milk containing iodine-131; 2) the implications of these results on the models and their parameters in HEDRIC; and 3) the implications of the demonstration on the Uncertainty and Sensitivity Analyses Plan.

- Results of the uncertainty and sensitivity analyses
 - The uncertainties in the dose to the thyroid from ingestion of cow's milk containing iodine-131 are very large. The 95 percent subjective upper confidence level on the dose is 17 or more times larger than the 5 percent subjective lower confidence level.
 - The important parameters, relative to their importance to the uncertainty in the dose, are shown in Figure 5.20. This figure is based on the sensitivity analysis of the reference adult male at each location. The ordering of importance will change slightly with location.
 - The dose factor for ingestion and feed-to-milk transfer factor are the most important parameters. Together they account for over 50 percent of the uncertainty in the predicted dose.
- Implications of these results on the models in HEDRIC and their parameters
 - The parameters shown in Figure 5.20 warrant reexamination, prior to the final runs of HEDRIC, with the greatest emphasis given to the dose factor for ingestion and the feed-to-milk transfer factor.
- Implications of the demonstration on the Uncertainty and Sensitivity Analyses Plan
 - The sensitivity analysis based on measures from multiple regression can be successfully applied to the models from DESCARTES and CIDER.
 - Explicit measures of the effect of the uncertainty of harvest dates on the uncertainty of dose were not developed. However, their importance is demonstrated implicitly by the importance of the uncertainty of the most

Figure 5.20. Ranking of Parameters by the Importance of Their Uncertainty on the Uncertainty in Dose from Ingestion of Milk from Cows on Feeding Regimes 1 and 3

recent (previous week's) surface deposition on the uncertainty of iodine-131 concentration in grain. To develop explicit measures would not be cost effective.

- Except for the iodine-131 partitioning within the major iodine-131 plume, the uncertainty in the surface depositions is not explained by measures from multiple regression. The uncertainty in the surface deposition, and the RATCHET model in general, must rely on a sensitivity analysis which partitions the uncertainty by holding subsets of the parameter constant.
- The use of boxplots to display the uncertainties is preferred to tables of numerical descriptors. The boxplot, in one picture, displays all the relevant information about the uncertainty (that is, ranges and 90 percent subjective confidence interval) along with the central predicted dose (either median or mean).

6.0 Recommendations, Data Quality Objectives, and Conclusions

The recommended approaches for conducting uncertainty and sensitivity analyses for the HEDR Project are provided in Sections 6.1 and 6.2, respectively. The recommended approaches were evaluated in terms of cost-effectiveness, necessary scientific rigor for given models, and effective communication of results.

6.1 Recommended Approach for Uncertainty Analyses of HEDR Models

Uncertainty analyses should be conducted for all dose predictions. These analyses lead to the most appropriate interpretation of the predicted doses because they provide a measure of the precision of the prediction.

The Monte Carlo technique should be used to estimate all dose uncertainties for the following reasons:

- consistency: the method can be used for all models from the simplest to the most complex
- cost effectiveness: off-the-shelf software (for example, Crystal Ball®) can be used for all but the most complex models
- accuracy: compared with the mathematical approach, the Monte Carlo technique gives more information about uncertainties. For example, subjective confidence intervals about the predictions can be developed directly from the output of the Monte Carlo run. However, additional assumptions must be made before such subjective confidence intervals can be made about the predictions derived from the mathematical approach.
- appropriate technique: the mathematical approach only approximates expected values (of doses) and variances. The more complex the model becomes, the greater the potential that results from the mathematical approach will be incorrect. The analytical technique is only comparable to Monte Carlo when the model is very simple and the parameters are independent.

Stratified random sampling is the recommended sampling strategy for those model parameters that are infrequently resampled. For those parameters that are frequently resampled, for example, parameters that are sampled on a daily basis, simple random sampling is recommended. The results of uncertainty analyses should be presented using boxplots for the following reasons:

- one boxplot figure provides all significant information: the range of predicted dose, the subjective confidence intervals, and estimates of the central value (the mean and median dose)
- are easily understood and interpreted

- can be used for all results
- can be directly obtained from the output of the Monte Carlo technique.

When a deterministic model is used, uncertainty analyses should be performed using the preceding recommendations for selected locations; pathways; and age, gender, and lifestyle categories. These results will be extended to specific cases by analogy.

6.2 Recommended Approach for Sensitivity Analyses of HEDR Mathematical Models

Sensitivity analyses should be performed on all HEDR computational models. Sensitivity analyses provide a method for 1) effectively interpreting the dose estimates, and 2) prioritizing individual parameters according to the uncertainty of the predicted dose. The results of the sensitivity analyses allow development of the most cost-effective strategy for evaluating the uncertainties in the value of model parameters.

For the most important parameters, the form of the probabilistic distribution (for example, normal or log-normal) for the parameter is very important. However, for those parameters whose distributions are not important to the uncertainty in dose, only the central estimate needs to be correct.

Sensitivity analyses should be performed using measures from multiple regression (coefficient of partial determination and the standardized regression coefficient). The multiple regressions should be performed on both the original parameter values and the rank transformed values. The results from the model that have the highest coefficient of multiple determination should be used. Multiple regression is cost-effective because the software is available and the approach is not labor intensive. The appropriateness of the multiple regression approach can be easily measured with the coefficient of determination. For the HEDRIC codes, sensitivity analyses using multiple regression have been effectively demonstrated for CIDER and DESCARTES. When the coefficient of determination is small, sensitivity analyses should be performed by holding subsets of the parameters constant and measuring the reduction in the uncertainty for each subset. An additional recommendation is for the uncertainty analyses to be done hierarchically, as demonstrated in Section 5.0. When a deterministic model is used, sensitivity analyses should be performed using these recommendations.

6.3 Extent to Which Sensitivity Analyses Should be Performed

Sensitivity analyses should be performed for all major pathways. The analyses should be performed on the predictions from several different locations within the HEDR study area because the sensitive parameter may change with location. Locations should include areas within the major radionuclide plumes, on the edge of these plumes, and outside of the plumes to determine whether a spatial effect on the uncertainties from the model is present.

Sensitivity analyses should be performed for all major time periods. For example, although the uncertainty in the start of the pasture season is not important to the uncertainty in the 1945 annual dose, it may be important to the uncertainty in the 1946 annual dose. The analyses must also be performed for distinct age, gender, and lifestyle categories.

6.4 Data Quality Objectives for Uncertainty and Sensitivity Analyses

The data quality objectives for the uncertainty and sensitivity analyses for the HEDR Project are defined in Shipler (1992) as follows:

Accuracy—The objective is to accurately quantify the uncertainties of the dose estimate. This will be achieved by accurately estimating the uncertainties of the parameters used by the model and propagating these uncertainties correctly throughout the computer models. The attainment of this objective will be assessed by peer reviews of the computer models and the parameters used within them.

This report describes the process and plan for estimating uncertainties and sensitivity of the HEDR mathematical models. The accuracies of the described formulas and estimations for uncertainty and sensitivity will be assessed during a peer review of this report. Accuracy of any uncertainties and sensitivity of specific HEDR results will be evaluated using methods described in this report when those results are computed and reported.

Completeness—The objective is to produce accurate uncertainties for all dose estimates. Additionally, sensitivity analysis will be performed for all distinct cases. The attainment of this objective will be accomplished by peer reviews.

This report describes the process and plan for estimating uncertainties and sensitivity of the HEDR mathematical models. The completeness of the described formulas and estimations for uncertainty and sensitivity will be assessed during a peer review of this report. Completeness for any specific HEDR results will be evaluated using methods described in this report when those results are computed and reported.

Representativeness—The objectives are that computer models accurately represent the conceptual models and that conceptual models accurately represent all the pathways by which a dose may have been received. The attainment of these objectives will be assessed by peer reviews.

Representativeness is not applicable to information in this report. The representativeness of HEDR models will be evaluated in accordance with the HEDR Verification and Validation Plan and will be reported in the Model Reliability Report.

Comparability—The objective is to ensure the models will be the most reliable that can be produced. The attainment of this objective will be assessed by peer review and validation studies where possible.

Comparability of HEDR models is not an objective of this report. Comparability of HEDR models will be accomplished in accordance with the HEDR Verification and Validation Plan, and will be reported in the Model Reliability Report.

6.5 Conclusions

The recommended approaches for performing uncertainty and sensitivity analyses within the HEDR Project have been effectively demonstrated in the analysis of the dose to the thyroid from the

ingestion of cow's milk containing iodine-131. These results indicate this approach can be used for all the dose predictions from HEDRIC. The reduced complexity of the additional models being developed by the HEDR Project should easily accommodate this approach. This report recommends the TSP approve the use of this plan.

7.0 References

- Agnew, C. E. 1985. "Multiple Probability Assessments by Dependent Experts." Journal of the American Statistical Association 80(390):343-347.
- Beck, D. M. R. F. Darwin, A. R. Erickson, and R. L. Eckert. 1992. Milk Cow Feed Intake and Milk Production and Distribution Estimates for Phase I. PNL-7227 HEDR, Pacific Northwest Laboratory, Richland, Washington.
- Bernreuter, D. L. 1980. Seismic Hazard Analysis: Solicitation of Expert Opinion. NUREG/CR-1582, TERA Corporation, Berkeley, California.
- Breshears, D. D., T. B. Kirchner, M. D. Otis, and F. W. Whicker. 1989. "Uncertainty in Predictions of Fallout Radionuclides in Foods and of Subsequent Ingestion." Health Physics 57(6):943-953.
- Crick, M. J., E. Hofer, J. A. Jones, and S. M. Haywood. 1988. Uncertainty Analysis of the Food-chain and Atmospheric Dispersion Modules of MARC. NRPB-R184, National Radiological Protection Board, Chilton, England.
- Dunning, D. E., and G. Schwarz. 1981. "Variability of Human Thyroid Characteristics and Estimates of Dose from Ingested Iodine-131." Health Physics 40(5):661-675.
- Eslinger, P. W. 1992. Software Development Plan for DESCARTES and CIDER. BN-SA-3667 HEDR, Battelle Pacific Northwest Laboratories, Richland, Washington.
- Eslinger, P. W., T. B. Miley, S. J. Ouderkirk, and W. E. Nichols. 1992. Preliminary Design Studies for the DESCARTES and CIDER Codes. BN-SA-3681 HEDR, Battelle Pacific Northwest Laboratories, Richland, Washington.
- Finkel, A. M. 1990. Confronting Uncertainty in Risk Management: A Guide for Decision-Makers. Center for Risk Management, Resources for the Future, Washington, D.C.
- Good, I. J. 1965. The Estimation of Probabilities. Research Monograph 30. M.I.T. Press, Cambridge, Massachusetts.
- Heeb, C. M. 1992. Iodine-131 Releases from the Hanford Site, 1944 Through 1947, Volume 1 - Text, Volume 2 - Data. PNWD-2033 HEDR Vol. 1 and 2, Battelle, Pacific Northwest Laboratories, Richland, Washington.
- Iman, R. L., and W. J. Conover. 1983. A Modern Approach to Statistics. Wiley and Sons, New York.
- Iman, R. L., and M. J. Shortencarier. 1984. A FORTRAN 77 Program and User's Guide for the Generation of Latin Hypercube and Random Samples for Use with Computer Models. NUREG/CR-362 SAND-83-2365, Sandia National Laboratories, Albuquerque, NM.
- Iman, R. L., and J. C. Helton. 1985. A Comparison of Uncertainty and Sensitivity Analysis Techniques for Computer Models. NUREG/CR-3904, Nuclear Regulatory Commission, Washington, D.C.
- Iman, R. L., and J. C. Helton. 1988. "An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models." Risk Analysis 8(1):71-90.
- International Atomic Energy Agency (IAEA). 1989. Evaluating the Reliability of Predictions Made Using Environmental Transfer Models. IAEA Safety Series No. 100, Vienna.

James, R. C., and G. James. 1968. Mathematics Dictionary. 3rd ed. Van Nostrand and Reinhold Company, New York.

McKay, M. D, 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(2):239-245.

Merkhofer, M. W., and P. McNamee. 1982. The SRI Probability Encoding Process: Experiences and Insights. Technical Report. SRI International, Menlo Park, California.

Miley, T. B., S. J. Ouderkirk, W. E. Nichols, and P. W. Eslinger. 1993. Data Model Description for the DESCARTES and CIDER Codes. BN-SA-3679 HEDR, Battelle Pacific Northwest Laboratories, Richland, Washington.

Mood, A. M., F. A. Graybill, and D. C. Boas. 1974. Introduction to the Theory of Statistics. 3rd ed. McGraw-Hill, Inc., New York.

Morgan, M. G., M. Henrion, and S. C. Morris. 1979. Expert Judgments for Policy Analysis. BNL-51358, Brookhaven National Laboratory, Upton, New York.

Morgan, M. G., and M. Henrion. 1990. Uncertainty. A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press, New York.

Morgan, M. G., M. Henrion, and S. C. Morris. 1980. Expert Judgments for Policy Analysis. BNL-51358, Brookhaven National Laboratory, Upton, New York.

Neter, J., W. Wasserman, and M. H. Kutner. 1985. Applied Linear Statistical Models. 2nd ed. Irwin, Inc., Homewood, Illinois.

Ramsdell, J. V., Jr., and K. W. Burk. 1992. Regional Atmospheric Transport Code for Hanford Emission Tracking (RATCHET). PNL-8003 HEDR, Pacific Northwest Laboratory, Richland, Washington.

Roberds, W. J. 1990. "Methods for Developing Defensible Subjective Probability Assessments." Transportation Research Record, No. 1288:183-190.

SENES Consultants Limited. 1985. Uncertainty Analysis and Probabilistic Modelling. INFO-0172, Atomic Energy Control Board, Ottawa, Canada.

Shipler, D. B. 1992. Integrated Task Plans for the Hanford Environmental Dose Reconstruction Project, FY 1992 Through May 1994. PNWD-2020 HEDR, Battelle, Pacific Northwest Laboratories, Richland, Washington.

Snyder, S. F., W. T. Farris, B. A. Napier, T. A. Ikenberry, and R. O. Gilbert. 1992. Parameters Used in the Environmental Pathways (DESCARTES) and Radiological Dose (CIDER) Modules of the Hanford Environmental Dose Reconstruction Integrated Codes (HEDRIC) for the Air Pathway. PNWD-2023 HEDR, Battelle Pacific Northwest Laboratories, Richland, Washington.

Spetzler, C. S., and C.A.S. von Holstein. 1972. "Probability Encoding in Decision Analysis." Example of a Probability Encoding Technique, K-258802, Stanford Research Institute, ORSA-TIMS-AIEEE Joint National Meeting, Atlantic City, New Jersey.

Stevens, W., J. E. Till, D. C. Thomas, J. L. Lyon, R. A. Kerber, S. Preston-Martin, S. L. Simon, M. L. Rallison, and R. D. Lloyd. 1992. Assessment of Leukemia and Thyroid Disease in Relation to Fallout in Utah: Report of a Cohort Study of Thyroid Disease and Radioactive Fallout from the Nevada Test Site. University of Utah, Salt Lake City, Utah.

Taylor, A. E., and W. R. Mann. 1983. Advanced Calculus. 3rd ed. John Wiley & Sons, Inc., New York.

TERA Corporation. 1980. Seismic Hazard Analysis: Solicitation of Expert Opinion. NUREG/CR-1582, Vol. 3. Nuclear Regulatory Commission, Washington, D.C.

Tukey, J. W. 1977. Exploratory Data Analysis. Addison-Wesley Publishing Company, Inc. Reading, Massachusetts.

von Holstein, C.A.S., and J. E. Matheson. 1979. A Manual for Encoding Probability Distributions. Final Report, SRI International, Menlo Park, California.

Winkler, R. L. 1969. "Scoring Rules and the Evaluation of Probability Assessors." Journal of the American Statistical Association 64(327):1073-1078.

Zamora, R. M. 1975. ONR Probability Encoding Task, Working Paper. SRI Project 4030, Office of Naval Research, Operational Decision Aids Project, Stanford Research Institute, Menlo Park, California.

Appendix A

Statistical Terminology Used in Uncertainty and Sensitivity Analyses

Appendix A

Statistical Terminology Used in Uncertainty and Sensitivity Analyses

This appendix defines the statistical terms and notation that are used in uncertainty and sensitivity analyses. The primary sources for this section are Mood, Graybill and Boas (1974), and Neter, Wasserman, and Kutner (1985). However, the reader is referred to any introductory books on theoretical and applied statistics for additional information on the following terminology.

A.1 Random Variable

A random variable associates an outcome of a random experiment with a real number. For example, tossing a coin is a random experiment. Two possible outcomes of this experiment are possible, a head or a tail. The random variable could be defined as the number of heads. In mathematical notation, this would be

$$X = \begin{cases} 1 & \text{if a head} \\ 0 & \text{if a tail} \end{cases} \quad (\text{A.1})$$

A random variable is usually denoted by a capital letter, such as X . Another random experiment is tossing a six-sided die. In this case, a number of different random variables can be defined. Two random variables from this experiment are:

$$\begin{aligned} X &= \begin{cases} 1 & \text{if a six appears} \\ 0 & \text{otherwise} \end{cases} \\ Y &= \begin{cases} 1 & \text{if a three or four appears} \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (\text{A.2}) \quad (\text{A.3})$$

As shown, the random variable is a function that transforms the results of a random experiment into a numerical value. The random variable is coupled with a probability distribution.

Another random experiment could be growing pasture grass on a 10-acre plot. The random variable, X , could be defined as the biomass per square meter. When this experiment is actually performed, the experimenter goes out in the field, cuts a square meter of pasture, dries it, and then weighs it. The outcome is then denoted as $x = 0.9 \text{ kg (dry)/m}^2$. The value of the random variable, obtained by performing the random experiment is usually denoted by the corresponding small letter.

If five different plots were sampled, the resulting biomass for each acre would form the vector $\underline{x} = [x_1, x_2, x_3, x_4, x_5]$, where x_i is the biomass that is associated with the i th plot.

A random experiment can have more than one random variable, for example biomass of pasture grass and concentration of iodine-131 per kg(dry) of the pasture grass. A multivariate random variable is usually denoted as a vector, for example, ($\underline{X} = [X_1, X_2]$) where X_1 is the biomass of the pasture grass and X_2 is the concentration of iodine-131 per L (dry) of the pasture grass. In the case of the multivariate sample, the results form an array, usually denoted by a bold lower-case letter. If five different plots of pasture grass are sampled, the resulting biomass and concentration of iodine-131 for each plot would form the array

$$\mathbf{x} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} & x_{15} \\ x_{21} & x_{22} & x_{23} & x_{24} & x_{25} \end{bmatrix} \quad (\text{A.4})$$

where x_{ij} is a real number that is associated with the j th plot of the i th random variable, $i=1$ is the biomass, and $i=2$ is the iodine concentration.

A.2 Cumulative Distribution Function

The cumulative distribution function (CDF) gives the probabilities of occurrence of certain events defined by the random variable. A CDF is usually denoted by $F_x(x) = P[X \leq x]$ which means that with probability P the random variable X has a value of less than or equal to x . For example, the holdup time of milk from the backyard cow, $Holdup_{backyard\ milk}$, is a random variable. It is assumed to have a uniform distribution with a minimum of one day and a maximum of three days (Snyder et al. 1992). Hence, the CDF for $Holdup_{backyard\ milk}$ is

$$F_x(x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{x-1}{2} & \text{if } 1 \leq x \leq 3 \\ 1 & \text{if } x > 3 \end{cases} \quad (\text{A.5})$$

Then

$$P[Holdup_{backyard\ milk} \leq 2 \text{ days}] = F_x(2) = \frac{2-1}{2} = \frac{1}{2}. \quad (\text{A.6})$$

It is common to display the CDF graphically, as in Figure A.1. The x-axis covers the range of possible values for the random variable and the y-axis has a range of $P[X \leq x]$ from 0 to 1.

CDFs for the distributions used by the HEDR Project are shown in Appendix B.

A.3 Percentiles

The p th percentile of the distribution is defined as the value x of the random variable X for which the solution is

$$P[X \leq x] = F_x(x) = \frac{p}{100}. \quad (\text{A.7})$$

Figure A.1. Cumulative Distribution Function and Probability Density Function for the Holdup Time for Milk from a Backyard Cow

As seen in Figure A.1, for the holdup time of milk from the backyard cow, *Holdup_{backyard milk}*, the 10th percentile is 1.2. That is, the probability is 0.1 that holdup time is less than or equal to 1.2 days. Other percentiles for this parameter include the 25th percentile (1.5 days), the 50th percentile (2.0 days), the 75th percentile (2.5 days), and the 90th percentile (2.8 days). The 50th percentile is also called the median, the 25th and 75th percentiles are also called the first and third quartiles, respectively.

A.4 Probability Density Function

If a random variable is continuous, the PDF, denoted by $f_X(x)$, is the function where

$$F_X(x) = \int_{-\infty}^x f_X(u) du \quad (A.8)$$

The density function of a continuous random variable is the function that a histogram of a random sample will tend to resemble as the number of samples becomes large and the class widths become small (Iman and Conover 1983).

For the holdup time of milk from the backyard cow, *Holdup_{backyard milk}*, which is assumed to have a uniform distribution with a minimum of one day and a maximum of three days, the PDF is

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{if } 1 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases} \quad (A.9)$$

It is common to display the PDF graphically, as in Figure A.1. The x-axis covers the range of possible values for the random variable and the y-axis has a range of 0 to the maximum $f_X(x)$, which is different for each distribution. The area under the PDF is always equal to one. The area under the PDF between two different values of the random variable is the probability the value of the random variable is between those two values. For example,

$$P[1.5 \text{ days} \leq \textit{Holdup}_{\textit{backyard milk}} \leq 2.5 \text{ days}] = \int_{1.5}^{2.5} \frac{1}{2} du = \frac{1}{2}. \quad (A.10)$$

The PDFs of the distributions used by the HEDR Project are shown in Appendix B.

A.5 Expectation

The expectation of a continuous random variable X is defined as follows:

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx \quad (A.11)$$

where $g(X)$ = a function of the random variable X

$E[g(X)]$ = the expected value of the function $g(X)$

$f_x(x)$ = the probability density function of X

A.6 Mean

The mean is the expected value of a random variable X . It is the center of gravity of the probability density function and is one of the measures of the center of the probability density function. The median, or 50th percentile, is another measure of the center. For symmetrical distributions, the mean and median are the same. The mean is also referred to as the first moment of the random variable.

The mean of a random variable is usually denoted as $E[X]$ or μ_x .

The mean for a continuous random variable is

$$E[X] = \int_{-\infty}^{+\infty} xf_x(x)dx \quad (A.12)$$

For the holdup time of milk from the backyard cow, $Holdup_{backyard\ milk}$, the mean is

$$E[Holdup_{backyard\ milk}] = \int_1^3 x \frac{1}{2} dx = \frac{1}{2} \left(\frac{1}{2} x^2 \Big|_1^3 \right) = \frac{1}{4} (3^2 - 1^2) = \frac{8}{4} = 2.0 \text{ days} \quad (A.13)$$

Fortunately, the mean for most probability distributions can be calculated directly from the parameters that describe the distribution. For example, the uniform distribution is described by two parameters, the minimum, a , and the maximum, b . The mean of the holdup time of milk from the backyard cow, $Holdup_{backyard\ milk}$, can simply be calculated as follows

$$E[Holdup_{backyard\ milk}] = \frac{a+b}{2} = \frac{1+3}{2} = 2.0 \text{ days} \quad (A.14)$$

The formulas for the mean of the PDFs used by the HEDR Project are shown in Appendix B.

A.7 Variance

The variance is a measure of the spread or dispersion of the random variable X . It is also referred to as the second central moment of the random variable. It is a central moment because the mean was subtracted from the random variable. It is defined as

$$E[(X - \mu_x)^2] = E[X^2] - (E[X])^2 \quad (A.15)$$

The variance is usually denoted as σ_x^2 , $\text{Var}[X]$ or $V[X]$.

The variance for a continuous random variable is

$$V[X] = \int_{-\infty}^{+\infty} x^2 f_X(x) dx - (E[X])^2 \quad (\text{A.16})$$

For the holdup time of milk from the backyard cow, $\text{Holdup}_{\text{backyard milk}}$, the variance is

$$\begin{aligned} V[\text{Holdup}_{\text{backyard milk}}] &= \int_1^3 x^2 \frac{1}{2} dx - (2)^2 = \frac{1}{2} \left(\frac{1}{3} x^3 \Big|_1^3 \right) - 4 \\ &= \frac{1}{6} (3^3 - 1^3) - 4 = \frac{26}{6} - 4 = \frac{1}{3} \end{aligned} \quad (\text{A.17})$$

The variance for most probability distributions can also be calculated directly from the parameters that describe the distribution. The variance of the holdup time of milk from the backyard cow, $\text{Holdup}_{\text{backyard milk}}$, can simply be calculated as follows

$$V[\text{Holdup}_{\text{backyard milk}}] = \frac{(b-a)^2}{12} = \frac{(3-1)^2}{12} = \frac{4}{12} = \frac{1}{3}. \quad (\text{A.18})$$

The formulas for the variance of the PDFs used by the HEDR Project are shown in Appendix B.

A.8 Standard Deviation

The standard deviation is another measure of the spread or dispersion of the random variable X . It is defined as the positive square root of the variance. For many applications, it is preferable to the variance because it is in the same measurement units as the random variable.

The standard deviation is usually denoted as σ_x , or SD_x .

If the random variable has a normal distribution then

$$P[\mu_x - 2\sigma_x < X < \mu_x + 2\sigma_x] = 0.9544. \quad (\text{A.19})$$

That is, the probability the random variable is within two standard deviations of the mean is 0.9544.

A.9 Covariance

The covariance measures the linear relationship of two random variables, X_1 and X_2 . It is defined as

$$E[(X_1 - \mu_{x_1})(X_2 - \mu_{x_2})] = E[X_1 X_2] - E[X_1]E[X_2]. \quad (\text{A.20})$$

The covariance is usually denoted as $\sigma_{x_1 x_2}$ or $COV[X_1, X_2]$.

The covariance will be positive when $(X_1 - \mu_{x_1})$ and $(X_2 - \mu_{x_2})$ tend to have the same sign with high probability and it will be negative when $(X_1 - \mu_{x_1})$ and $(X_2 - \mu_{x_2})$ tend to have the opposite sign with high probability. Unfortunately the magnitude of the covariance does not have much meaning because it depends on the variance of X_1 and X_2 .

A.10 Correlation Coefficient

The correlation coefficient removes the individual variability of X_1 and X_2 from the covariance and thus is a better measure of the linear relationship between two random variables. It is defined as

$$\frac{COV[X_1, X_2]}{\sqrt{V[X_1]V[X_2]}}. \quad (\text{A.21})$$

The correlation coefficient is usually denoted as $\rho_{x_1 x_2}$.

The correlation coefficient is unitless and has a range from -1 to +1. When no linear relationship exists between two random variables, they are said to be uncorrelated and the correlation coefficient is zero. A positive correlation coefficient indicates when one random variable increases in value, the other random variable increases in value. A negative correlation coefficient indicates when one random variable increases in value, the other random variable decreases in value. When the correlation coefficient is -1 or +1, if the random variables were plotted together, they would form a straight line with a negative or positive slope, respectively.

A.11 Independence

If two random variables are uncorrelated, their covariance is zero. However, covariance only characterizes the linear relationship between random variables. Independence is a much stronger notation where two or more random variables are independent if (and only if)

$$f_{x_1, x_2, \dots, x_k}(x_1, x_2, \dots, x_k) = \prod_{i=1}^k f_{x_i}(x_i) \quad (\text{A.22})$$

or

$$F_{x_1, x_2, \dots, x_k}(x_1, x_2, \dots, x_k) = \prod_{i=1}^k F_{x_i}(x_i) \quad (\text{A.23})$$

Thus, if two random variables are uncorrelated, they are not necessarily independent. However, if two random variables are independent, they are also uncorrelated.

A.12 Sample Statistics

When a random experiment is performed, the resulting outcomes are a sample, of size n , from the distribution that describes the random variable. A number of sample statistics can be calculated using these outcomes.

Sample mean

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (\text{A.24})$$

Sample variance

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \quad (\text{A.25})$$

Sample standard deviation

$$s_x = \sqrt{s_x^2} \quad (\text{A.26})$$

Sample correlation coefficient

$$r_{x_1 x_2} = \frac{\sum_{i=1}^n x_{1i} x_{2i} - n\bar{x}_1 \bar{x}_2}{\sqrt{\left(\sum_{i=1}^n x_{1i}^2 - n\bar{x}_1^2 \right) \left(\sum_{i=1}^n x_{2i}^2 - n\bar{x}_2^2 \right)}} \quad (\text{A.27})$$

The sample correlation coefficient is often called the Pearson correlation coefficient.

Often these sample statistics are used as estimates of the corresponding descriptors of their probability distribution that are discussed in this appendix, that is

$$\hat{\mu}_x = \bar{x} \quad (\text{A.28})$$

$$\hat{\sigma}_x^2 = s_x^2 \quad (\text{A.29})$$

$$\hat{\sigma}_x = s_x \quad (\text{A.30})$$

$$\hat{\rho}_{x_1 x_2} = r_{x_1 x_2} \quad (\text{A.31})$$

The little 'hat' on the notation indicates that it is an estimate.

A.13 Confidence Interval

For a random sample $(X_1, X_2, \dots, X_n)$ from a probability distribution parameterized by θ , the random interval between (T_1, T_2) is a statistical 100γ percent confidence interval for $\tau(\theta)$ if

$$P[T_1 < \tau(\theta) < T_2] = \gamma \quad (\text{A.32})$$

where

$T_1 = t_1(X_1, X_2, \dots, X_n)$ = a function of the random sample, called the lower confidence limit

$T_2 = t_2(X_1, X_2, \dots, X_n)$ = another function of the random sample called the upper confidence limit

$\tau(\theta)$ = a function of the parameter θ .

Each time a random sample is taken, the value of (T_1, T_2) will be different, but $\tau(\theta)$ is the same. A 95 percent confidence interval indicates for a specific random sample that a 95 percent chance exists for the calculated value of the interval (T_1, T_2) to include $\tau(\theta)$.

For example, if the random sample is from a normal distribution, then a 95 percent confidence interval for μ is

$$\left(\bar{x} - 1.96 \frac{s_x}{\sqrt{n}}, \bar{x} + 1.96 \frac{s_x}{\sqrt{n}} \right) \quad (\text{A.33})$$

because

$$P \left[\bar{x} - 1.96 \frac{s_x}{\sqrt{n}} < \mu < \bar{x} + 1.96 \frac{s_x}{\sqrt{n}} \right] = 0.95. \quad (\text{A.34})$$

In the uncertainty analysis, the confidence intervals are called subjective to distinguish them from the statistical confidence intervals. The statistical confidence interval is based on actual samples from the field (for instance, the dry weight of pasture and concentration of iodine in the pasture that is collected in the field, see Section A.1). The uncertainty analysis is based primarily on subjective probability distributions derived from expert judgment.

A.14 Boxplot

The boxplot shown in Figure A.2, is a generalization of the box-and-whisker plot developed by Tukey (1977). The boxplots used in this document have a box that contains the middle 50 percent of the predicted values, between the 25th and 75th percentiles. Within the box, the median (50th percentile) and mean are shown. The ends of the whiskers generally are the 5th and 95th percentile, which are the lower and upper subjective confidence limits of the 90 percent subjective confidence interval. The minimum and maximum predicted values are also generally shown. In practice, each investigator defines the percentiles used in the boxplot.

A.15 Rank Correlation Coefficient

The sample correlation coefficient describes the linear relationship between the samples of two random variables. The rank correlation coefficient describes the monotonic relationship between the samples of two random variables. A linear relationship is one that can be described by a straight line. A monotonic relationship is one that can be described by a line that is continually increasing (or continually decreasing), but is not necessarily straight.

Each of the n outcomes in each sample are sorted from the smallest to the largest. Then, the value of each outcome is replaced by its rank in the sample. That is, the smallest outcome in each

Figure A.2. Example of a Boxplot

sample is replaced by 1, the next smallest outcome in each sample is replaced by 2, and so forth. The rank correlation coefficient is then calculated using the equation for the sample correlation coefficient with the value of each outcome replaced by its rank.

Figure A.3a (a simple linear regression using actual values) shows the relationship between the concentration of iodine-131 in milk and the feed-to-milk transfer factor for an individual cow, which has a sample correlation coefficient of 0.618. Figure A.3b (a simple linear regression using the ranks of the values shown in Figure A.3a) shows the relationship between the ranks of the two variables in Figure A.3a, which has a rank correlation coefficient of 0.825.

The sample rank correlation coefficient is often called the Spearman rank correlation coefficient.

A.16 Sample Coefficient of Variation

The ratio of the sample standard deviation to the sample mean, $s_x / \bar{x}$, is the sample coefficient of variation. The sample coefficient of variation is often denoted by CV. This sample statistic is useful when comparing a series of samples where the sample mean and standard deviation tend to change together.

A.17 Simple Linear Regression

Simple linear regression is used to fit a straight line through n pairs of observations, $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$. The straight line in Figure A.3a is the result of the simple linear regression of the feed-to-milk transfer factor on milk concentration. The model for this straight line relationship is

$$Y = \beta_0 + \beta_1 x + \varepsilon \quad (\text{A.35})$$

where Y = dependent variable

x = independent variable

β_0 = intercept, the value of Y when x is zero

β_1 = slope of the straight line

ε = random error component.

The slope and intercept are often referred to as the regression coefficients.

The estimates of the parameters of the linear regression are

Figure A.3. Relationship Between the Concentration of Iodine-131 in Milk and the Feed-to-Milk Transfer Factor for the Individual Cow

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (\text{A.36})$$

and

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}. \quad (\text{A.37})$$

A.18 Multiple Linear Regression

Multiple linear regression is an extension of the simple linear regression model to allow for p independent variables,

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon. \quad (\text{A.38})$$

Often this function is called a response surface. The estimates of the regression coefficients are calculated using a computer since a $p+1$ dimension matrix must be inverted. The regression coefficients are also called partial regression coefficients.

A.19 Coefficient of Determination

The coefficient of determination, R^2 , is a measure of how well the linear regression model fits the observations. It is the proportion of the variability in the dependent variable that is explained by the linear regression.

For simple linear regression, the coefficient of determination (sometimes called the coefficient of simple determination) is the square of the sample correlation coefficient. For the example shown in Figure A.3a, $R^2 = 0.382$, or only 38.2 percent of the variability in the concentration of iodine-131 in milk is explained by a linear regression with the feed-to-milk transfer factor in the cow. However, when the observations are replaced by their ranks, in Figure A.3b, $R^2 = 0.680$, or 68 percent of the variability in the ranks of the concentration of iodine-131 in milk, is explained by a linear regression with the ranks of the feed-to-milk transfer factor in the cow.

For multiple linear regression, the coefficient of determination (sometimes called the coefficient of multiple determination) is

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (\text{A.39})$$

where

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_p x_{ip} . \quad (\text{A.40})$$

A.20 Coefficient of Partial Determination

The coefficient of partial determination is a measure of the marginal contribution of one of the independent variables in explaining the variability in the dependent variable, given that all the other independent variables are already included in the model. It is the proportion of the remaining variability in the independent variable, after a regression on $p-1$ independent variables, that is explained by the p th independent variable. The coefficient of partial correlation is the square root of the coefficient of partial determination.

A.21 Standardized Regression Coefficients

The standardized regression coefficients are defined as

$$B_k = \hat{\beta}_k \times \left(\frac{s_k}{s_y} \right) \quad (\text{A.41})$$

where s_k = standard deviation of the k th independent variable, x_k

s_y = standard deviation of the dependent variable.

The standardized regression coefficient gives the number of standard deviations the independent variable changes when the k th dependent variable is changed by one standard deviation. The standardized regress coefficients are also called beta coefficients and standardized partial regression coefficients.

Appendix B

Probability Distributions Used in Uncertainty and Sensitivity Analyses

Appendix B

Probability Distributions Used in Uncertainty and Sensitivity Analyses

This appendix describes the six probability density functions (PDFs) that are proposed in this document for modeling the uncertainty of parameters in HEDR Project models. These six distributions are:

- uniform
- loguniform
- triangular
- normal
- lognormal
- piecewise uniform

For each distribution the following information is provided: a graphical plot of the PDF and of the cumulative distribution function (CDF), the equation for the PDF and the CDF, and the expected value (true mean) and variance of the distribution. In addition, a numerical example illustrates the computation of these quantities. Further information about these distributions is provided in Iman and Shortencarier (1984).

DISTRIBUTION: Uniform(a,b)

$$-\infty < a \leq x \leq b < +\infty$$

$$f_X(x) = \frac{1}{b-a}$$

$$F_X(x) = \frac{x-a}{b-a}$$

$$E[X] = \frac{a+b}{2}$$

$$V[X] = \frac{(b-a)^2}{12}$$

$$\text{median} = E[X]$$

Example $a = 0$ (minimum)
 $b = 12$ (maximum)

$$f_X(x) = \frac{1}{12}$$

$$F_X(x) = \frac{x}{12}$$

$$E[X] = \frac{12}{2} = 6$$

$$V[X] = \frac{144}{12} = 12$$

$$\text{median} = 6$$

DISTRIBUTION: Loguniform(a,b)

$$f_X(x) = \frac{1}{x(\ln b - \ln a)} \quad 0 < a \leq x \leq b < +\infty$$

$$F_X(x) = \frac{\ln x - \ln a}{\ln b - \ln a}$$

$$E[X] = \frac{b-a}{\ln b - \ln a}$$

$$V[X] = \frac{(b-a)(\ln b - \ln a)(b+a) - 2(b-a)}{2(\ln b - \ln a)^2}$$

$$\text{median} = \sqrt{ab}$$

Example

$$a = 1 \quad (\text{minimum})$$

$$b = 36 \quad (\text{maximum})$$

$$f_X(x) = \frac{0.279}{x}$$

$$F_X(x) = 0.279 \ln x$$

$$E[X] = \frac{35}{3.5835} = 9.767$$

$$V[X] = \frac{2190.657}{25.683} = 85.295$$

$$\text{median} = \sqrt{36} = 6$$

DISTRIBUTION: Triangular(a,b,c)

B.4

$$f_X(x) = \begin{cases} \frac{2(x-a)}{(c-a)(b-a)} & \text{if } a \leq x \leq b \\ \frac{2(c-x)}{(c-a)(c-b)} & \text{if } b \leq x \leq c \end{cases} \quad -\infty < a \leq b \leq c < +\infty$$

$$F_X(x) = \begin{cases} \frac{(x-a)^2}{(c-a)(b-a)} & \text{if } a \leq x \leq b \\ \frac{b-a}{c-a} - \frac{(x+b-2c)(x-b)}{(c-a)(c-b)} & \text{if } b \leq x \leq c \end{cases}$$

$$E[X] = \frac{a+b+c}{3} \quad V[X] = \frac{a(a-b)+b(b-c)+c(c-a)}{18}$$

$$\text{median} = \begin{cases} a + \sqrt{(c-a)(b-a)}/2 & \text{if } b \geq (a+c)/2 \\ c - \sqrt{(c-b)(c-a)}/2 & \text{if } b \leq (a+c)/2 \end{cases}$$

Example

$$\begin{aligned} a &= 0 & (\text{minimum}) \\ b &= 4 & (\text{mode}) \\ c &= 12 & (\text{maximum}) \end{aligned}$$

$$f_X(x) = \begin{cases} \frac{x}{24} & \text{if } 0 \leq x \leq 4 \\ \frac{(12-x)}{48} & \text{if } 4 \leq x \leq 12 \end{cases}$$

$$F_X(x) = \begin{cases} \frac{x^2}{48} & \text{if } 0 \leq x \leq 4 \\ \frac{1}{3} - \frac{(x-20)(x-4)}{96} & \text{if } 4 \leq x \leq 12 \end{cases}$$

$$\begin{aligned} E[X] &= \frac{16}{3} = 5.333 & V[X] &= \frac{112}{18} = 6.222 \\ \text{median} &= 12 - \sqrt{48} = 5.072 \end{aligned}$$

DISTRIBUTION: Normal(μ, σ^2)

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad -\infty < x < +\infty$$

$$F_X(x) = \int_{-\infty}^x f_X(u) du \quad [\text{no closed form solution}]$$

$$E[X] = \mu$$

$$V[X] = \sigma^2$$

$$\text{median} = E[X]$$

Example

$$\mu = 6$$

$$\sigma^2 = 4$$

$$f_X(x) = \frac{1}{5.013} \exp\left[-\frac{(x-6)^2}{8}\right]$$

$$E[X] = 6$$

$$V[X] = 4$$

$$\text{median} = 6$$

DISTRIBUTION: Lognormal(μ, σ^2)

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[\frac{-(\ln x - \mu)^2}{2\sigma^2}\right] \quad 0 < x < +\infty$$

$$F_X(x) = \int_0^x f_X(u) du \quad [\text{no closed form solution}]$$

$$E[X] = \exp\left[\mu + \frac{\sigma^2}{2}\right]$$

$$V[X] = (\exp[\sigma^2] - 1) \exp[2\mu + \sigma^2]$$

Geometric Mean (GM) = $\exp[\mu]$

Geometric Standard Deviation (GSD) = $\exp[\sigma]$
median = GM

μ and σ are in units of $\ln$, the natural logarithm

Example

$$\mu = \ln 4 = 1.3863$$

$$\sigma = \ln 1.5 = 0.4055$$

$$f_X(x) = \frac{0.9839}{x} \exp\left[\frac{-(\ln x - 1.3863)^2}{0.3288}\right]$$

$$E[X] = \exp[1.4685] = 4.3427 \quad GM = 4$$

$$V[X] = 0.1787 \exp[2.9370] = 3.3699 \quad GSD = 1.5$$

median = 4

Cumulative Distribution Function, $P[X \leq x]$

Probability Density Function, $f_X(x)$

B.7

DISTRIBUTION: Piecewise Uniform($n, \underline{a}, \underline{p}$)

$$-\infty < a_0 < a_1 < \dots < a_n < +\infty \quad p_i \geq 0 \quad i = 1, \dots, n \quad \sum_{i=1}^n p_i = 1$$

$$f_X(x) = \frac{p_i}{a_i - a_{i-1}} \quad a_{i-1} \leq x \leq a_i$$

$$F_X(x) = \frac{p_i(x - a_{i-1})}{a_i - a_{i-1}} + \sum_{j=1}^{i-1} p_j \quad a_{i-1} \leq x \leq a_i$$

$$E[X] = \frac{1}{2} \sum_{i=1}^n p_i (a_i + a_{i-1})$$

$$V[X] = \frac{1}{3} \sum_{i=1}^n p_i (a_i^2 + a_i a_{i-1} + a_{i-1}^2) - (E[X])^2$$

Example

$$n = 3 \quad a_0 = 0 \quad a_1 = 2 \quad a_2 = 6 \quad a_3 = 12$$

$$p_1 = 0.1 \quad p_2 = 0.6 \quad p_3 = 0.3$$

$$f_X(x) = \begin{cases} 0.05 & \text{if } 0 \leq x \leq 2 \\ 0.15 & \text{if } 2 \leq x \leq 6 \\ 0.05 & \text{if } 6 \leq x \leq 12 \end{cases}$$

$$F_X(x) = \begin{cases} 0.05x & \text{if } 0 \leq x \leq 2 \\ 0.1 + 0.15(x - 2) & \text{if } 2 \leq x \leq 6 \\ 0.7 + 0.05(x - 6) & \text{if } 6 \leq x \leq 12 \end{cases}$$

$$E[X] = \frac{1}{2} [(0.1 \times 2) + (0.6 \times 8) + (0.3 \times 18)] = 5.2$$

$$V[X] = \frac{1}{3} [(0.1 \times 4) + (0.6 \times 52) + (0.3 \times 252)] - (5.2)^2 = 8.6933$$

$$\text{median} = 4.6667$$

Distribution

No. of
Copies

No. of
Copies

OFFSITE

Technical Steering Panel

D. S. Barth
University of Nevada
4505 Maryland Parkway
Las Vegas, NV 89154

W. A. Bishop
2503 Wedgewood Court S.E.
Olympia, WA 98501

M. L. Blazek
Oregon Department of Energy
625 Marion Street N.E.
Salem, OR 97310

G. G. Caldwell
Tulsa City-County Health Dept.
4616 East 15th Street
Tulsa, OK 74112

S. N. Davis
Dept. of Hydrology and Water
Resources
Building 11
University of Arizona
Tucson, AZ 85721

N. J. Germond
224 Iron Mountain Blvd.
Lake Oswego, OR 97034

P. C. Klingeman
Civil Engineering Dept.
Apperson Hall 202
Oregon State University
Corvallis, OR 97331-2302

K. J. Kopecky
Fred Hutchinson Cancer
Research Center
1124 Columbia Street
Seattle, WA 98104

P. D. McGavran
Dept. of Health and Welfare
450 W. State Street, 4th Floor
Boise, ID 83720-5450

R. L. Morrill
Dept. of Geography, DP-10
University of Washington
Seattle, WA 98195

A. H. Murphy
3115 NW McKinley Drive
Corvallis, OR 97330

D. W. Price
Agricultural Economics
Hulbert Hall Room 211
Washington State University
Pullman, WA 99164-6210

M. A. Robkin
Radiological Sciences, SB-75
University of Washington
Seattle, WA 98195

G. S. Roessler
Route 1, Box 139H
Elysian, MN 56028

B. Shleien
2421 Homestead Drive
Silver Springs, MD 20902

A. P. Slickpoo, Sr.
809 Nez Perce Lane
P.O. Box 331
Kamiah, ID 83536

J. E. Till
Route 2 Box 122
Neeses, SC 29107

D. E. Walker, Jr.
P.O. Box 4147
Boulder, CO 80306

No. of
CopiesNo. of
CopiesOther

- 2 DOE Office of Scientific and
Technical Information
Technical Information Center
P.O. Box 62
Oak Ridge, TN 37830
- B. G. Brooks, EH-421
Room J-112
Department of Energy
Germantown, MD 20545
- 18 K. CharLee
Office of Nuclear Waste Mgmt.
Department of Ecology
719 Sleater Kinney Road, S.E.
Suite 200
Olympia, WA 98504
- W. J. Conover
College of Business Administration
Texas Tech University
15th and Flint
Lubbock, TX 79409
- M. R. Donnelly
Public Health Service
Centers for Disease Control
2201 Sixth Avenue, RX-22
Seattle, WA 98121
- H. A. Haerer
Golden Associates, Inc.
4104 148th N.E.
Redmond, WA 98052
- M. Henrion
Lumina Decision Systems, Inc.
350 Cambridge Avenue
Suite 390
Palo Alto, CA 94306

E. Hofer
Gesellschaft für Anlagen- und
Reaktorsicherheit GmbH
Forschungsgelände
8046 Garching
Germany

F. O. Hoffman
SENES, Oak Ridge Inc.
Center for Risk Analysis
677 Emory Valley Road
Suite C
Oak Ridge, TN 37830

L. E. Sewell
Centers for Disease Control
4770 Buford Highway N.E., F-35
Atlanta, GA 30341-3724

S. L. Simon
Nationwide Radiological Study
P.O. Box 1808
Republic of Marshall Islands
Majuro, Marshall Islands 96960

J. P. Thomas
HEAL
1720 N. Ash
Spokane, WA 99205

ONSITE

- 6 DOE Richland Field Office
- R. F. Brich, TSD A5-55
Public Reading Room (5) A1-65
- 53 Battelle, Pacific Northwest
Laboratories
- D. M. Anderson K6-57
G. L. Black (5) K1-25

No. of
Copies

S. D. Cannon	K1-25
D. H. Denham	K3-56
D. E. Deonigi	K6-54
P. W. Eslinger	K6-96
W. T. Farris	K3-54
M. D. Freshley	K6-77
N. D. Foote	K7-70
R. O. Gilbert	K7-34
W. A. Glass (14)	K4-13
R. H. Gray	K1-33
S. P. Gydesen	P8-55
G. L. Harvey	K1-77
C. M. Heeb	K6-42
T. A. Ikenberry	K3-54
E. B. Liebow	HARC

No. of
Copies

C. A. Lopresti	K7-34
B. A. Napier	K3-54
J. V. Ramsdell	K6-03
J. A. Shaw	K1-25
D. B. Shipler	K1-25
J. C. Simpson (3)	K7-34
S. F. Snyder	K3-54
S. Stage	K6-03
C. D. Taylor	B1-40
W. L. Templeton	K1-30
M. E. Thiede	K6-13
W. H. Walters	K6-09
Publishing Coordination	K1-06
Records Center (2)	K3-70
Technical Library (2)	P8-55

END

DATE
FILMED
9 / 7 / 193