ABSTRACT

MOKHTARI, AMIRHOSSEIN. Evaluation of Sensitivity Analysis Methods for Application to Microbial Food Safety Process Risk Models (Under the direction of Dr. H. Christopher Frey).

With the emergence of quantitative risk assessment models in the field of food safety, there has also been a growing recognition of the need for sensitivity analysis of such models. Key questions that must be addressed in performing sensitivity analysis with food safety risk models include the following: What are the key criteria for sensitivity analysis methods applied to food safety risk assessment?; What sensitivity analysis methods are most promising for application to food safety and risk assessment?; and What are the key needs for implementation and demonstration of such methods?

Microbial food safety process risk (MFSPR) models are examples in which the process of sensitivity analysis is challenged by typical characteristics of those models such as non-linearity, thresholds, interactions, use of both continuous and categorical inputs, and two-dimensional simulation of variability and uncertainty.

In June 2001, NCSU hosted a workshop on sensitivity analysis. The workshop supported the need for case studies with existing MFSPR models to demonstrate to practitioners how sensitivity analysis methods can be used and to evaluate various specific methods. Recommendations also were made regarding need for a guidance document to assist practitioners in selecting, applying, interpreting, and reporting the results of sensitivity analysis.

The three main purposes of this dissertation are to: (1) evaluate sensitivity analysis methods; (2) present examples of how sensitivity analysis can be applied to MFSPR models and how the results can be presented and interpreted; and (3) provide guidance to risk analysis

practitioners regarding the application of sensitivity analysis, particularly with regard to MFSPR models as well as other risk models with similar characteristics. This study included a review of existing methods, and a detailed series of quantitative case studies of multiple sensitivity analysis methods applied to an *E. coli O157:H7* food safety process risk model in ground beef. Methods evaluated include Pearson and Spearman correlation analyses, sample and rank regression analyses, sample and rank stepwise regression analyses, analysis of variance, and classification and regression trees. The report presents results and insights from the application of these methods to the model.

The methods were compared and evaluated with regard to multiple criteria, including applicability, robustness of results, ability to rank and discriminate among inputs, providing insights that can be used for risk management decisions, and others. The guidance provided here represents a synthesis of rigorous yet practical information across many disciplines, including engineering and science fields. The guidance is intended to assist exposure and risk modeling practitioners to make effective use of sensitivity analysis methods and the insights they provide. The guidance provides insight regarding: (1) why and when to perform sensitivity analysis; (2) preparation of existing or new models to facilitate sensitivity analysis; (3) defining objectives and case study scenarios; (4) selection of methods; (5) general principles for application of methods; and (6) presentation and interpretation of results.

[This work was supported by a cooperative agreement with U.S. Department of Agriculture's Office of Risk Assessment and Cost-Benefit Analysis]

EVALUATION OF SENSITIVITY ANALYSIS METHODS FOR APPLICATION TO MICROBIAL FOOD SAFETY PROCESS RISK MODELS

By

AMIRHOSSEIN MOKHTARI

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

DEPARTMENT OF CIVIL, CONSTRUCTION, AND ENVIRONMENTAL ENGINEERING

Raleigh
2004

APPROVED BY

_____

_____ _____

_____

Chair of Advisory Committee

To my Lovely Mother, Father, Sister, and Brother

BIOGRAPHY

Amirhossein Mokhtari was born on 13th April, 1975 at Tehran, Iran. He got his Bachelor's Degree in Civil Engineering from University of Tehran, Tehran, Iran in 1998. Upon graduation, he continued for his Master's degree in Hydraulic Structures in the same institute where he graduated in 2000. He came to North Carolina State University for his PhD degree in Environmental Engineering in 2001. He also obtained a minor in Geographic Information Systems (GIS). He was a research assistant in Dr. H. Christopher Frey's research group, and worked on use of sensitivity analysis in food safety risk assessment. He will start working as a postdoctoral research associate in NC State University after completion of his graduation as a full-time employee. His work will be concerned with sensitivity analysis of biological and environmental risk assessment models.

ACKNOWLEDGMENTS

This research was supported by United States Department of Agriculture under a cooperative agreement (No. 58-0111-0-005) with Department of Civil, Construction, and Environmental Engineering at North Carolina State University.

I would like to express my appreciation to Dr. H. Christopher Frey for his continuous guidance and valuable technical suggestions throughout the course of this work. Without Dr Frey's suggestion and guidance, this work would not have been in its current form. I would like to thank my other committee members, Dr. E. Downey Brill, Dr. Peter Cowen, and Dr. Casson Stallings. Whenever I needed any help from these individuals regarding my academic interests, they were always there to help me out.

I would also like to thank my research group, Tanwir, Yunghua, Maggie, Alper, Amr, Kaishan, and Allen for helping me with any questions that I had and giving me company in the times of hard work.

Last, but certainly not least, I would like to thank my parents, Forough and Mohammad Sadegh Mokhtari for their sincere love, enormous support and enthusiasm for my education, and to my sister and brother Farahnaz and Abbas, who emotionally supported me with the pains and the happiness during the last three years; without their endless encouragement and support, this could not have been possible.

Amirhossein Mokhtari

TABLE OF CONTENTS

List of Figures……...ix

List of Tables……xi

Part I. Introduction

1.0 Introduction..... 2

1.1 Objectives 3

1.2 Overview of Research..... 4

1.3 Organization..... 8

1.4 References..... 11

Part II. Guidance for Preparation of Existing or New Models to Facilitate Sensitivity Analysis

2.0 Introduction..... 15

2.1 Preparation of Existing Models for Sensitivity Analysis..... 16

2.1.1 Identification of Model Structure..... 16

2.1.2 Identification of Inputs..... 22

2.1.3 Selection of Model Output for Sensitivity Analysis..... 23

2.1.4 Probabilistic Simulation and Its Implication for Sensitivity Analysis 24

2.1.5 Modification of Models 26

2.2 Preparation of New Models for Sensitivity Analysis 27

2.2.1 Modeling Environment 27

2.2.2 Characterizing Variability and Uncertainty in a Probabilistic Simulation..... 30

2.2.3 Modeling Strategies 31

2.2.4 Model Documentation 32

2.3 References..... 33

Part III. Recommended Practice Regarding Selection of Sensitivity Analysis Methods Applied to Microbial Food Safety Process Risk Models

1.0 Introduction..... 37

2.0 When to Perform Sensitivity Analysis? 38

2.1 Prioritization of Potential Critical Control Points..... 38

2.2 Identification of Key Sources of Variability and Uncertainty 39
2.3 Model Refinement, Verification, and Validation 39
3.0 Selection of Sensitivity Analysis Methods 41
3.1 Key Questions for Selection of Sensitivity Analysis Methods 42
3.1.1 What are the Objectives of Sensitivity Analysis? 42
3.1.2 Based Upon the Objectives What Information is Needed from Sensitivity Analysis? 44
3.1.3 What are the Characteristics of the Model that Constrain or Indicate Preference Regarding Method Selection? 45
3.1.4 How Detailed is the Analysis? 47
3.1.5 Can Commonly Used Methods Adequately Address the Characteristics of Interest in the Analysis? 49
3.1.6 Is the Implementation of the Selected Sensitivity Analysis Method Post Hoc? 49
3.2 Decision Frameworks to Assist in Selecting Sensitivity Analysis Methods 50
4.0 Acknowledgment 53
5.0 References 54
Part IV. Evaluation of Classification and Regression Trees for Sensitivity Analysis: Case Study for the *E. coli* Food Safety Process Risk Model
1.0 Introduction 63
2.0 Overview of Classification and Regression Trees 67
3.0 Overview of the *E. coli O157:H7* Food Safety Process Risk Model 70
4.0 Probabilistic Analysis Scenarios 73
5.0 Application of CART 74
5.1 Application of CART to the Cooking Effect Part 74
5.1.1 Addressing Interactions Using CART 75
5.1.2 Addressing Thresholds Using CART 77
5.2 Application of CART to the Slaughter Module to Address High Exposure Scenarios 78
6.0 Discussion and Conclusions 81
7.0 Acknowledgment 83

8.0 References 84

Part V. Two-Dimensional Sensitivity Analysis: Application of Analysis of Variance to the *E. coli* Food Safety Process Risk Model

1.0 Introduction 103

1.1 Food Safety Risk Assessment Modeling 104

1.2 Need for Sensitivity Analysis 105

1.3 Selection of the Sensitivity Analysis Method and the MFSPR Model for the Case Studies 105

2.0 Overview of ANOVA 107

3.0 Overview of the *E. coli* Food Safety Process Risk Model 108

4.0 Probabilistic Analysis Scenarios 110

4.1 Variability Analysis for Different Uncertainty Realizations 110

4.2 Co-mingled One-Dimensional Variability and Uncertainty Analysis 111

5.0 Application of ANOVA 112

5.1 Variability Analysis for Different Uncertainty Realizations 113

5.2 Co-Mingled Analysis of Variability and Uncertainty 116

5.3 Verification of ANOVA Results 120

6.0 Discussion and Conclusions 122

7.0 Acknowledgment 126

8.0 References 127

Part VI. Evaluation of Sampling-Based Methods for Sensitivity Analysis: Case Study for the *E. coli* Food Safety Process Risk Model

1.0 Introduction 149

2.0 Methodology 150

2.1 Choice of a MFSPR Model 150

2.2 Definition of Case Scenarios for Example Case Studies 151

2.3 Probabilistic Analysis 152

2.4 Selected Sensitivity Analysis Methods 153

2.4.1 Correlation Analysis 154

2.4.2 Regression Analysis 154
2.4.3 Stepwise Regression Analysis 156
3.0 Comparison of Sampling-Based Sensitivity Analysis Methods 156
3.1 Correlation Analysis 157
3.2 Regression Analysis 159
3.3 Stepwise Regression Analysis 161
3.4 Comparison of Results from Selected Sampling-Based Sensitivity Analysis Methods 166
4.0 Discussion and Conclusions 166
5.0 Acknowledgment 170
6.0 References 171
Part VII. Conclusions and Recommendations
7.0 Summary and Conclusions 191
7.1 When Should Sensitivity Analysis be Performed? 191
7.2 What are Key Considerations in Preparation of a Model to Facilitate Sensitivity Analysis? 192
7.3 How Can One Select Among Typical Sensitivity Analysis Methods? 193
7.4 What are the Considerations in Presenting Results of Sensitivity Analysis Methods? 194
7.5 Can Conventional Sensitivity Analysis Methods Provide Robust Insights in Spite of their Apparent Limitations? 197
7.6 How Can One Discriminate the Importance Between Closely Ranked Inputs? 198
7.7 What are the Implications of Sensitivity Analysis in a Two-Dimensional Probabilistic Framework Compared to the One-Dimensional Analysis of Variability and Uncertainty? 199
8.0 Recommendations for Future Work 201
Appendix *A*. Glossary of Technical Terms……..204

LIST OF FIGURES

Part II. Guidance for Preparation of Existing or New Models to Facilitate Sensitivity Analysis

Figure 1. Conceptual Example of a Modular Modeling Framework Comprised of Multiple Modules in Series. 17
Figure 2. Conceptual Framework of Modules in Parallel. 18
Figure 3. Conceptual Example of the Binning Approach, Leading to Loss of One-to-One Correspondence 20
Figure 4. Possible Classifications of Inputs in Food Safety Process Risk Models. 22

Part III. Recommended Practice Regarding Selection of Sensitivity Analysis Methods Applied to Microbial Food Safety Process Risk Models

Figure 1. Decision Framework for Selecting an Appropriate Sensitivity Analysis Method. 58
Figure 2. Decision Framework for Selecting Appropriate Sensitivity Analysis Method for Identifying Key Sources of Variability and Uncertainty and Model refinement as Key Objectives of the Analysis 59

Part IV. Evaluation of Classification and Regression Trees for Sensitivity Analysis: Case Study for the *E. coli* Food Safety Process Risk Model

Figure 1. Schematic Diagram of a Classification and Regression Tree Illustrating Root Node, Internal Nodes, and Terminal Nodes (Leaves) 88
Figure 2. Schematic Diagram of a Split with n_0 Data Points. 89
Figure 3. Schematic Diagram of the Connection Between the Exposure and Different Modules and Parts of the *E. coli* Model. 90
Figure 4. The Regression Tree for the Cooking Effect Part. 91
Figure 5. Scatter Plot for the Log Reduction in the Number of *E. coli* Organisms versus the Cooking Temperature at Home 92

Part V. Two-Dimensional Sensitivity Analysis: Application of Analysis of Variance to the *E. coli* Food Safety Process Risk Model

Figure 1. Schematic Diagram of the Farm-to-Table Risk Assessment Model for *E. coli O157:H7* in Ground Beef. Source: FSIS, 2001. 139
Figure 2. Case Study Scenario for Application of Sensitivity Analysis to Variability Analysis for Different Uncertainty Realizations of a Model. 140
Figure 3. Case Study Scenario for Application of Sensitivity Analysis to the Co-mingled Variability and Uncertainty Analysis of a Model. 141
Figure 4. Cumulative Probability Function for Maximum Density. 142
Figure 5. Frequencies Associated with Identification of Each Factor in an Importance Group Based on ANOVA. Group 1: The Most Important Factor; Group 2: Factors with Secondary Importance; Group 3: Factors with Minor Importance; and Group 4: Unimportant Factors. 143

Figure 6. Cumulative Probability Distribution of R^2 Values in the Two-Dimensional Analysis for a Model including Main Effects and a Model Including Main and Interaction Effects. 144
Figure 7. Frequencies Associated with Identification of Each Input in an Importance Group Based on: (a) Sample Correlation; and (b) Rank Correlation. Group 1: The Most Important Input; Group 2: Inputs with Secondary Importance; Group 3: Inputs with Minor Importance; and Group 4: Unimportant Inputs....145
Figure 8. Conditional Complementary Cumulative Distribution Functions: (a) Inputs are Successively Conditioned Based on ANOVA Results; and (b) Inputs are Successively Conditioned Based on Pearson Correlation Analysis Results. The Output is Conditioned by Holding 0, 1, 2, 3, 4, 5, and 6 Most Important Inputs at Their Mean Values....146
Figure 1. Case Study Scenario for Application of Sensitivity Analysis to Variability Analysis for Different Uncertainty Realizations of a Model. 181
Figure 2. Frequencies Associated with Identification of Each Input in an Importance Group Based on: (a) Pearson Correlation; and (b) Spearman Correlation. Group 1: The Most Important Input; Group 2: Inputs with Secondary Importance; Group 3: Inputs with Minor Importance; and Group 4: Unimportant Inputs....182
Figure 3. Mean Ranks and Corresponding 95 Percent Confidence Intervals Based on Pearson Correlation Analysis (PCA) and Spearman Correlation Analysis (SCA). 183
Figure 4. Frequencies Associated with Identification of Each Input in an Importance Group Based on: (a) Sample Regression; and (b) Rank Regression. Group 1: The Most Important Input; Group 2: Inputs with Secondary Importance; Group 3: Inputs with Minor Importance; and Group 4: Unimportant Inputs....184
Figure 5. Mean Ranks and Corresponding 95 Percent Confidence Intervals Based on Sample Regression Analysis (SRA) and Rank Regression Analysis (RRA). 185
Figure 6. Cumulative Probability for Coefficient of Determination, R^2, in Sample Regression Analysis (SRA) and Rank Regression Analysis (RRA). 186
Figure 7. Frequencies Associated with Identification of Each Input in an Importance Group Based on: (a) Sample Stepwise Regression; and (b) Rank Stepwise Regression. Group 1: The Most Important Input; Group 2: Inputs with Secondary Importance; Group 3: Inputs with Minor Importance; and Group 4: Unimportant Inputs....187
Figure 8. Mean Ranks and Corresponding 95 Percent Confidence Intervals Based on Sample Stepwise Regression Analysis (SSRA) and Rank Stepwise Regression Analysis (RSRA). 188
Figure 9. Mean Ranks and Corresponding 95 Percent Confidence Intervals Based on Pearson Correlation (PCA), Spearman Correlation (SCA), Sample Regression (SRA), Rank Regression (RRA), Sample Stepwise Regression (SSRA), and Rank Stepwise Regression (RSRA) Analyses for: (a) CHeff; (b) Norg; (c) TNC; and (d) TNI. 189

LIST OF TABLES

Part III. Recommended Practice Regarding Selection of Sensitivity Analysis Methods Applied to Microbial Food Safety Process Risk Models

Table 1. Summary of Key Characteristics of Selected Sensitivity Analysis Methods 60

Part IV. Evaluation of Classification and Regression Trees for Sensitivity Analysis: Case Study for the *E. coli* Food Safety Process Risk Model

Table 1. Input Variables and the Corresponding Mean Value and 95 Percent Probability Range in the Cooking Effect Part 94
Table 2. Input Variables and the Corresponding Mean Value and 95 Percent Probability Range in the Slaughter Module.....95
Table 3. Specification of Precooking Treatments in the Cooking Effect Part.....96
Table 4. Reduction in Deviance Associated with Selected Inputs in the Regression Tree Generated in the Cooking Effect Part 97
Table 5. Comparison of the Results for Stepwise Regression Analysis and CART in the Cooking Effect Part.....98
Table 6. Reduction in Deviance Associated with Selected Inputs in the Regression Tree Generated in the Co-mingled Analysis of Variability and Uncertainty in the Slaughter Module.....99
Table 7. Comparison of the Results for Stepwise Regression Analysis, Correlation Analysis, and CART in the Slaughter Module 100

Part V. Two-Dimensional Sensitivity Analysis: Application of Analysis of Variance to the *E. coli* Food Safety Process Risk Model

Table 1. Input Variables and Corresponding Mean Values, 95% Probability Ranges, and Distribution Types in the Growth Estimation Part of the *E. coli* Model (FSIS, 2001) 132
Table 2. Levels Defined for Factors in the Growth Estimation Part of the *E. coli* Model 133
Table 3. Summary of the ANOVA Results for Two-Dimensional Variability Simulation for Different Uncertainty Realizations.....134
Table 4. Summary of Sample (Pearson) and Rank (Spearman) Correlation Analyses Results for Two-Dimensional Variability Simulation for Different Uncertainty Realizations.....135
Table 5. Summary of the ANOVA Results for Co-Mingled Analysis of Variability and Uncertainty ($R^2 = 0.82$).....136
Table 6. Summary of the ANOVA Results for 200 Bootstrap Simulations for F value Statistics 137
Table 7. Evaluation of ANOVA Contrasts for the Growth Estimation Regarding the Interactions Between Storage Temperature and Storage Time at Stage 3.....138

Part VI. Evaluation of Sampling-Based Methods for Sensitivity Analysis: Case Study for the *E. coli* Food Safety Process Risk Model

Table 1. Input Variables and Corresponding Mean Values, 95% Probability Ranges, and Distribution Types in the Slaughter Module of the *E. coli* Model.....175
Table 2. Summary of Pearson and Spearman Correlation Analyses Results 176
Table 3. Summary of Sample Standardized Regression Analysis Results.....177
Table 4. Summary of Rank Regression Analysis Results.....178
Table 5. Summary of Sample Stepwise Regression Analysis and Rank Stepwise Regression Analysis Results.....179
Table 6. Summary of Key Characteristics of Selected Sampling-Based Sensitivity Analysis Methods.....180

PART I

INTRODUCTION

1.0 Introduction

Quantitative risk assessment is an emerging tool in the field of microbial food safety (Rose and Sobsey, 1993; Jaykus, 1996). In particular, there has been increasing development and use of such models for food safety process risk assessment (Nauta, 2000). Examples include models for *Salmonella* (SERAT, 1998), *Vibrio parahaemolytics* (CFSAN, 2001), *E. coli O157:H7* (FSIS, 2001), and *Listeria monocytogenes* (CFSAN, 2003). These models typically include hundreds of parameters and inputs that have uncertainty and/or variability. These models can be large and complex with typical characteristics such as nonlinearities, interactions between inputs, thresholds or saturation points, and use of both continuous and categorical inputs (Frey, 2002). Interaction refers to a case in which the effect of an input to the model response depends on the value of another input, while nonlinearity is a relationship between two variables in which the change in each variable is not simply proportional to the change in the other variable. Due to the characteristics of MFSPR models, it can be difficult to prioritize controllable variables that are most promising with respect to risk management goals. As a matter of good practice, it is important to evaluate how a model responds to changes in its inputs as part of the process of model development, verification, and validation.

Sensitivity analysis is a methodology for identifying the key model inputs that contribute the most to variation in a selected model output or to the highest values of the output. Thus, an analyst is often interested in identifying which one or subset of potentially many controllable variables is of greatest importance with regard to human exposure and risk. In the context of food safety, an important input is one that not only has a significant influence on the variation of the output or any desired measure of sensitivity (Hofer, 1999; Saltelli et al., 2000; Helton and Davis, 2002), but also is controllable in practice and could be changed in order to reduce risk. For example, Critical Control Points (CCPs) are points, steps, or procedures in the process of

bringing food from farm to the table at which direct control can be applied, and a food safety hazard can be prevented, eliminated, or reduced to an acceptable level (Hulebak and Schlosser, 2002; Seward, 2000). Potential CCPs are typically incorporated in microbial food safety process risk models. Sensitivity analysis results can lead to identification of important CCPs among the list of inputs. Other reasons for performing sensitivity analysis include: assisting the process of model development; prioritizing additional work aimed at reducing uncertainty in an assessment; and providing insight into the robustness of model results when making decisions (Cullen and Frey, 1999).

1.1 Objectives

The objectives of this project are to identify, review, and evaluate sensitivity analysis methods for application to microbial food safety process risk models and provide guidance for practitioners regarding application of sensitivity analysis based on case studies with the *E. coli O157:H7* food safety process risk model developed primarily by USDA (FSIS, 2001). These objectives serve as an aid in identifying potential control points along the farm-to-table continuum, to inform decisions about food safety research and data acquisition priorities, and to contribute to the development of sound food safety regulations.

The analyses of this dissertation are targeted to answer several key questions that address the overall objectives of the research. The answers to these questions are discussed in Part VII:

- When should sensitivity analysis be performed?
- What are key considerations in preparation of a model to facilitate sensitivity analysis?
- How can one select among typical sensitivity analysis methods?
- What are the considerations in presenting results of sensitivity analysis methods?

- Can conventional sensitivity analysis methods provide robust insights in spite of their apparent limitations?
- How can one discriminate the importance between closely ranked inputs?
- What are the implications of sensitivity analysis in a two-dimensional probabilistic framework compared to the one-dimensional analysis of variability and uncertainty?

1.2 Overview of Research

This dissertation is the result of a three phase project. An objective of the project was to transfer, apply, and adapt sensitivity analysis methods developed in other disciplines (e.g. complex engineering systems) to microbial food safety risk assessment.

During the first year of this three year cooperative agreement, NCSU prepared a literature review of sensitivity analysis methods and convened a workshop of sensitivity analysis experts and food safety process risk model practitioners. Several “white” papers were presented during the first day of the workshop (Frey and Patil, 2002; Helton and Davis, 2002; Hulebak and Schlosser, 2002; Kohn, 2002; Pate-Cornell, 2002; Thompson, 2002, Saltelli, 2002). These papers have been revised and published in Risk Analysis as a special section on sensitivity analysis, supplemented with an introduction that summarized the recommendations of the workshop (Frey, 2002). The participants of the workshop formulated recommendations in response to specific trigger questions:

- What are the key criteria for sensitivity analysis methods applied to food safety risk assessment?
- What sensitivity analysis methods are most promising for application to food safety risk assessment?
- What are the key needs for implementation and demonstration of such methods?

The first question addresses the key challenges associated with sensitivity analysis applied to food safety risk assessment, while the latter two questions pertain to recommendations for identification, evaluation, and selection of methods and for transferring results into the hands of practitioners.

A key criterion for sensitivity analysis and for the risk model and analysis in general, is that it must be relevant to a decision. This means that the model output of interest must be directly related to the decision. For example, if a decision is informed by whether risk is above or below a threshold, then the model output should be a variable indicating the probability that the estimated risk is above or below the threshold. The sensitivity analysis should pertain to variation in inputs that causes a change in the value of this output that would lead to a different decision.

Technical requirements of a sensitivity analysis method are manifold and may differ from one application to another, and from one decision application to another. An ideal sensitivity analysis method would be model independent. Specifically, the sensitivity analysis method should not require the introduction of any assumptions regarding the functional form of the risk model and, therefore, should be applicable to a wide range of different model formulations. Although sensitivity analysis methods that are based upon linear formulations and additivity are relatively convenient and easy to apply, it is important that methods be used that take into account the simultaneous interaction among multiple inputs, especially for nonlinear models that contain thresholds.

The method should provide not just a rank ordering of key inputs, but also some quantitative measure of the sensitivity of each input so that it is possible to distinguish the most strongly sensitive inputs from those with weaker influence on the selected model output. For

example, is the most sensitive of the inputs substantially more important than the second ranked input, or do the top two inputs have approximately equal influence on the model output?

The results of both the risk analysis and the sensitivity analysis should be interpreted in terms of plausibility with respect to governing biological, physical, chemical, and other processes. The results should be explainable to key target audiences, such as risk managers and stakeholders.

Because it is unlikely that one sensitivity analysis method will meet all of the criteria for an ideal method, the group agreed that it would be necessary to apply two or more sensitivity analysis methods in a given context in pursuit of obtaining robust insight regarding key sensitivities based upon different sensitivity analysis measures. Methods that have been peer reviewed, have been clearly demonstrated, and that are readily available will be more readily accepted by practitioners.

In response to the second trigger question, the group did not identify specific methods. Instead, the group emphasized the key criteria that were generated in response to the first question. For example, methods that can deal with interactions, nonlinearities, discontinuities, and discrete inputs would be preferred over methods that cannot. Methods that are global or generic are likely to be more promising than other types of methods.

Before applying a sensitivity analysis method, it may help to reduce the computational burden by narrowing down the search space among the input parameters. For example, if adverse consequences do not occur unless a temperature exceeds a threshold above which microbial growth becomes significant, it may not be necessary or important to analyze model behavior when the temperature is below the threshold. Thus, the search space could be narrowed to cases where the temperature is above the threshold in order to reduce computational time.

In response to the third trigger question, the group agreed that there was a need to explore multiple sensitivity analysis methods and apply them to more than one food safety risk model. The methods should be tested at research institutes and efforts should be made to confirm or validate the results. The process of testing methods will help in establishing a track record for specific methods applied to food safety process risk models.

Based upon experience with sensitivity analysis methods and representative food safety process risk models, practitioners indicated that it would be useful to have a guideline regarding the use of sensitivity analysis methods. The guideline should not be too prescriptive, but should provide useful boundaries and principles for selecting, using, and interpreting sensitivity analysis methods, as well as in reporting results. A comparison of methods, taking into account real life constraints, should be part of the guideline. The guideline should outline a tiered approach to sensitivity analysis. The guideline could make a useful contribution by clearly defining terminology. Because of the interdisciplinary nature of risk assessment and of sensitivity analysis, terms may be defined differently by different experts or practitioners. Therefore, a common reference regarding such definitions is important.

In Phase 2, several sensitivity analysis methods were evaluated based upon application to two microbial food safety process risk models (Frey et al., 2003): (1) *Listeria monocytogenes* in Ready-to-Eat (RTE) foods (CFSAN, 2003); and (2) *E. coli O157:H7* in ground beef (FSIS, 2001). Methods were selected from each of the three categories of sensitivity analysis techniques defined by Frey and Patil (2002) including mathematical, statistical, and graphical. Selected methods covered a wide range of techniques from simplistic sensitivity analysis approaches such as nominal range sensitivity analysis (NRSA) (Cullen and Frey, 1999) to complex techniques such as analysis of variance (ANOVA) (Neter et al., 1996) and classification and regression trees

(CART) (Breiman et al., 1981). A key output of Phase 2 was a summary table comparing the capabilities of many sensitivity analysis methods with regard to the characteristics of MFSPR models and with regard to analytic objectives.

The objective of the third and final phase of the project was to develop a document for practitioners regarding guidance on the selection and application of sensitivity analysis methods for food safety process risk models. The guidance was based upon input from: (1) Phase 1 of the project, including the review of sensitivity analysis methods by NCSU and recommendations from the workshop; (2) Phase 2 of the project, including results from the applications of multiple sensitivity analysis methods with respect to two food safety process risk models; and (3) a workshop conducted at Office of Cost-Benefit Analysis and Risk Assessment, USDA, Washington (March, 2003).

1.3 Organization

Part II discusses the key considerations in preparing or developing a model in order to facilitate sensitivity analysis. This part is a summary of the experience obtained in Phase 2 of the project from application of several sensitivity analysis methods to the *Listeria* and *E. coli* models. Key issues include model structure, identification and accessing of model inputs, decision-based selection of model outputs, specification of the type of probabilistic simulation to be performed, and suggestions for how to modify an existing model. Recommendations also are made regarding the development of new food safety process risk models. These recommendations involve modeling environment, determination of the type of probabilistic simulation, modeling strategies, and documentation of models.

Four manuscripts that the author has submitted or plans to submit for publication in peer-reviewed journals comprise Parts III through VI of this dissertation. More detail is available in project reports prepared for USDA (Frey et al., 2003; Frey et al., 2004).

The paper given in Part III of this dissertation provides a synthesis of information and lessons learned across various disciplines to provide a framework to practitioners regarding selection of sensitivity analysis methods applied to microbial food safety process risk models. However, the framework can be applied to other exposure and risk models with similar characteristics. The situations in which the application of sensitivity analysis to microbial food safety process risk models is recommended are discussed. A decision framework for method selection is introduced. The decision framework can substantially facilitate the process of selecting a sensitivity analysis method for a practitioner.

The manuscript presented in Part IV evaluates classification and regression trees (CART) for sensitivity analysis of microbial food safety process risk models. CART is evaluated on the basis of its ability to address typical characteristics of microbial food safety process risk models such as nonlinearities, interaction, threshold and saturation points, and categorical inputs. Moreover, this method is evaluated with respect to identification of high exposure scenarios and corresponding key inputs and critical limits. Two alternative approaches for ranking sensitive inputs are explored and compared. The ability to identify key inputs and critical values of such inputs that produce especially high (or low) values of an output was illustrated. The implications of sensitivity analysis based on CART with respect to development of risk management alternatives are discussed.

The manuscript presented in Part V focuses on the concept of sensitivity analysis in a two-dimensional probabilistic framework and its implication in providing more robust results

and confidence in insights regarding key model inputs. Analysis of variance (ANOVA) is selected as a sensitivity analysis method due to its ability to address the typical characteristics of microbial food safety process risk models. Uncertainty associated with the ranks of inputs is quantified. Bootstrap simulation is demonstrated as an approach for discrimination of closely ranked inputs.

The manuscript presented in Part VI evaluates selected sampling-based sensitivity analysis methods. The case studies in this paper develop insights with respect to the behavior and effectiveness of the techniques under consideration when applied to microbial food safety process risk models. A qualitative comparison of the selected sampling-based methods is provided.

Part VII summarizes the conclusions and recommendations based on the analyses provided in Parts III to VI in order to answer the questions raised in Section 1.1. Appendix A provides a glossary of technical terms to assist the reader.

Each part of this manuscript has its own list of references cited.

1.4 References

Breiman, L., Friedman, J.H, Olshen, R.A., and Stone, C.J., 1984. Classification and regression trees. Belmont, Calif.: Wadsworth International Group.

CFSAN, 2001. Draft risk assessment on the public health impact of *Vibrio parahaemolyticus* in raw molluscan shellfish. Prepared by the *Vibrio parahaemolyticus* risk assessment task force (Center for Food Safety and Applied Nutrition) for Food and Drug Administration (FDA), USDA, Washington, DC.

CFSAN, 2003. Quantitative assessment of relative risk to public health from foodborne *Listeria monocytogenes* among selected categories of ready-to-eat foods. Prepared by Center for Food Safety and Applied Nutrition, Food and Drugs Administration and Food Safety and Inspection Service, U. S. Department of Agriculture, Washington DC, September 2003, http://www.foodsafety.gov/~dms/lmr2-toc.html.

Cullen, A.C., Frey, H.C., 1999. Probabilistic techniques in exposure assessment. Plenum Press: New York.

Frey, H.C., 2002. Introduction to special section on sensitivity analysis and summary of NCSU/USDA workshop on sensitivity analysis. Risk Analysis, 22(3): 539-545.

Frey, H.C., Mokhtari, A., Tanwir, D., 2003. Evaluation of selected sensitivity analysis methods based upon applications to two food safety process risk models. Prepared by North Carolina State University for office of risk assessment and cost-benefit analysis, U.S. Department of Agriculture, Washington, DC, June 2003, www.ce.ncsu.edu/risk.

Frey, H.C., Mokhtari, A., Zheng, J., 2004. Recommended practice regarding selection, application, and interpretation of sensitivity analysis methods applied to food safety process risk models. Prepared by North Carolina State University for office of risk

assessment and cost-benefit analysis, U.S. Department of Agriculture, Washington DC , January 2004, www.ce.ncsu.edu/risk.

Frey, H.C., Patil, R., 2002. Identification and review of sensitivity analysis methods. Risk Analysis, 22(3): 553-577.

FSIS, 2001. Draft risk assessment of the public health impact of *E. coli O157:H7* in ground beef. Prepared by the *E. coli O157:H7* Risk Assessment Team for the Food Safety and Inspection Service and United States Department of Agriculture, Washington, DC. http://www.fsis.usda.gov/OPPDE/rdad/FRPubs/00-023N/00-023NReport.pdf.

Helton, J.C., Davis, F.J., 2002. Illustration of sampling-based methods for uncertainty and sensitivity analysis. Risk Analysis, 22(3): 591-622.

Hofer, E., 1999. Sensitivity analysis in the context of uncertainty analysis for computationally intensive models. Computer Physics Communications, 117(1-2): 21-34.

Hulebak, K.L, Schlosser, W., 2002. Hazard analysis and critical control point (HACCP) history and conceptual overview. Risk Analysis, 22(3): 553-578.

Jaykus L.A., 1996. The application of quantitative risk assessment to microbial food safety risks. Critical Reviews in Microbiology; 22: 279-93.

Kohn, M.C., 2002. Use of sensitivity analysis to assess reliability of metabolic and physiological models. Risk Analysis, 22(3): 623-631.

Nauta, M.J., 2000. Separation of uncertainty and variability in quantitative microbial risk assessment models. International Journal of Food Microbiology, 57(1-2): 9-18.

Neter, J., Kutner, M.H., Nachtsheim, C.J., and Wasserman, W. 1996. Applied Linear Statistical Models, Fourth Edition. McGraw-Hill: Chicago, IL.

Pate-Cornell, E., 2002. Risk and uncertainty analysis in government safety decisions. Risk Analysis, 22(3): 633-646.

Rose, J.B., Sobsey, M.D., 1993. Quantitative risk assessment for viral contamination of shellfish and coastal waters. Journal of Food Protection, 56(12): 1043-1050.

Saltelli, A., 2002. Sensitivity analysis for importance assessment. Risk Analysis, 22(3): 579-590.

Saltelli A., Chan, K., Scott, M., 2000. Sensitivity analysis, probability and statistics series. John Wiley & Sons: New York.

SERAT, 1998. Salmonella Enteritidis risk assessment for shell eggs and egg products: Final Report. Prepared by Salmonella Enteritidis Risk Assessment Team for The Food Safety and Inspection Service (FSIS), USDA, Washington, DC.

Seward, S., 2000. Application of HACCP in food service. Irish Journal of Agricultural and Food Research, 39(2): 221-227.

Thompson, K.M., 2002. Variability and uncertainty meet risk management and risk communication. Risk Analysis, 22(3): 647-654.

PART II

GUIDANCE FOR PREPARATION OF EXISTING OR NEW MODELS TO FACILITATE SENSITIVITY ANALYSIS

2.0 Introduction

Prior to application of any sensitivity analysis methods, the model under study should be prepared for the analysis. In particular, it is important that the model is coded in a manner such that the inputs and outputs are clearly identifiable and accessible. Furthermore, the characteristics of the model, such as modularity, binning, aggregation, and probabilistic simulation options, may constrain the use of particular sensitivity analysis methods. Sensitivity analysis should be included in the list of primary modeling objectives at the time of model development. The implementation of specific model development strategies will facilitate sensitivity analysis. For an existing model, the practitioner is typically interested in applying sensitivity analysis with minimum modification to the model. However, in some situations, if the model has not been designed to facilitate sensitivity analysis, substantial modifications may be required (e.g., Patil and Frey, 2003).

Generally, a thorough understanding of the model and its limitations is essential to select well-suited sensitivity analysis methods and to determine the scope of sensitivity analysis application. The scope of sensitivity analysis may include the entire model or could be focused on specific modules or parts of a model. For example, an analyst may focus sensitivity analysis on the exposure module, which is a typical part of food safety process risk models.

This chapter focuses on the preparation of food safety process risk models for sensitivity analysis. These models are classified into two categories; (1) existing models; and (2) new models. For existing models, discussed in Section 2.1, the key steps prior to application of sensitivity analysis methods are given in Sections 2.1.1 through 2.1.5.

Section 2.2 focuses on modeling strategies and recommendations for development of new food safety process risk models in order to facilitate the application of sensitivity analysis methods.

2.1 Preparation of Existing Models for Sensitivity Analysis

The characteristics of existing models have a direct influence on the choice of sensitivity analysis methods and the scope of sensitivity analysis. In some cases, the modelers might not have anticipated the application of sensitivity analysis, and hence, the model may not have been developed in a manner that facilitates sensitivity analysis. To identify whether the modeling methodology used is compatible with application of sensitivity analysis, the model has to be thoroughly reviewed and characterized. The following sections discuss the key features that need to be studied in the process of understanding a model.

2.1.1 Identification of Model Structure

The identification of model structure helps determine the scope of sensitivity analysis. Important model characteristics that are related to model structure include modularity, binning, and aggregation.

A general framework for performing quantitative food safety risk assessment is the modular process risk model (MPRM) (Nauta 2001a&b). Modularity is a way of organizing a model by breaking the overall model into sub-components. For example, each step or key activity in a food safety process risk model can be modeled in a separate module characterized by inputs and outputs. An example of modularity in food safety risk assessment is the *E. coli* model in which there are separate modules for processes such as slaughter of cattle and preparation of ground beef servings (FSIS, 2001). Conceptually, independent processes or components can be modeled as separate modules. Procedures that are repeated within the model

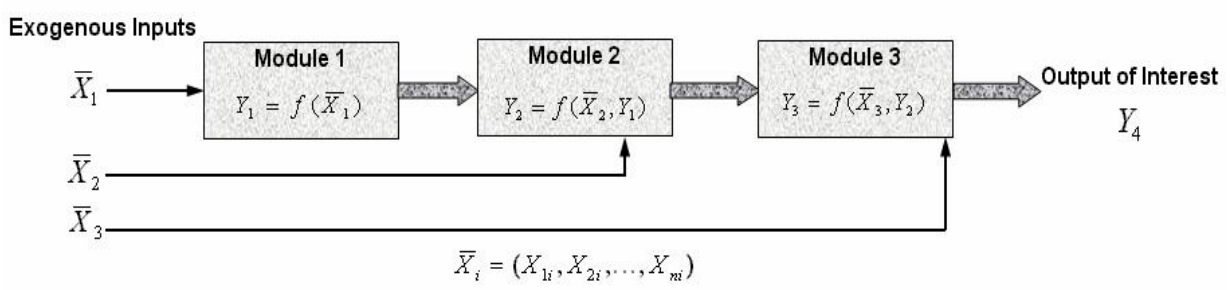


Figure 1. Conceptual Example of a Modular Modeling Framework Comprised of Multiple Modules in Series.

also can be represented using modules. In a modular framework, an output from a module may be an input to other modules. In this manner, modules are connected to each other. For example, an output representing the contamination level of meat trim from the slaughter module can be an input to the preparation module representing the initial contamination of meat trim when modeling the grinding process.

The ability to apply sensitivity analysis typically depends upon a one-to-one correspondence between values of an output and values of an input. Thus, if the existing model utilizes a modular structure, in order to apply sensitivity analysis across modules in a framework composed of multiple modules, there should be one-for-one mapping of every input value to the output values, where the input of interest may be in one module and the output of interest could be in a dependent module. In a modular framework, modules can be connected in two ways: (1) serial connection; and (2) parallel connection. Figure 2-1 depicts a conceptual modeling framework comprised of three modules in series. Each module has exogenous inputs (i.e., $\overline{X}_1$, $\overline{X}_2$, and $\overline{X}_3$). Furthermore, modules that depend upon other modules also have internal inputs. The internal inputs are the outputs from the predecessor module (i.e., Y_1, and Y_2). Thus, a module output is influenced directly by internal and exogenous inputs for the immediately preceding module, and indirectly by the exogenous inputs for each predecessor module. Since the exogenous inputs typically include variability and uncertainty of interest to the analyst or

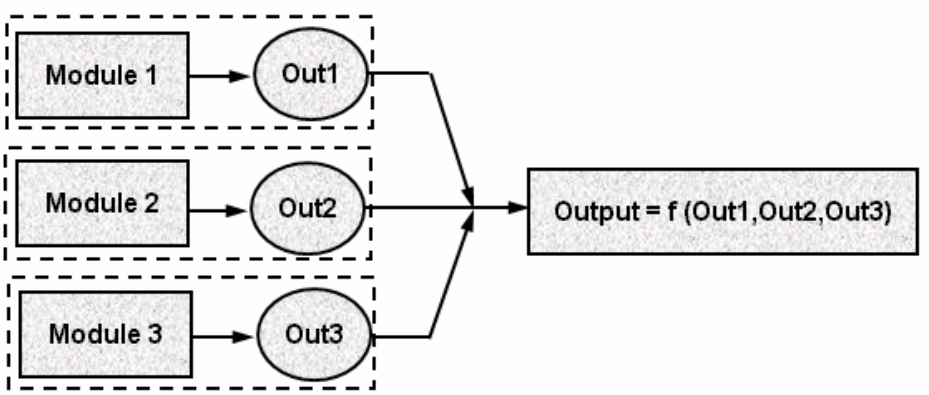


Figure 2. Conceptual Framework of Modules in Parallel.

decision maker, the typical desired objective of sensitivity analysis would be to evaluate the sensitivity of the selected output with respect to the exogenous inputs for all of the predecessor modules. A serial model structure that maintains the one-to-one correspondence between exogenous inputs and the selected output facilitates sensitivity analysis.

Modules can also be in parallel, as illustrated in Figure 2-2. The final model output is function of the outputs from each of several parallel modules. The key difference between the parallel and series frameworks is that in the parallel framework each module is independent from the other modules. This structure permits identification of the module to which the model output has the highest sensitivity on a relative basis. Frey et al. (2003) evaluated a case study with the *E. coli* model in which serving contamination, growth effect, and cooking effect modules had a parallel connection to the model output. A case study was prepared to prioritize the effect of these modules on the final exposure in ground beef serving. For the case study, each module was replaced by its corresponding output probability distribution. For evaluation of the priority rank of the distributions of outputs with respect to variability in exposure, four cases were examined using Monte Carlo simulation. Case *Zero* represented the situation in which all three variables (i.e., serving contamination, growth effect, and cooking effect) varied based on their distributions. For each of the other three cases, the output of one of the modules varied based on its

distribution, while the other two outputs were conditioned at their mean values. Comparison of the results for these four cases helped identify which of the four cases had the largest maximum contamination or the highest probability of exceeding contamination levels that would be considered to be high. Results of the analysis indicated that cooking caused a maximum range of variation in the exposure to *E. coli O157:H7* organisms and that growth resulted in the highest number of consumer exposures to positive doses of *E. coli O157:H7* organisms in ground beef.

Sometimes modelers employ techniques that result in a loss of one-to-one correspondence between exogenous inputs to predecessor modules and the output of interest. These techniques might have been selected because they offered a computational convenience, provided a useful intermediate summary, or facilitated exchange of data between modules. Examples of these situations may include: (1) parameterizing empirical values of internal inputs into theoretical distributions; (2) summarizing continuous internal inputs into intervals (i.e., binning); and (3) aggregating which leads to many-to-one relationship between input and output values.

The one-to-one correspondence between an output and model inputs may be lost, for example, by using the inputs to estimate the parameters of a theoretical distribution that serves as an intermediate output for the model.

Another example leading to loss of one-to-one correspondence between an output of a module and its exogenous predecessor inputs is the use of binning for internal inputs. Binning is a commonly used summarization technique. For example, in the *E. coli* model the contaminant concentration in beef trimings is simulated as a continuous variable, but subsequently is binned by 0.5 log increments from 0 to 8 logs (FSIS, 2001). The estimated contamination for the combo bin is rounded to the next upper level. For instance, if the estimated contamination is 0.1 logs, it

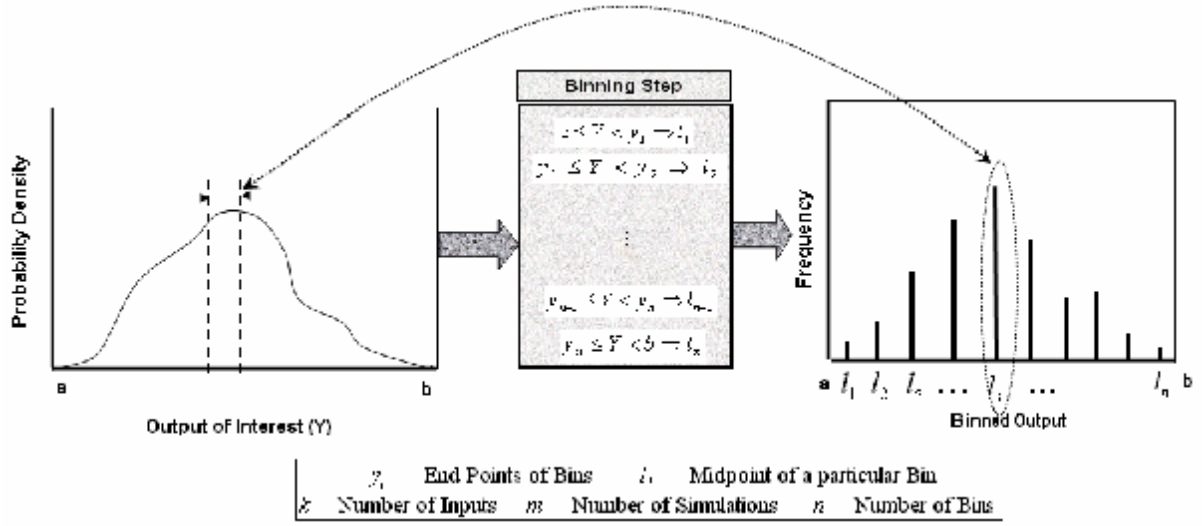


Figure 3. Conceptual Example of the Binning Approach, Leading to Loss of One-to-One Correspondence.

is binned into a range of 0 to 0.5 logs and is quantified as 0.5 logs. Another example where binning was used is the exposure module in *Listeria* model in which several individual meal servings were mapped to the same dose bin. After binning, the specific simulated meal serving that resulted in a particular dose cannot be identified. The exposure range is binned into half logs. The fraction of serving contained in each bin is calculated (CFSAN, 2003).

Figure 2-3 illustrates how the binning approach leads to loss of one-to-one correspondence between an output and model inputs. The figure shows a continuous distribution of the pre-binned output values. In the binning step, *n* bins are defined using equal intervals. There is a one-to-one correspondence between the pre-binned output values and sampled values of the model inputs. For example, a value of y_j of the pre-binned output corresponds to a unique sample from the distributions of the model inputs. This figure shows that the problem arises when binned output values are used instead of the pre-binned output values. For example, if the j^{th} bin is selected from the binned output distribution, it is not possible to trace back the values of the inputs associated with the selected binned output value. This figure shows that there is a range of pre-binned output values associated with the j^{th} bin, and hence, there are not unique

values of the model inputs associated with the estimated output value. In order to eliminate the binning in an existing model, it might be necessary to substantially change the model structure. Such a change may be beyond the scope of the analysis. Therefore, under this circumstance, modular-based sensitivity analysis may be considered.

Aggregation refers to situations in which multiple numerical values are combined into one numerical value. For example, in the *E. coli* model, the contamination level in a combo bin is estimated based upon contamination contributed by individual meat trims from multiple slaughtered cattle. Meat trim are from cattle that are slaughtered, dehided, and eviscerated, and are the result of a fabrication process. These steps take place in different parts of the slaughter plant. Each of these steps can be a source of contamination. The combo bin contamination level might be influenced by only a small proportion of the incoming meat trim. The combo bin contamination level is calculated based upon sum of the number of organisms from each of the possible sources of contamination. This type of situation is challenging to sensitivity analysis because there is a many-to-one correspondence between meat trim and the combo bin, whereas sensitivity analysis typically considers one-to-one correspondences. Sensitivity analysis could be based on one or more summary statistics. For example, the sensitivity of the combo bin contamination to the average meat trim contamination, or the most contaminated meat trim could be assessed. However, this would result in some loss of information. A simpler option is to perform sensitivity analysis only on internal inputs that depend on the model output of interest (e.g., the aggregated combo bin contamination levels), This would exclude consideration of inputs prior to the aggregation in the sensitivity analysis. For example, the original code in the *E. coli* slaughter module was modified to introduce a new composite animal that represents the characteristics of all contaminated cattle contributing to contamination in a combo bin. This

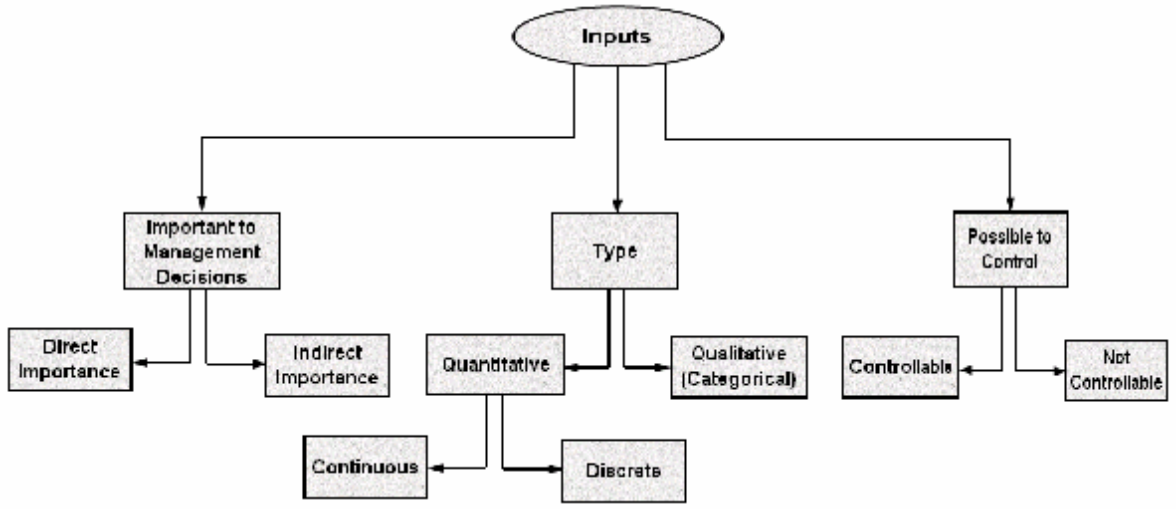


Figure 4. Possible Classifications of Inputs in Food Safety Process Risk Models.

composite animal has the total contamination from different contamination sources, and hence contributes the total number of *E. coli* organisms to a combo bin. Sensitivity analysis for combo bin contamination was performed with respect to the characteristics of the composite animal (Frey et al., 2003).

2.1.2 Identification of Inputs

To perform sensitivity analysis, the inputs of interest must be identified. Inputs can be classified in different ways. Figure 2-4 shows a possible classification of inputs in food safety process risk models. The inputs may be of direct or indirect importance to management decisions (e.g., serving size versus sanitation control). Some inputs may be controllable and others not (e.g., storage temperature versus lag period). Also, the degree of control may vary from one input to another. Inputs may be qualitative (categorical) (e.g., ground beef consumption types) or quantitative (e.g., storage temperature). Quantitative inputs can be continuous (e.g., storage time) or discrete (e.g., number of infected animals in a truck). Some quantitative inputs may be described by empirical distributions, while other inputs may be represented by parametric distributions.

Model inputs may represent variability, uncertainty, or both in a probabilistic simulation. The manner in which uncertainty is modeled may vary among inputs. In some cases, uncertainty may be considered in the parameters of the variability distributions. For example, a triangular distribution may be used to represent variability in storage temperature. A triangular distribution is defined by three parameters representing minimum, mode, and maximum values of the distribution. The mode of this distribution may be uncertain. A uniform distribution may be used to represent the uncertainty for the most likely value. In contrast, uncertainty can be incorporated in an input as a choice of alternative variability distributions. For example, in a case of laboratory data for initial concentration of a pathogen, a few parametric distributions may be good fits to available data. There may exist uncertainty in selecting which parametric distribution better represents the data. Uncertainty can be addressed for the input by weighting fitted variability distributions and randomly selecting each variability distribution proportional to its weight during simulation. For example, four parametric distributions are considered for representation of variability in the initial concentration of the *Listeria monocytogenes* pathogens (CFSAN, 2003). These distributions include Beta, Weibull, Triangular, and Lognormal. Each parametric variability distribution is selected randomly in alternative realizations of uncertainty of the model to represent the uncertainty in the choice of a variability distribution.

2.1.3 Selection of Model Output for Sensitivity Analysis

The selection of a model output for sensitivity analysis depends on the assessment objectives. For example, the output of interest may be different when objective varies from reduction in the risk of mortality in the population to reduction in the risk of mortality per serving. The relative importance of inputs is likely to vary according to the choice of output. An analyst should clearly determine the type of management objective that is expected before

selecting the output of interest. Frey et al. (2003) investigated case studies with the *E. coli* model in which the relative importance of inputs varied based upon the selection of output related to the risk reduction strategies. For example, when the number of contaminated ground beef servings was considered as the output of interest, the cooking effect and related inputs, including cooking temperature and pre-cooking handlings, were identified as key. In contrast, when the amount of contamination in ground beef servings was selected as the output of interest, growth and related inputs, including storage times and temperatures, were characterized as key.

2.1.4 Probabilistic Simulation and Its Implication for Sensitivity Analysis

The scope and detail of inferences drawn from sensitivity analysis depend on the choice of probabilistic simulation dimensions. The presence or absence of variability and uncertainty dimensions in the modeling process should be identified prior to selection of a sensitivity analysis approach. Food safety process risk models typically have one-dimensional or two-dimensional probabilistic frameworks incorporating variability, uncertainty, or both in the simulation. The selection of the specific probabilistic dimension of a model for sensitivity analysis depends on the assessment objective. Figure 2-5 shows a schematic diagram of alternative assessment objectives that are appropriate for the two modeling frameworks. The assessment objectives in this figure include: (1) variability only; (2) uncertainty only; (3) variability for different realizations of uncertainty; (4) uncertainty for different realizations of variability; and (5) co-mingled variability and uncertainty. The insights regarding sensitivity that can be obtained from these assessment objectives are discussed in Part III.

For food safety process risk models that have a one-dimensional probabilistic framework, a practitioner may select variability only, uncertainty only, or co-mingled variability and uncertainty analysis. For example, Patil and Frey (2003) evaluated case studies with the FDA

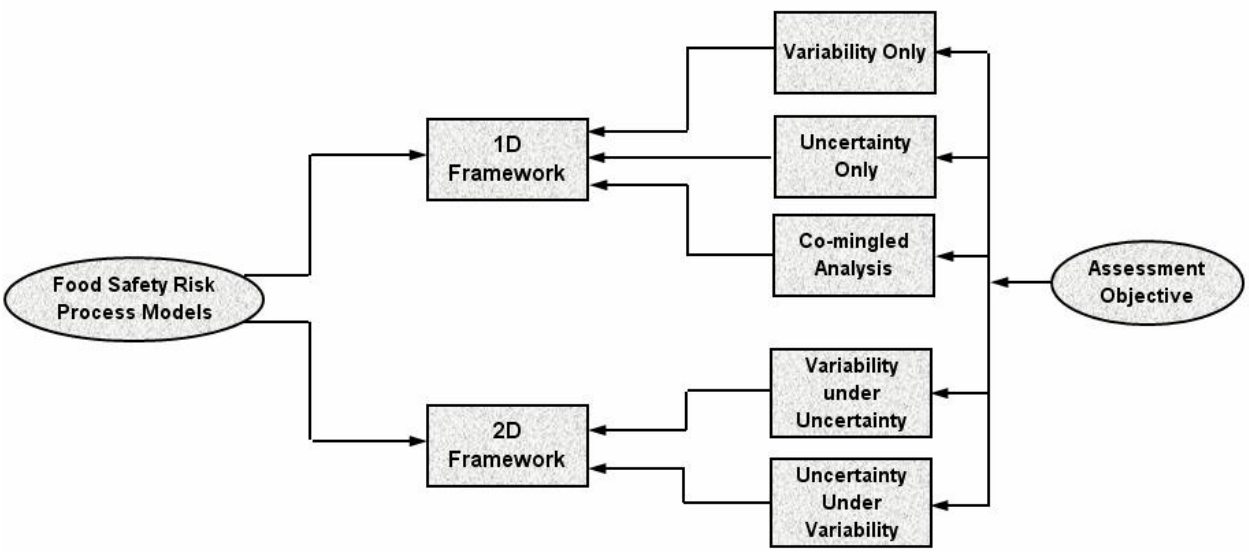


Figure 5. Alternative Frameworks for Food Safety Process Risk Models and Selection of the Probabilistic Dimension for Sensitivity Analysis Based on the Assessment Objective.

Vibrio (CFSAN, 2001) model that has a one-dimensional probabilistic framework considering only variability in the model inputs, and hence, sensitivity analysis was focused on variability. A two-dimensional framework may be used to quantify variability and uncertainty simultaneously. When analyzing the results, the analyst may prefer to focus on characterizing the range of uncertainty regarding estimated distributions for inter-individual variability in exposure and risk, or to focus on quantifying uncertainty in the exposure or risk for a particular percentile or statistic from the variability dimension. In a two-dimensional framework, it is also possible to accommodate one-dimensional cases such as for variability only, uncertainty only, or co-mingled variability and uncertainty. For studies with major policy-based consequences, a practitioner may need to apply comprehensive sensitivity analyses that fully distinguish between variability and uncertainty depending on decision criterion. For example, when the decision criterion is to provide specific confidence that selected percentile in the exposed population falls below a threshold of concern, a practitioner would typically need to fully distinguish between variability and uncertainty. In contrast, for preliminary studies a practitioner may decide to focus only on variability or uncertainty.

Available resources also may affect the selection of the number of dimensions of the probabilistic simulation when performing sensitivity analysis. Two-dimensional sensitivity analysis is generally resource intensive. It demands computational resources both during the creation of datasets and the application of sensitivity analysis. The two dimensional analysis may not be practicable for sensitivity analysis methods that cannot be automated. In contrast, a one-dimensional analysis often requires fewer resources. Although application of sensitivity analysis on a single probabilistic dimension of the model is substantially easier, the credibility and robustness of the insights obtained with respect to the sensitivity may be reduced because the one-dimensional analysis often assumes that the other probabilistic dimension of the model can be replaced with deterministic values. For example, case studies performed by Frey et al. (2003) in which variability and uncertainty were fully differentiated show substantial ambiguity in ranks of key variability sources when uncertainty is considered. Quantifying the degree of ambiguity in the ranks of inputs helps in making management decisions that are robust to uncertainty.

2.1.5 Modification of Models

In some situations a model must be modified to apply sensitivity analysis. Some common modifications are discussed in this section. These modifications often demand changing data storage procedures, including model inputs, outputs, and internal inputs. Additional modifications may be required for some specific models.

In order to perform sensitivity analysis, a dataset including values generated during the probabilistic simulation for each input and corresponding output values should be available. As a common practice in programming, generated values of each input are stored dynamically in the model. Typically, old values of an input from a previous iteration are replaced with new values to conserve the memory space and run time. Thus, it may be necessary to modify the model in

order to store these values for future reference. Based upon the modeling environment, these values can be stored in a separate data file, spread sheet, or inside the model in an array. The software package used for application of sensitivity analysis may influence the format that would be used for data storage. For example, SAS© can read formats such as comma-separated values (.csv).

Storage of intermediate values estimated during the simulation of a model is useful in some cases. For example, in situations in which values of an internal input are binned before using in other parts of the model, it is helpful to store individual values of the intermediate input prior to binning. If the output of interest is calculated in the original model based upon a binned internal input, then the model should be modified, as discussed in Section 2.1.1, so that the output is calculated for individual values of the internal input.

2.2 Preparation of New Models for Sensitivity Analysis

To facilitate the application of sensitivity analysis when developing a new model, it is essential to consider the proper use of modeling techniques and model implementation. Good software engineering practices in general help application of sensitivity analysis (Frey et al., 2003). This section discusses some measures that can be used to facilitate sensitivity analysis when building a new model. These measures involve modeling environments, different approaches for characterizing variability and uncertainty, modeling strategies, and model documentation, which are discussed in Sections, 2.2.1 through 2.2.4, respectively.

2.2.1 Modeling Environment

Modeling environments involve: (1) the format of data storage; (2) the use of add-in software packages; (3) programming environments; and (4) aggregation and binning.

It is desirable to store data as pairs of inputs and outputs to facilitate the application of sensitivity analysis. The stored format should be compatible or easily convertible to the formats accepted by software packages commonly used for sensitivity analysis. Examples of such formats are comma separated values (.csv) and free text format (.txt).

Storing data as pairs of inputs and outputs may not be an optimal solution, especially for models with an extensive number of inputs and outputs or a large number of iterations during probabilistic simulations. As an alternative storing approach, application of relational databases is recommended. Using a database for storing the results of the analysis can facilitate the process of managing and retrieving the information in the form of outputs and corresponding inputs. A well-defined database can decrease the required storage space and increase the speed of retrieving data. An example of this strategy to store data for performing sensitivity analysis is the Environmental Protection Agency (EPA), Stochastic Human Exposure Dose Simulation, (SHEDS) Pesticide model (Zheng and Frey, 2003).

"Add-in" is a term used, especially by Microsoft, for a software utility or other program that can be added to a primary program. Examples of add-in software packages are Crystal Ball™ or @Risk, which are applicable to Microsoft Excel. The use of add-in software packages helps reduce the amount of programming. However, the add-in software packages typically do not provide access to the code used to perform the various operations, and thus restrict the flexibility of the model with respect to possible modifications for sensitivity analysis. Performing the operations inside spread sheets using add-in software packages makes it accessible to those without knowledge of the programming language, but may make it difficult to audit the model. For complex and large models coded in Excel sheets, it becomes very difficult to understand the flow of the model and connection between different parts and modules inside the model.

The choice of programming environments depends on the skill of the modeler, use of add-ins, and the scope of the analysis. For models that are extensive and that will be used for multiple analyses, a programming language environment and good software engineering practices are recommended. The choice of modeling environment should account for a trade-off, if any, between the skills of the analyst, resources, anticipated needs for future model refinements, and desired flexibility with regard to sensitivity analysis. The codes for most of the operations used in food safety process risk models, such as Monte Carlo simulation, are available as open source codes on the internet that can be downloaded free of charge. Open source codes of other available codes, such as R are also available that could be used for sensitivity analysis. The use of object-oriented or modular codes to perform various operations facilitates modification and enhancement of models. Also, well-documented codes can be easier for a knowledgeable analyst to audit compared to the add-in software.

If there is some rationale for application of binning in a model, it should be provided as an extra functionality. The user should be able to access the actual numerical values of an output before binning without having to modify the model itself. Hence, pre-binned values should be stored. Aggregation results in similar restrictions for application of sensitivity analysis as does binning. If aggregation is part of a real food safety process under study, such as the aggregation of meat trims to combo bins in slaughter plants, then the aggregation may be unavoidable. In some cases, aggregation is done because of lack of information about individual processes and insufficient understanding of what is exactly happening from a biological viewpoint. Hence, different processes in the continuum of bringing food from farm-to-table may be aggregated into a black box. In these cases it may be practical to refrain from aggregation by collecting more detailed data regarding the process under study.

2.2.2 Characterizing Variability and Uncertainty in a Probabilistic Simulation

The choice of probabilistic dimensionality depends on the assessment objective. If the objectives include characterizing inter-individual variability in exposure or risk, or identifying key sources of controllable variability in order to inform risk management decisions, a probabilistic simulation of variability should be included. If the objectives include quantification of uncertainty due to lack of knowledge, evaluation of the robustness of risk management options, or identification of key sources of uncertainty, then a probabilistic simulation of uncertainty should be included. If the assessment objective motivates inclusion of both variability and uncertainty, then a choice can be made between a one-dimensional simulation in which variability and uncertainty are co-mingled versus a two-dimensional simulation in which they are distinguished. The former may be interpreted as treating variability as a source of uncertainty about a representative member of a population, while the latter is typically preferred if the objective includes characterizing exposure or risk to different member of a population. Although the number of dimensions of probability simulation for a food safety risk assessment model depends on specific problems under study, ideally, both variability and uncertainty should be characterized since the explicit separation of variability and uncertainty helps analysts or decision-makers to understand how model outputs might improve.

When both variability and uncertainty dimensions are incorporated into a model, special attention should be paid to the modeling strategies to make the separation of variability and uncertainty practical. When distinguishing between variability and uncertainty, an analyst may be interested in quantifying the range of uncertainty regarding the population distribution of variability and/or key sources of uncertainty for specific fractiles of the variability distribution of the output. The former is easily addressed in a two-dimensional simulation framework in which there are multiple realizations of the distributions of inter-individual variability. The latter case

requires that the set of random numbers associated with the variability dimension of each model input remain constant in multiple realizations of uncertainty. Part III provides further discussion regarding the latter case. An example of a food safety process risk model following this modeling strategy is the *Listeria* model in RTE foods (CFSAN, 2003). In this model, the set of random numbers generated for sampling from variability distributions of inputs are stored. At each uncertainty realization of the model, the model used the stored random numbers for sampling from variability distributions of inputs.

2.2.3 Modeling Strategies

A basic strategy in building a new model is to develop independent modules to facilitate sensitivity analysis. An independent module is one that does not share inputs or internal data with its predecessor or successor modules, and hence, the module is self-sufficient for calculation of its output. This modeling strategy is most easily achieved when using modules in parallel, as shown in Figure 2-2. For modules in series, as shown in Figure 2-1, as long as the one-to-one correspondence can be preserved, it is possible to attribute variation in the output to the inputs of one or more modules.

The main functionalities in a model should be implemented in separate modules. For example, there might be separate modules for random number generation, core model simulation, simulation result summarization and presentation, sensitivity analysis module, and others. An example of the use of such a strategy is the C++ version of the EPA/SHEDS/Pesticide model (Zheng and Frey, 2003). A modular framework can easily incorporate future additions of new functionalities simply by adding new modules without requiring extensive modification of the original model.

2.2.4 Model Documentation

Good documentation of the model structure and implementation is necessary to facilitate future application or modifications of a model. Good documentation includes:

- Introducing the model and its probabilistic framework
- Explaining of the model equations
- Introducing the model inputs and corresponding probability distributions and units
- Introducing the model outputs and corresponding units

It is necessary to introduce a model clearly and in detail, including mathematical equations used to describe the model and how the model was implemented in a computer program. Good documentation will also help the application of sensitivity analysis. For example, it will help an analyst to better understand a model including its inputs and outputs, and to select appropriate methods to perform sensitivity analysis. Good documentation requires that the model structure, the composition of model components, assumptions used, and software structure design implementing the model be clearly presented. For example, each module in a model should be clearly explained regarding its purpose, associated assumptions, its interior functional relationships, and its inputs and outputs. Each input should be clearly described regarding its type (e.g., probability distribution, numerical value or logical decision variables, and quantitative versus qualitative), values, associated variability and/or uncertainty dimensions, and its unit.

2.3 References

CFSAN, 2003. Quantitative assessment of relative risk to public health from Foodborne *Listeria monocytogenes* among selected categories of ready-to-eat foods. Prepared by Center for Food Safety and Nutrition/Food and Drugs Administration, U. S. Department of Agriculture, September 2003.

Frey, H.C., Mokhtari, A., and Tanwir, D., 2003. Evaluation of selected sensitivity analysis methods based upon applications to two food safety process risk models. Prepared by North Carolina State University for Office of Risk Assessment and Cost-Benefit Analysis, U.S. Department of Agriculture. Washington DC. www.ce.ncsu.edu/risk.

FSIS, 2001. Draft risk assessment of the public health impact of Escherichia Coli O157:H7 in ground beef. Prepared for the Food Safety and Inspection Service and United States Department of Agriculture, Washington, DC. www.fsis.usda.gov/OPPDE/rdad/FRPubs/00-023N/00-023NReport.pdf

Nauta, M.J., 2001a. A Modular process risk model structure for quantitative microbiological risk assessment and its application in an exposure assessment of *Bacillus Cereus* in a REPFED. RIVM Report 149106 007, National Institute for Public Health and the Environment, Bilthoven.

Nauta, M.J., 2001b. Modeling bacterial growth in quantitative microbiological risk assessment: Is it possible? International Journal of Food Microbiology, 73(2-3): 297-304.

Patil, S. R. and Frey, H. C., 2003. Comparison of sensitivity analysis methods based upon applications to a food safety risk model. Risk Analysis, Submitted December 19, 2002, accepted July 29, 2003.

Zheng, J., Frey, H. C., 2003. Design and development report of the C++ version of EPA/SHEDS/Pesticide model. Prepared by Department of Civil, Construction and

Environmental Engineering at North Carolina State University for the Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, June 2003.

PART III

RECOMMENDED PRACTICE REGARDING SELECTION OF SENSITIVITY ANALYSIS METHODS APPLIED TO MICROBIAL FOOD SAFETY PROCESS RISK MODELS

Accepted to be Published at:

Human and Ecological Risk Assessment

ABSTRACT

We present a guideline for selection of sensitivity analysis methods applied to microbial food safety process risk (MFSPR) models. The guideline provides useful boundaries and principles for selecting sensitivity analysis methods for MSFPR models. Although the guideline is predicated on a specific branch of risk assessment models related to food-borne diseases, the principles and recommendations provided are typically generally applicable to other types of risk models. Applicable situations include: prioritizing potential critical control points; identifying key sources of variability and uncertainty; and refinement, verification, and validation of a model. Based on objective of the analysis, characteristics of the model under study, amount of detail expected from sensitivity analysis, and characteristics of the sensitivity analysis method, recommendations for selection of sensitivity analysis methods are provided. A decision framework for method selection is introduced. The decision framework can substantially facilitate the process of selecting a sensitivity analysis method for a practitioner.

Key Words: Sensitivity analysis, microbial food safety process risk model, variability, uncertainty

1.0 Introduction

There has been increasing development and use of quantitative models for microbial food safety process risk (MFSPR) assessment (Rose and Sobsey, 1993; Jaykus, 1996). The purpose of MFSPR models is typically to provide insight regarding possible risk management strategies. These models can be large and complex. Thus, it can be difficult to prioritize controllable inputs that are most promising with respect to risk management goals. Sensitivity analysis can be used as a methodology for identifying the key model inputs that contribute the most to variation in a selected model output or to the highest values of the output. An analyst is often interested in identifying which one or subset of potentially many controllable inputs is of greatest importance with regard to human exposure and risk. Other reasons for performing sensitivity analysis include assisting in the process of model development. For example, when a model is developed by a team or in response to suggestion from multiple stakeholders, it is often the case that many features incorporated into the model are not essential. The identification of inputs that are of insignificant sensitivity could be used to guide the elimination of particular inputs or components of the model. Critical evaluation and reduction of the size of the model can help in preventing the model from becoming so large and unwieldy that it is no longer practical. However, sensitivity analysis is not currently used in practice as effectively as it could be because of lack of a comprehensive framework for choosing and applying such methods.

There are several techniques for sensitivity analysis used by practitioners and analysts in numerous fields, including microbial risk assessment. Some may be useful and more rigorous for MFSPR model (Crick *et al.*, 1987; Morris, 1991; Helton, 1993; Saltelli *et al.*, 1999). This paper is the result of a research to transfer, apply, and adapt sensitivity analysis methods developed in other disciplines (e.g. complex engineering systems) to microbial food safety risk assessment. Promising sensitivity analysis methods were identified, reviewed, and evaluated (Frey and Patil,

2002; Frey *et al.*, 2003) and input from experts and practitioners were obtained (Helton and Davis, 2002; Hulebak and Schlosser, 2002; Kohn, 2002; Pate-Cornell, 2002; Thompson, 2002, Saltelli, 2002). The need for guidance to practitioners regarding selection, application, interpretation, reporting, and documentation of sensitivity analysis methods is addressed (Frey *et al.*, 2004). Considering the wide range of sensitivity analysis methods, a practitioner should have a resource via which to more clearly understand which methods are appropriate for a specific application. This paper summarizes the key findings in response to this concern.

This paper is intended to assist the practitioner regarding the following key questions:

- When should I perform sensitivity analysis?
- How can I select among typical sensitivity analysis methods?

2.0 When to Perform Sensitivity Analysis?

Sensitivity analysis of a MFSPR model is useful for: (1) prioritization of potential critical control points; (2) identification of key sources of uncertainty and variability; and (3) model refinement, verification, and validation.

2.1 Prioritization of Potential Critical Control Points

A critical control point (CCP) is a point, step, or procedure at which control can be applied, and a food safety hazard can be prevented, eliminated, or reduced to an acceptable level (Hulebak and Schlosser, 2002; Seward, 2000). Examples of CCPs include cooking, chilling, prevention of cross contamination, and product formulation controls.

Sensitivity analysis helps prioritize the possible CCPs that are incorporated into the model. Furthermore, sensitivity analysis can provide insight regarding critical limits for a particular CCP in order to develop preventive measures. A critical limit is a criterion that must be met for each preventive measure associated with a CCP (e.g., the maximum allowed food

storage temperature). In the case where the CCP is a measurable variable in food production, handling, or preparation, insight regarding the critical limit is of direct risk management relevance.

2.2 Identification of Key Sources of Variability and Uncertainty

Uncertainty represents the lack of perfect knowledge of an input and can be reduced by further measurements. Variability represents a true heterogeneity in the population, and cannot be reduced by further measurement (Murphy, 1998; Anderson and Hattis, 1999; Cullen and Frey, 1999). Variability is also interpreted to refer to the certainty that different individuals have different exposures and different risks (NRC, 1994).

In order to prioritize data collection activities, it is useful to prioritize the key sources of uncertainty and variability. In many cases, the variability in exposures is typically influenced by only a subset of the model inputs that are subject to variability. Similarly, the uncertainty in a selected model output may be influenced by only a subset of the model inputs that are subject to uncertainty. It would be a poor allocation of scarce resources to spend an equal amount of time developing probability distributions for all model inputs, if the output is sensitive to only a small number of inputs. Sensitivity analysis can be applied to a probabilistic risk assessment model to provide insight regarding which model inputs contribute the most to uncertainty, variability, or both, for a particular model output. This insight can then be used to allocate scarce resources preferentially to data collection or research for those inputs that matter the most to the assessment.

2.3 Model Refinement, Verification, and Validation

During the process of developing or refining a model, sensitivity analysis can be applied for verification and validation of model-based analysis (Saltelli, 2002).

Verification is a process of checking whether a model is implemented as it was intended to be. Sensitivity analysis can be used to assess whether the model output responds appropriately to a change in model inputs. For example, if storage temperature increases, one typically expects an increase in the growth of microbial pathogens and, therefore, in exposure. If a model responds in an unexpected way to changes in one or more inputs, then trouble-shooting efforts can be focused to identify and correct the source of the problem.

Sensitivity analysis can assist in the validation process. Validation typically includes comparison of model predictions to independent empirical data for the model output under a known set of conditions. When validating a model, it is typically necessary to specify input values to reproduce a particular scenario for which real-world data are available for the model output. However, in many cases, the corresponding real world values are not known for all of the model inputs. Thus, judgments often must be made regarding what values to assign to some model inputs for which empirical data are not readily available. In order to determine how accurately a particular input must be estimated, it is useful to understand how substantially the model output responds to it.

The overall assessment objectives should be clearly laid out before the model is built. Once the objectives are defined, the model must be designed to address these objectives. For example, if the objective is to develop insight into possible risk management strategies, variables that can be controlled should be incorporated into the model. The answers sought from application of sensitivity analysis should be clearly listed. The usefulness of sensitivity analysis can then be assessed based on whether the available methods can address the questions under consideration in a manner that is appropriate to the characteristics of the model.

The choice of an appropriate sensitivity analysis method should depend on the objectives, model characteristics, amount of detail expected from sensitivity analysis, and characteristics of the method. In general, methods that are not sensitive to specific model characteristics, such as non-linearity, interaction, thresholds, and categorical inputs, are likely to be more robust. Criteria for selecting an appropriate sensitivity analysis method are incorporated into a decision framework presented here. When the framework suggests more than one applicable method, it is recommended that more than one method be applied and the results compared, if resources allow, in order to obtain more robust insight with respect to sensitivity.

3.0 Selection of Sensitivity Analysis Methods

Sensitivity analysis methods are typically categorized based on their scope, applicability, or characteristics. For example, Frey and Patil (2002) broadly classify sensitivity analysis methods with regard to characteristics as mathematical, statistical, and graphical methods. Alternatively, Saltelli *et al.* (2000) classify sensitivity analysis methods with respect to the application and scope as screening, local, and global. The former approach is used here.

Examples of mathematical methods include nominal range sensitivity analysis (NRSA) (Cullen and Frey, 1999) and differential sensitivity analysis (DSA) (Hwang *et al*., 1997; Grievank, 2000). Examples of statistical sensitivity analysis methods include sample (Pearson) and rank (Spearman) correlation analysis (Edwards, 1976; Gibbons 1985), sample and rank regression analysis (Draper and Smith, 1981; Neter *et al.*, 1996), analysis of variance (ANOVA) (Neter et. al., 1996; Steel et. al., 1997), classification and regression tree (CART) (Breiman *et al.*, 1984), response surface method (RSM) (Khuri and Cornell, 1987; Myers and Montgomery, 1995), Fourier amplitude sensitivity test (FAST) (Saltelli *et al.*, 2000), mutual information index (MII) (Jelinek, 1970; Merz *et al.*, 1992), and Sobol's indices (Sobol, 1993, Saltelli *et al*., 2000).

Examples of graphical sensitivity analysis methods include scatter plots (Kleijnen and Helton, 1999) and conditional sensitivity analysis (Frey *et al.,* 2003). These methods are used in the decision framework introduced in this section. Further discussion of these methods are provided in Frey and Patil (2002) and Frey *et al.* (2003, 2004).

3.1 Key Questions for Selection of Sensitivity Analysis Methods

The selection of a sensitivity analysis method appropriate for a particular case study depends on the characteristics of the model and the case study. This section discusses a series of six key questions that the practitioner should consider before choosing a method. These questions should be considered a starting point for selecting a method. The analyst should determine whether there are additional considerations in a particular case study beyond the issues addressed here.

3.1.1 What are the Objectives of Sensitivity Analysis?

Knowledge of the objectives helps an analyst efficiently allocate the available time and resources to select an appropriate sensitivity analysis method that can provide relevant information and insight. Some common objectives of sensitivity analysis are to: (1) rank the importance of model inputs (e.g., critical control points); (2) identify combinations of input values that contribute to high exposure and/or risk scenarios; (3) identify and prioritize key sources of variability and uncertainty; (4) identify critical limits; and (5) evaluate the validity of the model. Some of these objectives, such as (3) and (5), might primarily be research issues, while others are typically associated with risk management or regulatory decision-making.

Rank ordering the model inputs based on their contribution to the variation in the output of interest is a typically employed approach. An analyst can provide recommendations for model refinement through further data collection to improve model input distributions using the insight

from rank ordering the inputs. Mathematical and statistical sensitivity analysis methods in general provide quantitative ranking of model inputs (Frey *et al.,* 2003). Graphical sensitivity analysis techniques typically do not provide quantitative rankings of inputs.

An analyst may be interested in identifying ranges of key inputs that lead to high exposure and/or risk in MFSPR models. For example, decision makers and regulatory agencies may be interested in identifying conditions responsible for high contamination levels in the process of bringing foods from the farm to the table. Some statistical methods, such as CART and ANOVA, can be used to identify combinations of inputs that lead to high exposures and risk (Frey *et al.*, 2003).

Objectives based upon identification and prioritization of key sources of variability and uncertainty in a probabilistic model are typically fulfilled using statistically-based sensitivity analysis methods. In addition, the analyst typically prefers that the sensitivity analysis method should provide quantitative measures of sensitivity for ranking key sources of variability and uncertainty. Depending upon the assessment objectives with regard to whether and how variability and uncertainty should be distinguished, sensitivity analysis can be applied in a one or two-dimensional probabilistic simulation framework. For example, if the assessment objective includes distinguishing between inter-individual variability in exposure versus uncertainty, then a two-dimensional probabilistic simulation framework would typically be employed. Ideally, in such situations the process of sensitivity analysis should be automated because of the potentially large number of model iterations required.

If the objective is to identify critical limits for controllable inputs, it is often necessary to use relatively sophisticated sensitivity analysis methods. This is particularly the case if the model has substantial nonlinearity, interactions, or thresholds. These methods often involve a

probabilistic simulation (e.g., Monte Carlo simulation). Critical limits are sometimes associated with saturation points or thresholds in the input domain. Sensitivity analysis methods based upon linearity assumptions are not well-suited for identification of these kinds of limits. However, some statistical (e.g., ANOVA, CART) and graphical methods (e.g., scatter plots) can provide useful insights in these situations (Frey *et al.*, 2003).

If the objective is to evaluate the validity of the model, then the analyst is typically interested in: (1) qualitative characteristics of how the output responds to changes in one or more inputs; and/or (2) how the output varies quantitatively as a result of varying inputs. Qualitative characteristics can include: (a) does the output increase or decrease if an input increases?; (b) does the output vary in a linear, monotonic, or non-monotonic manner?; (c) are there threshold or saturation effects, or other characteristics of specific interest (e.g., inflection points, singularity points)?; and (d) are there interactions between two or more inputs that vary simultaneously? With regard to each of these characteristics, or perhaps others, the analyst will typically assess whether the model output is responding in an appropriate or explainable manner. If so, then confidence in the model tends to increase. If not, then corrective measures with respect to the model structure and assumptions are indicated. Sources of errors in a model response could include incorrect specification of the inputs or a problem with the analytical model or the computer code.

3.1.2 Based Upon the Objectives What Information is Needed from Sensitivity Analysis?

The analyst should decide prior to analysis what information is of most interest. Examples of information that could be useful, depending upon the objectives, include: (1) qualitative or quantitative ranking of inputs; (2) discrimination of the importance among different inputs; (3) grouping of inputs that are of comparable importance; (4) identification of

inputs that are not important; (5) identification of critical limits; (6) identification of inputs and ranges that produce high exposure or risk.

The analyst will typically want to be able to distinguish whether an input is of high, moderate, low, or no importance with regard to a particular output. Therefore, it is often useful to be able to discriminate regarding whether one input is more important than another input, even if the output is sensitive to both. Some statistical methods provide measures of whether one input is significantly more sensitive than another input. Several inputs may have comparable importance with regard to a particular output. The analyst may wish to group inputs that are of comparable importance based upon similar sensitivity index values. Similarly, the analyst may wish to identify inputs that are not important to the assessment. Information regarding unimportant inputs is useful because it is typically not necessary to devote additional resources to refining the assumptions for such inputs. Such resources can be prioritized elsewhere.

In order to identify critical limits in a model, the selected sensitivity analysis method should have the capability of comparing the model responses for different intervals for one or more model inputs. Similarly, if the objective is to identify high exposures, the selected sensitivity analysis method should identify the inputs and their ranges that are responsible for such exposures. Methods such as CART or ANOVA have capabilities to provide these kinds of insights.

3.1.3 What are the Characteristics of the Model that Constrain or Indicate Preference Regarding Method Selection?

MFSPR models usually have specific characteristics that may constrain the application of sensitivity analysis methods. These characteristics include: (1) nonlinearities; (2) interactions; (3) thresholds and saturation points; and (4) categorical inputs (Frey, 2002).

An ideal sensitivity analysis method should be model independent. Specifically, a sensitivity analysis method should not require any assumptions regarding the functional form of the risk model and should be applicable to different model formulations. Some methods are considered to be global and model-independent. An example is CART. However, some methods, including many commonly used methods, are based upon assumptions regarding the functional form of the model. For example, sample correlation analysis or linear regression is based upon the assumption of a linear model. In some cases, a linearizing transformation could be made, such as a log transformation, in order to have a more nearly linear model response to changes in inputs. Thus, with appropriate transformations, sometimes it is possible to apply these techniques. Rank correlation coefficients or rank regression are based upon the assumption of a monotonic model. If a sensitivity analysis method based upon an assumed functional form of a model is applied to a model with different characteristics, then the results of the sensitivity analysis may not be valid.

Because interaction between inputs is often one of the common characteristics of MFSPR models, an analyst should select a sensitivity analysis method that can deal with interactions. The most commonly used mathematical sensitivity analysis methods (e.g., NRSA, DSA) are not able to capture interaction effects between model inputs. However, many statistical methods, such as ANOVA, CART, Sobol, or appropriately specified regression approaches, can identify interactions between model inputs. Graphical methods may provide insights regarding interaction effects; however, the use of graphical methods typically requires expert judgment based upon visual inference and therefore is a qualitative approach to sensitivity analysis.

Identification of thresholds and saturation points that have a substantial effect on the model output (e.g., exposure or risk) is useful because they may imply critical limits. Sensitivity

analysis methods that enable an analyst to compare variation of the model output in different regions of an input domain are appropriate for identification of these model characteristics. Such methods include ANOVA, CART and some graphical sensitivity analysis methods. ANOVA provides direct statistical measures for tracking thresholds and saturation points. Graphical sensitivity analysis methods provide visual insight regarding possible thresholds or saturation points in an input domain. For example, scatter plots can assist in identification of thresholds and saturation points taking into account simultaneous variation of other inputs. Graphical analysis based upon conditional sensitivity analysis can also help identify thresholds and saturation points for an input conditional on point values from other inputs.

Categorical or qualitative inputs are commonly used in MFSPR models. Examples include food categories, season, consumer age, eating location, and consumption type. Mathematical methods such as NRSA and DSA do not accommodate these types of inputs. However, an analyst could simply run the model with different values of such an input while other inputs are held constant and assess the range of variation in the model output compared to that obtained when each of several other inputs are varied individually. Among statistical sensitivity analysis methods, ANOVA and CART can deal with categorical inputs. Regression analysis can deal with categorical inputs by using dummy variables (Neter *et. al.,*1996). Graphical sensitivity analysis methods can be used to visualize the variation of model output versus different categorical values of inputs.

3.1.4 How Detailed is the Analysis?

Screening or refined sensitivity analyses can be conducted based upon the objectives of an analysis. The choice between the two is typically governed by resource availability, the importance of the analysis, and the stage of the assessment. Screening analyses require fewer

resources than refined analyses. A screening analysis might be used in the early stages of model development to help refine the model and its inputs and to assess model validity. Typical screening analysis methods include local mathematical methods such as DSA or NRSA, as well as sample or rank correlation coefficients available in commercial statistical software packages. A refined analysis might typically be used in the later stages of analyses with a model that has undergone previous screening analyses. A refined analysis is preferred if the results will be used to make decisions regarding commitments of large amounts of resources for further model development, data collection for model inputs, or risk management strategies. However, the choice of a refined analysis should account for a trade-off, if any, between the skills of the analyst, resources, anticipated needs for custom programming, and potentially higher required time with regard to sensitivity analysis.

Screening and refined analysis can be used together. The time and effort to execute a refined analysis often depends on the number of inputs that are included in the analysis. Therefore, it is often useful to use a screening method to identify model inputs that are not important with regard to variation in the output of interest. The refined analysis can then be applied to a smaller set of inputs for which there is reason to believe that the model output has at least some sensitivity.

When applying a screening analysis method, the analyst should keep in mind that there are several potentially critical limitations of such methods, including: (1) screening methods are often not model-independent, and therefore the insights may be inaccurate if the model structure deviates significantly from the theoretical basis of the method; (2) some screening methods are local, and cannot adequately account for simultaneous variation in multiple inputs; and (3) for local screening methods, it may not be possible to identify the inputs and ranges that lead to the

highest risk or to gain useful insight regarding critical control points. Thus, it is preferable to confirm results from a screening method with a different method or to follow-up with a refined analysis.

3.1.5 Can Commonly Used Methods Adequately Address the Characteristics of Interest in the Analysis?

Under pressure of deadlines and faced with constraints of an existing model, an analyst may prefer to use readily available features of existing software tools rather than attempt to implement more refined techniques. Such features typically include sample (Pearson) and rank (Spearman) correlation coefficients applied in the context of a one-dimensional probabilistic simulation. These methods are applicable to global sensitivity of linear or monotonic models, respectively. Sample correlations are not a robust method for dealing with thresholds or interactions and cannot provide useful insights when applied to categorical inputs. Rank correlations are more robust than sample correlations with regard to thresholds but are not capable of providing insight regarding the existence of a threshold or the value of the threshold. If an analyst chooses to proceed with an available method whose theoretical basis differs from the key characteristic of the model to which it is applied, then the accuracy and robustness of the results cannot be assured.

3.1.6 Is the Implementation of the Selected Sensitivity Analysis Method Post Hoc?

A sensitivity analysis method is referred to as post hoc if it is applied to previously prepared results from probabilistic or deterministic simulation of a model but is not included as a component of the model itself. Therefore, the application of a post hoc sensitivity analysis method does not contribute to the process of model simulation, but it may impose requirements regarding the type and format of data that should be stored from the simulation. For example, regression analysis can be applied to a dataset of randomly generated values of the inputs and

corresponding output values. In these cases, it is possible for the sensitivity analysis to be performed in a different software environment and by different personnel compared to the simulation modeling. Examples of methods that can be applied post hoc include sample and rank correlations, regression analysis, ANOVA, CART, and graphical techniques.

In contrast, some methods require a different simulation strategy and therefore may have to be programmed or implemented manually or interactively with the model. Some of this will depend upon how the model is structured. For instance, among statistical methods for sensitivity analysis, Sobol's method, Fourier amplitude sensitivity test (FAST), and mutual information index (MII) typically are not post hoc. These methods require approaches to model simulation different than the conventional Monte Carlo methods and therefore must be incorporated into the simulation process. The application of these kinds of methods typically requires more advance planning and coding than do the post hoc methods. Although these are powerful methods that offer advantages over push-button methods, their widespread practical application is limited until software becomes available by which these methods can be easily incorporated into a risk model.

3.2 Decision Frameworks to Assist in Selecting Sensitivity Analysis Methods

Figure 1 shows that the first step in selecting an appropriate sensitivity analysis method is to decide on the level of detail expected from sensitivity analysis. This figure presents two levels of sensitivity analysis: (1) screening analysis; and (2) refined analysis. For screening analysis, the choice of sensitivity analysis method depends on the simulation approach of a model. If a deterministic approach is selected for screening analysis, particularly for local sensitivity analysis, then mathematical sensitivity analysis methods are recommended. For probabilistic approaches, push-button techniques including sample and rank correlation coefficients are listed as appropriate methods. In contrast, if a practitioner decides to perform a refined analysis, the

choice of a method depends on the objective of sensitivity analysis. Three objectives are listed in the decision framework for refined analysis: (1) model refinement; (2) identifying key sources of variability and uncertainty; and (3) identifying high exposure scenarios. For the latter objective, two classes of methods are introduced. Class *A* introduces methods that provide explicit measures for addressing high exposure scenarios. CART is the only method introduced in this class capable of addressing high exposure scenarios directly (Frey ***et. al.,*** 2003; Mishra ***et al.,*** 2003). Methods introduced in Class *B* (including ANOVA, conditional sensitivity analysis, and scatter plots) require judgment of an analyst for interpretation of the results and identification of inputs responsible for high exposure scenarios. The decision framework for method selection considering the first two objectives is illustrated in Figure 2.

When the objective of sensitivity analysis is either model refinement or identifying key sources of variability and uncertainty, the choice of an appropriate method for sensitivity analysis depends on the characteristics of the model under study. Figure 2 considers four characteristics for a food safety model: (1) non-linearity; (2) interaction; (3) categorical inputs; and (4) threshold and saturation points.

For non-linear models, the choice of a method first depends on whether or not the model is monotonic. The decision framework is further classified based upon the condition of whether a sensitivity analysis method is post hoc or integrated with software. For models with interactions, the selection of method depends on whether the implementation of the selected method is post hoc. Post hoc methods are classified as of Class *C* or *D.* Methods in Class *C* directly take into account interaction between model inputs, while the results of methods in Class *D* involve analyst judgment in order to address interaction between inputs. Methods appropriate for models with categorical inputs are classified based upon post hoc or integrated implementation. When

the model under study has thresholds or saturation points, sensitivity analysis methods are classified based on whether they explicitly address such characteristics. Methods in class *C* include ANOVA and scatter plots (SP), which explicitly handle possible thresholds or saturation points in a model, while methods in Class *D* include CART and conditional sensitivity analysis, which require analyst judgment for quantifying available thresholds and saturation points.

As a summary, Table 1 provides key characteristics addressed by selected sensitivity analysis methods. These methods were evaluated with respect to multiple characteristics based upon case studies performed by Frey and Patil (2002) and Frey *et al* (2003). Ideally, a sensitivity analysis method should respond to the effects of simultaneous variation in all inputs. The methods were evaluated to determine if they address nonlinearities in response to an input. The identification of the presence or absence of threshold in the model response was evaluated. The ability of sensitivity analysis methods to identify and provide detailed insights regarding the existence of interaction among inputs is considered to be critical for MFSPR models. Some methods handle only quantitative inputs and others can address both quantitative and categorical inputs.

Because high exposure cases are often of special interest, methods that can help identify and characterize conditions leading to high exposures may be preferable. The ability to address uncertainty about importance of inputs via two-dimensional analysis can provide an indication of robustness and confidence in the efficiency of an identified control measure.

Some methods are easier to apply in practice than others. The ease of application may often constrain the feasibility of a method. A method is typically easier to implement when software tools for implementing the method already exist, especially if they have user-friendly

interfaces. Of course, ease of implementation will be a function of software availability and programming skill level.

The ability to produce quantitative rankings and the ability to evaluate the statistical significance of the rankings are useful to identify the relative importance of inputs and the confidence that should be imputed to the rankings. Some methods produce more useful measures by which to discriminate the importance among similarly ranked inputs. Finally, although each method has a different theoretical basis, the bottom line for practical use of the methods is whether they produce reasonable results even if there are departures from key assumptions of the method.

4.0 Acknowledgment

The work reported here is based on work supported by Cooperative Agreement No. 58-0111-0-005 between the U.S. Department of Agriculture, Office of the Chief Economist, Office of the Risk Assessment and Cost Benefit Analysis (USDA/OCE/ORACBA) and the Department of Civil, Construction, and Environmental Engineering at North Carolina State University. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Agriculture.

5.0 References

Anderson EL, and Hattis DH. 1999. When and how can you specify a probability distribution when you don't know much? Risk Anal 19(1): 43-68.

Breiman L, Friedman JH, Olshen RA, *et al*. 1984. Classification and Regression Trees. Wadsworth International Group, Belmont, CA, USA.

Crick MJ, Hill MD, and Charles D. 1987. The role of sensitivity analysis in assessing uncertainty. Proceedings, NEA Workshop on Uncertainty Analysis for Performance Assessment of Radioactive Waste Disposal Systems, OECD, Paris.

Cullen AC, and Frey HC. 1999. Probabilistic Techniques in Exposure Assessment. Plenum Press, New York, USA.

Draper NR, and Smith H. 1981. Applied Regression Analysis. Second Edition. John Wiley and Sons, New York, USA.

Edwards AL. 1976. An Introduction to Linear Regression and Correlation. W. H. Freeman, San Francisco, CA, USA.

Frey HC, Mokhtari A, and Tanwir D. 2003. Evaluation of Selected Sensitivity Analysis Methods Based Upon Applications to Two Food Safety Process Risk Models. Prepared by North Carolina State University for Office of Risk Assessment and Cost-Benefit Analysis, U.S. Department of Agriculture. Washington DC. Available at: http://www.ncsu.edu/risk.

Frey HC, Mokhtari A, and Zheng J. 2004. Recommended Practice Regarding Selection, Application, and Interpretation of Sensitivity Analysis Methods Applied to Food Safety Process Risk Models. Prepared by North Carolina State University for Office of Risk Assessment and Cost-Benefit Analysis, U.S. Department of Agriculture. Washington DC. Available at: http://www.ncsu.edu/risk.

Frey HC, and Patil R. 2002. Identification and review of sensitivity analysis methods. Risk Anal 22(3): 553-577.

Gibbons JD. 1985. Nonparametric Statistical Inference, 2nd Edition. Marcel Dekker Inc, New York, USA.

Grievank, A. 2000. Principles and Techniques of Algorithmic Differentiation: Evaluating Derivatives. SIAM Publisher, Philadelphia, PA, USA.

Helton JC. 1993. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. Reliab Eng Syst Safe 42(2-3): 327-367.

Helton JC, and Davis FJ. 2002. Illustration of sampling-based methods for sensitivity analysis. Risk Anal 22(3): 591-622.

Hulebak KL, and Schlosser W. 2002. Hazard analysis and critical control point (HACCP) history and conceptual overview. Risk Anal 22(3): 553-578.

Hwang D, Byun DW, and Odman MT. 1997. An automatic differentiation technique for sensitivity analysis of numerical advection schemes in air quality models. Atmos Environ 31(6): 879-888.

Jaykus LA. 1996. The application of quantitative risk assessment to microbial food safety risks. Crit Rev Microbiol, 22:279-93.

Jelinek F. 1970. Probabilistic Information Theory. McGraw Hill Book Company, New York, USA.

Khuri AJ, and Cornell JA. 1987. Response Surfaces. Marcel Dekker Inc, New York, USA.

Kleijnen JPC, and Helton JC. 1999. Statistical analyses of scatterplots to identify important factors in large-scale simulations, 1: Review and comparison of techniques. Reliab Eng Syst Safe 65(2): 147-185.

Kohn MC. 2002. Use of sensitivity analysis to assess reliability of metabolic and physiological models. Risk Anal 22(3): 623-631.

Merz JF, Small MJ, and Fischbeck PS. 1992. Measuring decision sensitivity: a combined Monte Carlo-Logistic regression approach. Med Decis Making 12(3): 189-196.

Mishra S, Deeds NE, and RamaRao BS. 2003. Application of classification trees in the sensitivity analysis of probabilistic model results. Reliab Eng Syst Safe 79(2): 123-129.

Morris MD. 1991. Factorial sampling plans for preliminary computational experiments. Technometrics 33(2): 161-174.

Murphy BL. 1998. Dealing with uncertainty in risk assessment. Hum Ecol Risk Assess 4(3): 685-699.

Myers RH, and Montgomery DC. 1995. Response Surface Methodology: Process and Product Optimization Using Designed Experiments. Wiley and Sons Ltd, New York, USA.

Neter J, Kutner MH, Nachtsheim CJ, *et al.* 1996. Applied Linear Statistical Models, Fourth Edition. McGraw-Hill, Chicago, IL, USA.

NRC (National Resource Council) 1994. Science and Judgment in Risk Assessment. National Academy Press. Washington, DC, USA.

Pate-Cornell E. 2002. Risk and uncertainty analysis in government safety decisions. Risk Anal 22(3): 633-646.

Rose JB, and Sobsey MD. 1993. Quantitative risk assessment for viral contamination of shellfish and coastal waters. J Food Protect 56(12): 1043-1050.

Saltelli A, Tarantola S, and Chan K. 1999. A quantitative, model independent method for global sensitivity analysis of model output. Technometrics 41(1): 39-56.

Saltelli A, Chan K, and Scott, M. 2000. Sensitivity Analysis: Probability and Statistics Series. John Wiley & Sons, New York, USA.

Saltelli A. 2002. Sensitivity analysis for importance assessment. Risk Anal 22(3): 579-590.

Seward S. 2000. Application of HACCP in food service. Irish J Agr Food Res 39(2): 221-227.

Sobol IM. 1993. Sensitivity analysis for nonlinear mathematical models. Math Modeling Com Exp 1(4): 407-414.

Steel RG, Torrie JH, and Dickey DA. 1997. Principals and Procedures of Statistics; A Biometric Approach. 3rd Edition. WCB McGraw-Hill, Boston, MA, USA.

Thompson KM. 2002. Variability and uncertainty meet risk management and risk communication. Risk Anal 22(3): 647-654.

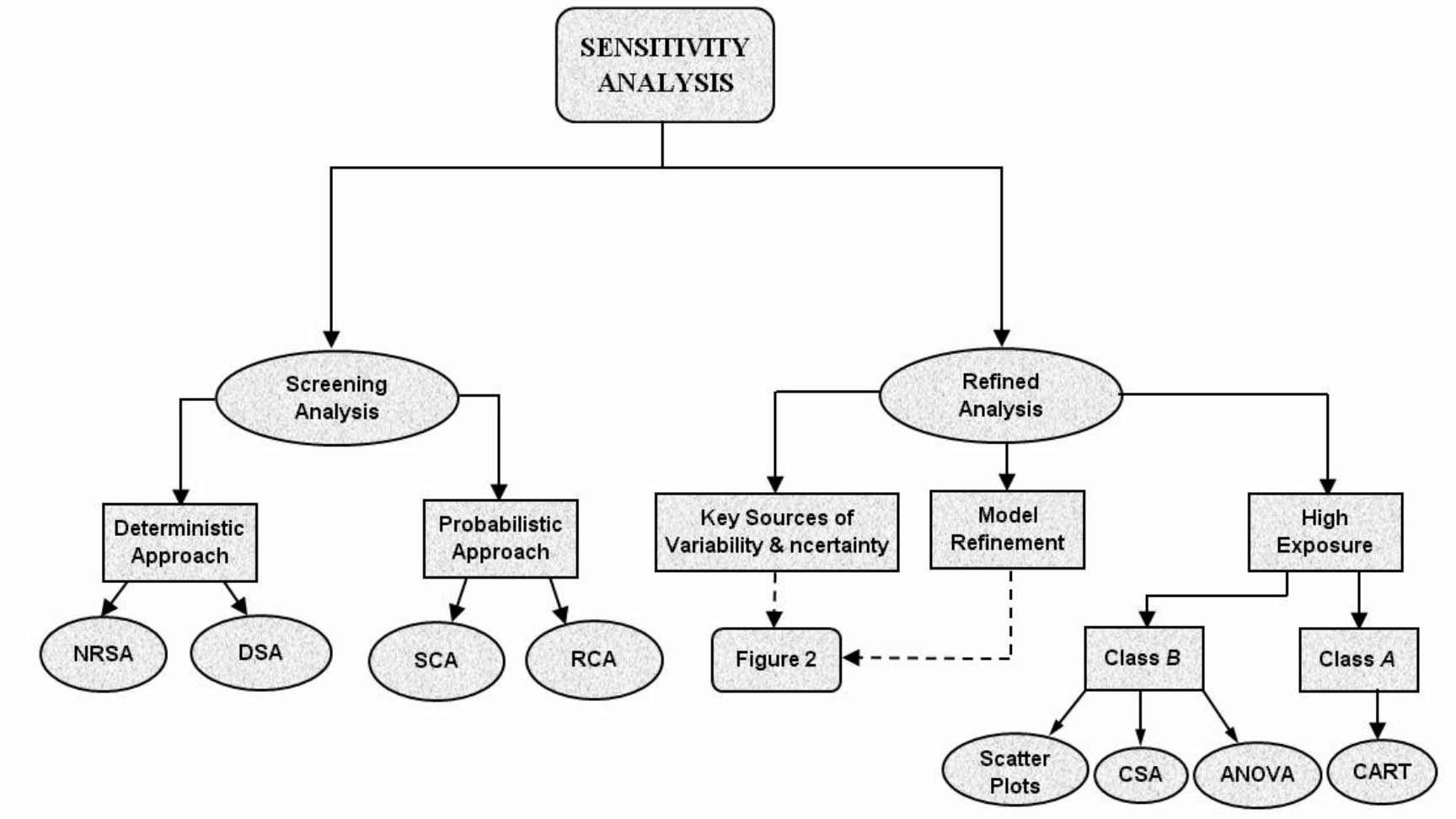


Figure 1. Decision Framework for Selecting an Appropriate Sensitivity Analysis Method.

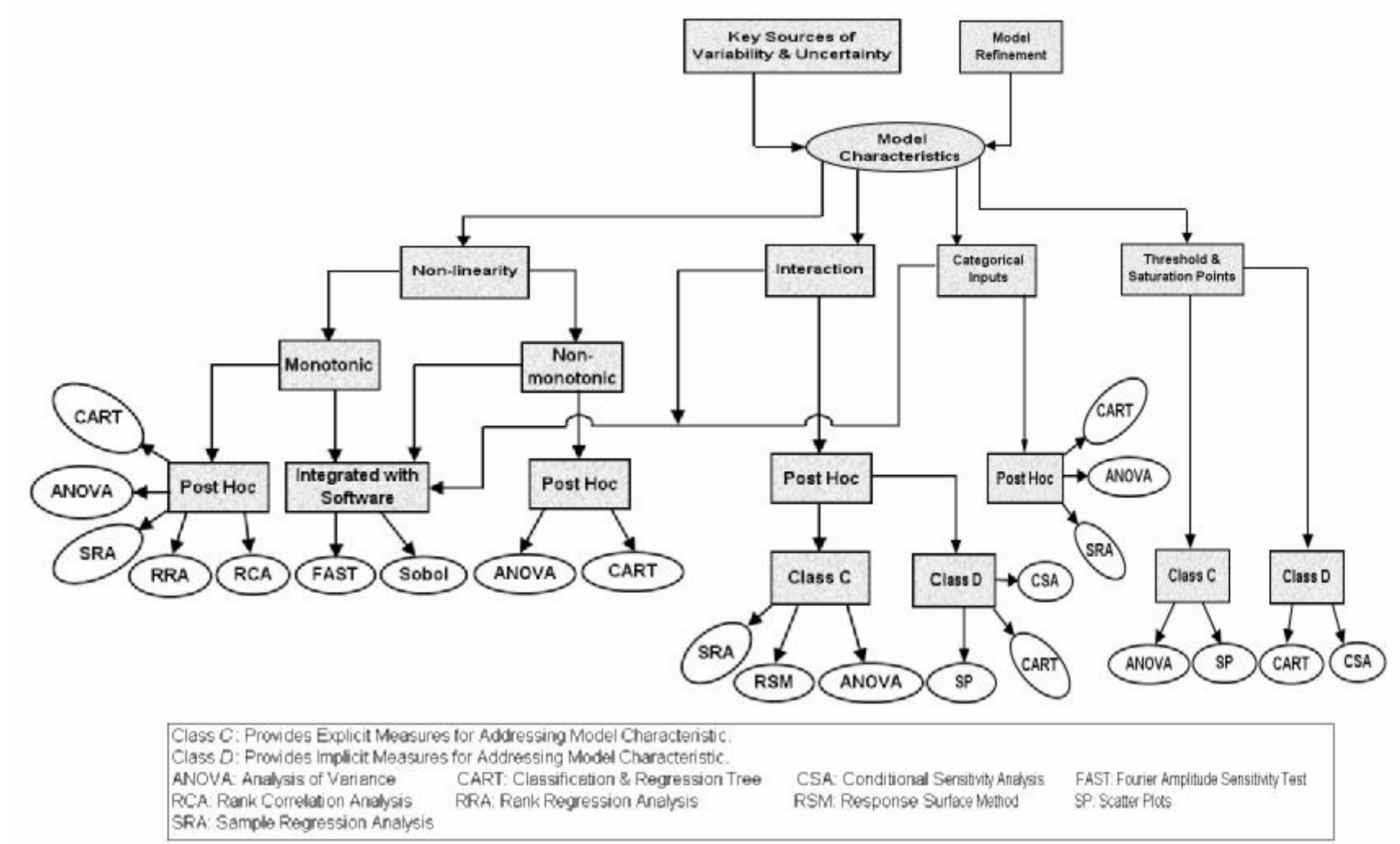


Figure 2. Decision Framework for Selecting Appropriate Sensitivity Analysis Method for Identifying Key Sources of Variability and Uncertainty and Model refinement as Key Objectives of the Analysis.

Table 1. Summary of Key Characteristics of Selected Sensitivity Analysis Methods

Characteristic	Sensitivity Analysis Method													
	NRSA	DSA	Correlation		Regression		ANOVA	CART	FAST	Sobol	RSM	MII	SP	CSA
			Sample	Rank	Linear	Rank								
Simultaneous Variation	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Non-linearity	No	No	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Threshold	No	No	No	No	No	No	Yes [c]	Yes	No	No	No	No	Yes	Yes
Interaction	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Qualitative vs. Quantitative inputs	No	No	No	No	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
High Exposure	No	No	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
Two-Dimensional Analysis	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No
Ease of Implementation	No	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	Yes	No
Quantitative Ranking of Inputs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No[a]	Yes	Yes	Yes	Yes	No	No
Measure of Statistical Significance	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Discrimination of Important Inputs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No [b]	No
Robust in Practice	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

[a] A method for ranking the input based upon the contribution of each input to reduction in total deviance was explored and is promising for future development

[b] Can be based upon expert judgment

[c] Depends on proper definition of factor levels

NRSA: Nominal Sensitivity Analysis DSA: Differential Sensitivity Analysis ANOVA: Analysis of Variance CART: Classification & Regression Tree
CSA: Conditional Sensitivity Analysis RSM: Response Surface Method MII: Mutual Information Index SP: Scatter Plots
FAST: Fourier Amplitude Sensitivity Test

PART IV

EVALUATION OF CLASSIFICATION AND REGRESSION TREES FOR SENSITIVITY ANALYSIS: CASE STUDY FOR THE *E. COLI* FOOD SAFETY PROCESS RISK MODEL

Submitted to:

International Journal of Food Microbiology

Abstract

Classification and Regression Trees (CART) is evaluated for sensitivity analysis of microbial food safety process risk (MFSPR) models. Sensitivity analysis is a tool for: quantifying the impact of changes in input values on model outputs; evaluating how the variation in the output of a model can be apportioned, qualitatively or quantitatively, among model inputs; and identifying factors contributing to particular outcomes (i.e., worst, best) of interest. CART is evaluated on the basis of its ability to address typical characteristics of MFSPR models such as nonlinearities, interaction, threshold and saturation points, and categorical inputs. Moreover, this method is evaluated with respect to identification of high exposure scenarios and corresponding key inputs and critical limits. Application of this method is illustrated with examples from the cooking effect part and the slaughter module of an *E. coli* food safety process risk model. Two approaches are presented for ranking sensitive inputs: (1) visualization of the tree; and (2) contribution of selected inputs to the reduction in the total impurity (e.g., deviance) of simulation results. These approaches typically produce similar but not always identical rankings. For example, an input that appears in the first split in a regression tree does not necessarily have the greatest cumulative contribution to reduction in deviance. The ability of CART to address interactions and thresholds was evaluated. The ability to identify key inputs and critical values of such inputs that produce especially high (or low) values of an output was illustrated. The implications of sensitivity analysis based on CART with respect to development of risk management alternatives are discussed.

Key Words: Sensitivity Analysis, Classification and Regression Trees, Interaction, Threshold, High Exposure, Microbial Food Safety Process Risk Models

1.0 Introduction

The objective of this paper is to evaluate classification and regression trees (CART) as a sensitivity analysis method for application to microbial food safety process risk models (MFSPR). CART enables an analyst to identify key inputs or interactions between inputs and the corresponding ranges associated with particular output values (Breiman *et al.,* 1984). This method also can identify ranges of key inputs that are associated with high exposure and/or risk. Mishra *et al.* (2003) used CART for sensitivity analysis of a probabilistic model for a total system performance assessment (TSPA) of the Yucca Mountain repository system in Nevada, demonstrating the capabilities of CART to handle categorical data and account for input interactions. However, CART has not been demonstrated in applications to MFSPRs. MFSPR models, sensitivity analysis, available methods for sensitivity analysis, and the rationale for focusing on CART are briefly discussed. Several key motivating questions are presented. The answers to these questions are developed based upon case studies of the application of CART to the output of a probabilistic MFSPR model.

Quantitative risk assessment is an emerging tool in the field of microbial food safety (Rose *et al.,* 1993; Jaykus, 1996). In particular, there has been increasing development and use of such models for food safety process risk assessment (Nauta, 2000). Examples include models for *Salmonella* (SERAT, 1998), *Vibrio parahaemolytics* (CFSAN, 2001), *E. coli O157:H7* (FSIS, 2001), and *Listeria monocytogenes* (CFSAN, 2003). These models typically include hundreds of parameters and inputs that have uncertainty and/or variability.

MFSPR models typically are characterized by nonlinearities, interactions between inputs, thresholds or saturation points, and use of both continuous and categorical inputs (Frey, 2002). Non-linearity is a relationship between two variables in which the change in one variable is not

simply proportional to the change in the other variable. A threshold is a value in an input domain below which a model output does not respond to changes in the input. Similarly, a saturation point is an input value above which the model output does not respond. Interaction is a case in which the effect of an input depends on the value of another input. Categorical inputs, such as gender and season, have values that are mutually exclusive.

Due to the typical characteristics of these models, it is often difficult to get reliable insight regarding which model inputs contribute the most to uncertainty, variability, or both, for a particular model output using a simple evaluation of model results. Uncertainty represents lack of perfect knowledge regarding an input value or distribution. Uncertainty can be reduced by further measurements or information. Variability represents true heterogeneity in a population, and cannot be reduced by further measurement (Murphy, 1998; Anderson and Hattis, 1999; Cullen and Frey, 1999). Some sources of variability are controllable (e.g., refrigeration temperatures during food storage) whereas others are inherent properties that are referred to as stochastic variability (e.g., differing susceptibility among members of an exposed human population). Knowledge of key sources of uncertainty is useful in prioritizing additional data collection or research. Knowledge of key sources of variability is useful in identifying control measures or prioritizing information gathering to produce more accurate estimates.

Sensitivity analysis is used to quantify the impact of changes in input values on model outputs (Cullen and Frey, 1999). For probabilistic models, sensitivity analysis can be based upon attribution of variation of the model response to variation in model inputs. In this context, sensitivity analysis is the study of how the variation in the output of a model can be apportioned, qualitatively or quantitatively, among model inputs (Saltelli ***et al.***, 2000). Although many sensitivity analysis methods focus on apportionment of the variance in an output to variance in

model inputs, other measures of sensitivity can be considered. For example, one could focus on characterizing the combinations of model input values that produce high-end exposure estimates.

In interpreting the results of sensitivity analysis, a distinction can be made between an input that is *sensitive* versus one that is *important* to a particular assessment objective. A sensitive input is any input that has a significant influence on the variation of the output (Saltelli *et al.,* 2000). An input is more sensitive than another if it produces a larger contribution to the measure of sensitivity being used, such as contribution to variance in an output (Hofer, 1999; Helton and Davis, 2002) or association with high-end values of particular concern (Mishra *et al.,* 2003).

In the context of food safety, an important input is one that is sensitive but that is also controllable in practice and could be changed in order to reduce risk. For example, Critical Control Points (CCPs) are points, steps, or procedures in the process of bringing food from farm to the table at which direct control can be applied, and a food safety hazard can be prevented, eliminated, or reduced to an acceptable level (Hulebak and Schlosser, 2002; Seward, 2000). Potential CCPs are typically incorporated in MFSPR models. Sensitivity analysis results can lead to identification of important CCPs among the list of sensitive inputs.

Sensitivity analysis methods are typically classified into categories based on their scope, applicability, and characteristics. For example, Frey and Patil (2002) broadly classify sensitivity analysis methods with regard to characteristics as mathematical, statistical, and graphical methods. Alternatively, Saltelli *et al.* (2000) classify sensitivity analysis methods with respect to the application and scope as screening, local, and global. Examples of commonly used sensitivity analysis methods are regression-based techniques (e.g., Helton, 1993), variance decomposition

methods (e.g., Saltelli *et al.*, 1999), and scatter plots (e.g., Kleijnen and Helton, 1999; Saltelli *et al.*, 2000; Frey and Patil, 2002).

Frey *et al.* (2004) presented a comparison of the capabilities of many sensitivity analysis methods with regard to the characteristics of MFSPR models and with regard to analytic objectives. Methods were compared with respect to characteristics such as ability to: consider simultaneous variation in inputs; address non-linearity, interactions, and thresholds; handle alternative input types; address high exposure scenarios; provide quantitative measures of importance; and implement easily. A key conclusion was that methods conventionally used for sensitivity analysis and available in commercial statistical software packages (e.g., correlation analysis and stepwise regression analysis) may not provide robust results when applied to MFSPR models, and hence, should be used carefully, if at all. However, CART was identified as a promising method that can deal with nonlinearity, interaction, thresholds, and categorical inputs, as well as provide insight into situations that produce high exposure or risk. Hence, CART is selected here as a focus for further evaluation.

The application of CART for sensitivity analysis is demonstrated via case studies with the *E. coli O157:H7* food safety process risk model in ground beef. This model was selected because it represents an important public policy issue and has key characteristics of typical MFSPR models. Ground beef servings contaminated with *E. coli O157:H7* organisms are estimated to be responsible for some 73,500 cases of infection, 2,150 hospitalizations, and 61 deaths in the United States each year (Mead *el al.*, 1999). Children under 5 and elderly people over 65 years old have the greatest concern regarding the infection with the *E. coli O157:*H7 organisms (FSIS, 2001). The case studies demonstrate the capabilities of CART in addressing

specific model characteristics, such as non-linearity, thresholds, interaction between inputs, and categorical inputs.

The results presented in this paper address the following key questions:

- What are the advantages of CART as a method for sensitivity analysis when applied to MFSPR models?
- How are key characteristics of MFSPR models addressed using CART?
- How can CART be used to gain insight into the contribution of sensitive inputs to high exposure or risk scenarios?

A brief description of CART is provided. The *E. coli* model is briefly discussed and a specific probabilistic analysis scenario is defined for evaluation of CART. Results based on the application of CART are presented. Comparisons are provided between the CART results and those from two commonly-used sensitivity analysis methods, including correlation analysis and stepwise regression analysis. Finally, comments regarding the general applicability of CART to MFSPR models are presented.

2.0 Overview of Classification and Regression Trees

CART has several advantages as a sensitivity analysis method. CART is nonparametric and does not require assumptions of a particular distribution for the error term. CART is resistant to the effects of outliers since splits usually occur at non-outlier values (Roberts et al., 1999). CART results are invariant with respect to monotonic transformations of the independent variables. As a result, it is not necessary to test a number of transformations to find the "best" fit (Hallmark et al., 2002). CART can capture non-additive behaviors (Mishra *et al*, 2003). Because the data in the case study are from a complex non-linear MFSPR model, this characteristic is a distinct advantage over conventional regression-based sensitivity analysis methods.

Tree-based models are statistical models designed for supervised prediction problems, and can be used for partitioning data (Breiman *et al.,* 1984). In a supervised problem a set of inputs is used to predict the output values. When the output is categorical, the model is a *classification tree*. For a continuous output, the tree-based model is a *regression tree.* The combined methodology is referred to as classification and regression trees (CART).

Figure 1 shows a schematic diagram of a tree-based model. Each tree is read from top to bottom starting at the *root node*. Each internal node represents a split base on the values of an input. Each input can be selected once or multiple times in a tree. Each split at an internal node represents an inequality. A case moves left if the inequality is true and right otherwise. The terminal nodes of the tree are called *leaves*. Each leaf represents a predicted mean output value. For a categorical output, each leaf shows specific classes of the output, while for a continuous output, leaves show statistically significantly different mean values for the output. Each path from the root node to a leaf is a "classification rule" that specifies specific cut-off values of selected inputs that lead to a leaf.

Each split in a tree is selected in order to reduce some measure of node impurity (Loh *et al.,* 1997). In classification trees, three common measures of node impurity are Gini Index (Breiman *et al,* 1984), entropy (Quinlan, 1993), and the chi-squared test (Kass, 1980). For regression trees, deviance is typically used as a measure of node impurity (Morgan and Sonquist 1963, Breiman *et al.,* 1984). A splitting criterion is typically defined based on the reduction in the value of a measure of node impurity associated with a specific input and a cut-off value selected for the split. An optimal split is the one that maximizes the reduction in the measure of node impurity. For the case studies presented in this paper, regression trees are used since the selected outputs are continuous.

Figure 2 shows a schematic diagram of a split, in which a dataset in a parent node with n_0 data points is classified into two child nodes with n_1 and n_2 data points, respectively, where $n_0 = n_1 + n_2$. The split is defined based on an input and a cut-off value within the domain of the input. The cut-off value is used to classify n_0 data points into n_1 and n_2 data points. The number of data points that are less than or equal to the cut-off values is n_1. Conversely, n_2 is the number of data points for which values are greater than the cut-off value. A parent node can be a root node or an internal node, while a child node can be an internal node or a leaf. The measure of impurity in the split based on the change in deviance (ΔD) is defined as:

$$\Delta D = D(0) - \{D(1) + D(2)\} \tag{1}$$

D(t) represents the total deviance in the node *t* and is calculated as:

$$D(t) = \sum_{j=1}^{n_t} (y_{jt} - \bar{y}_t)^2 \tag{2}$$

where,

n_t = Number of data points at node *t*; *t* = 0, 1, 2.

y_{jt} = j^{th} output value at node *t*; *t* = 0,1, 2, and *j* = 1,…, n_t.

$\bar{y}_t$ = Mean output value at node *t*; *t* = 0, 1, 2.

An appropriate split and corresponding cut-off value are those that maximize the *DD* value in Equation 1. Termination of splitting typically is specified to occur when the number of data points at a node drops below a selected minimum, or when the maximum possible reduction in the dataset impurity for a particular node is less than a selected minimum (Breiman *et al.*, 1984).

When using CART for sensitivity analysis, the nodes are typically evaluated qualitatively based on their relative contribution in determining sensitive inputs. The earliest splits contribute most to the reduction of impurity in the dataset, and hence, inputs selected in those splits are

typically considered to be most sensitive. Unless there is collinearity between inputs, CART selects only the most sensitive inputs and values of these inputs that result in the maximum reduction in the dataset impurity (Breiman *et al.*, 1984). For MFSPR models, sample values of inputs are typically simulated as independent; hence, model inputs are not collinear. Typically, but not always, inputs used as a basis for splits in the top nodes have more influence on the response variable in comparison with inputs in the lower nodes. Thus, input precedence in the tree provides insight regarding sensitivity. Another indication of sensitivity for an input is if it is selected repeatedly at multiple levels within the tree.

Breiman *et al.* (1984) devised a measure of input sensitivity for CART based upon the contributions of inputs to the reduction in the dataset impurity. Based on this measure, the amount of cumulative contribution of selected inputs in a tree to the reduction of the total dataset impurity indicates input sensitivity. For regression trees and for cases that an input is selected multiple times in a tree, corresponding to each selection, a value for *DD* is calculated using Equation (1). Inputs are prioritized with regard to sensitivity based on the relative magnitude of their total contribution to the reduction of the dataset impurity (e.g., deviance). Hence, the higher the contribution of the input to the reduction of the total impurity, the higher is the rank.

CART was applied to the case studies using S-PLUS© Version 6.1. This software has the ability to perform CART analysis on a dataset using a Graphical User Interface (GUI) (S-PLUS, 2000).

3.0 Overview of the *E. coli O157:H7* Food Safety Process Risk Model

The US Department of Agriculture (USDA) developed a farm-to-table process risk model to evaluate the public health impact from *E. coli O157:H7* in ground beef (FSIS, 2001). This

model was developed in Microsoft Excel using @Risk, an add-in software package, as the basis for performing probabilistic simulations.

The *E. coli O157:H7* risk assessment includes hazard identification, exposure assessment, hazard characterization, and risk characterization steps. The *E. coli* model has a modular framework. In a modular framework, each key activity in the model is modeled in a separate module. Each module may have inputs that are calculated based upon a predecessor module, and may produce outputs that are fed to a successor module. Within modules, binning is often used. Binning is a summarization technique in which values of an output are binned into predefined intervals. For example, in the *E. coli* model the contaminant concentration in beef trimings is estimated as a continuous output, but subsequently is binned by 0.5 log increments from 0 to 8 logs (FSIS, 2001). Because of binning, there is a loss of one-to-one correspondence between outputs of one module and the inputs of its predecessors. Therefore, binning is an obstacle to global sensitivity analysis of an entire model. Thus, for the case studies, CART is applied to specific modules in the exposure assessment step.

In the *E. coli O157:H7* model, the exposure step is the foundation for other steps. For example, in the hazard characterization step, the dose response relationship is estimated based on the results of the exposure assessment step and available surveillance data regarding infection with *E. coli O157:H7*. Uncertainty in exposure results leads to uncertainty in the dose-response relationship. The exposure step also quantifies the relationship between potential CCPs and exposures. Thus, the case studies were focused on this step.

The exposure assessment step consists of three major modules: (1) production; (2) slaughter; and (3) preparation. The preparation module consists of three parts: (1) growth estimation; (2) cooking effect; and (3) serving contamination. The exposure assessment is based

upon a probabilistic approach for modeling the prevalence and concentration of the *E. coli* pathogen in live cattle, carcasses, beef trim, and a single serving of cooked ground beef.

Figure 3 shows a schematic diagram of the connection between the final model output and different modules and parts of the model in the exposure assessment step. Two sections of the exposure assessment step were selected for the case studies: (1) cooking effect part of the preparation module; and (2) slaughter module. These sections have characteristics such as thresholds and interactions between inputs. The mean values and 95 percent probability ranges for each input are given in Tables 1 and 2 for the cooking effect part and the slaughter module, respectively. Results from application of CART to other parts and modules of the exposure assessment step are available in Frey *et al.* (2003).

The cooking effect part evaluates the effect of cooking ground beef servings on the total number of *E. coli O157:H7* organisms considering the cooking temperature, precooking treatments, and the cooking place (i.e., home versus away from home). Precooking treatments are characterized based upon precooking storage temperature and storage time. Nine alternative precooking treatments are given in Table 3. The output in the cooking effect part is the mean log reduction in the number of *E. coli* organisms due to cooking per ground beef serving.

In the slaughter plant, combo bins are filled with meat trims from cattle harvested from feedlots and breeding herds. Combo bins are containers that hold 2,000 lbs of meat trims. The output from the slaughter module is a distribution describing the frequency of the *E. coli* organisms in combo bins made during high and low prevalence seasons (i.e., summer and winter).

4.0 Probabilistic Analysis Scenarios

This section explains the probabilistic analysis scenarios defined for each case study. The cooking effect part has a one-dimensional probabilistic simulation incorporating variability. The number of iterations selected for the variability analysis simulation is 65,000, based on the upper limit of what can be stored in each Excel sheet.

In the slaughter module, inputs incorporate both variability and uncertainty. Alternative approaches can be used to distinguish between variability and uncertainty in inputs and their contribution to the output variation (Cullen and Frey, 1999; Frey *et al.*, 2004). These approaches include: (1) variability only; (2) uncertainty only; (3) variability for different realizations of uncertainty; (4) uncertainty for different realizations of variability; and (5) co-mingled of variability and uncertainty. The choice of an appropriate approach should be made taking into account the assessment objectives, the data quality objectives, the availability of data, and the importance of the assessment. For the purpose of demonstrating the capabilities of CART, a one-dimensional probabilistic simulation is adequate.

Co-mingled analysis of variability and uncertainty is selected as the probabilistic approach in the slaughter module. In this approach, variability and uncertainty in inputs are incorporated but not distinguished, thereby producing in only one dimension the largest range of values that might be assigned to each model input. This approach could be a preliminary step in prioritizing data collection or the development of distributions for model inputs (Frey *et al.*, 2004). An analyst may wish to assess the key sources of variation regardless of whether they represent variability or uncertainty. Moreover, an analyst may make a judgment that it is difficult to separate variability from uncertainty (e.g. Nauta, 2000) or that a two-dimensional probabilistic approach is impractical or not necessary based upon the assessment objectives. For the case study, 65,000 realizations are performed.

5.0 Application of CART

This section presents results from application of CART to the probabilistic analysis scenarios defined in Section 4. Section 5.1 shows how CART addresses interactions between inputs and can be used to identify thresholds in the model response based on a case study with the cooking effect part. Section 5.2 demonstrates the capability of CART to address high exposure scenarios with a case study for the slaughter module. When applicable, CART results are compared to results from conventional sensitivity analysis methods, including stepwise regression analysis and correlation analysis.

5.1 Application of CART to the Cooking Effect Part

The regression tree from application of CART to the cooking effect part is shown in Figure 4. The vertical distance below each split indicates the portion of deviance that is reduced because of the split. Cooking temperature is inferred to be the most sensitive input because it is selected in the first node and appears repeatedly throughout many of the lower nodes of the tree. Precooking treatment appears several times in the tree but not in the first node, suggesting that this input is of secondary sensitivity.

The total deviance for the simulated output values is 639,900. The regression tree reduces approximately 95 percent of the total deviance. The contribution of each input to reduction in deviance is summarized in Table 4. Each appearance of an input in the tree is denoted with a numerical entry in the table. For example, cooking temperature appears three times in the fourth level of the tree at nodes 6, 7, and 8, and hence, three values are calculated for the reduction in the total deviance corresponding to these nodes. The total reduction in the deviance associated with each input is based upon the cumulative effect of repeated splits at lower levels of the tree.

Cooking temperature has a larger contribution to the reduction of the total deviance. Hence, this input is selected as the most sensitive one.

Stepwise regression analysis is selected as a conventional sensitivity analysis method for comparison with the CART results. There are categorical inputs in the cooking effect part. Thus, correlation analysis cannot be used for sensitivity analysis. The results given in Table 5 show that a regression model that includes cooking temperature and precooking treatment accounts for almost 76 percent of the output variation. Moreover, results show that there is a statistically significant interaction between selected inputs in the regression model that explains almost 5 percent of output variation. The interaction is modeled based upon a multiplicative term.

Inputs are ranked based on their partial R^2 value. The partial R^2 value for each step is determined as the difference between the R^2 values for the current step and the R^2 values for the previous step. Based on the partial R^2 values, cooking temperature and precooking treatments are ranked as first and second inputs, respectively. In stepwise regression, only the inputs that have P values less than 0.05 are included indicating that selected inputs and the interaction term are statistically significant. Stepwise regression analysis provides similar results with respect to identification of inputs with highest sensitivity when compared to CART. Despite this similarity, Sections 5.1.1 and 5.1.2 demonstrate that CART provides insights not available with commonly used methods such as stepwise regression.

5.1.1 Addressing Interactions Using CART

The ability of CART to account for interaction between sensitive inputs makes it a useful method compared to linear or additive models such as regression models. CART can provide explicit insight regarding the nature of the interaction effect between inputs by identifying how inputs affect model response, without any assumptions regarding the model structure. In contrast,

when using regression analysis, the analyst must hypothesize what the interaction is and include appropriate terms in the regression model to attempt to account for such interactions. As the number, complexity, or both, of interactions increases, the development of appropriate regression models becomes more difficult. Thus, an advantage of CART is that important interactions are represented without need for prior specification of interaction terms.

Although the regression results in Table 5 imply that there is a significant interaction between cooking temperature and pretreatment, the inferred characteristics of the interaction are predicated upon the assumed regression model. By comparison, the CART results enable discovery of more detailed and specific insights as to the interaction, which is more complex than simply a multiplication of two terms. Figure 4 shows that the partitions with respect to precooking treatment typically separate the first few such treatments (i.e., *a*, *b*, and *c*) from the remainder. The mean log reduction in *E. coli* organisms is greatest for the leaf number 12, corresponding to an average reduction of 12.1. This average reduction is based upon a cooking temperature of greater than 85 ^{o}C and precooking treatments of *c* through *i*. In contrast, the lowest reduction is based upon a cooking temperature of less than 53.2 ^{o}C irrespective of the type of pretreatment, as indicated the leaf number 1. The larger average reductions are typically associated with higher cooking temperatures and with precooking treatments other than *a* and *b*.

This interaction pattern indicates that not only increasing cooking temperature, but also defrosting ground beef servings at room temperature prior to cooking could lead to high values of log reduction in the number of *E. coli* organisms. Table 3 shows that for precooking treatments *c* to *i* ground beef servings are kept in room temperature (e.g., temperature between 21^{o}C and 30^{o}C) for few hours prior to cooking. This insight can be used as a recommended practice for cooking ground beef servings such as hamburger patties.

Thus, when comparing CART to a more typically used methods, key advantages are that CART not only enables identification of the presence of an interaction, but provides insights into the complexity of the interaction and regarding specific cut-points for selected inputs that could be identified as the basis for eliminating or reducing undesirable outcomes. A sensitivity analysis method such as regression that involves making assumptions regarding model form potentially can produce unreliable or misleading insights regarding cut-points.

5.1.2 Addressing Thresholds Using CART

The ability of CART to identify a threshold is demonstrated via an example in the cooking effect part. Conventional sensitivity analysis methods, such as regression analysis are not able to identify thresholds unless an analyst has prior knowledge of the threshold and specifies a functional form to properly capture it. However, in typical practice, sensitivity analysis results based on correlation and regression methods are undermined if thresholds are present since thresholds introduce non-linearity in the model. For example, the sample (Pearson) correlation coefficient for output values versus cooking temperature is 0.92 for precooking treatment *c*, implying that there is a strong linear association between the output and input values. However, in actuality there is a nonlinear relationship that includes a threshold. Furthermore, the correlation coefficient method cannot provide insight regarding the numerical value of the threshold.

The partitioning of the data with respect to cooking temperature in Figure 4 suggests that there could be a temperature threshold below which there is low reduction in *E. coli* organisms. In particular, for temperatures less than 58 ^{o}C the log reduction in *E. coli* organisms was substantially smaller than for any other temperature range.

In order to confirm the identified threshold, a scatter plot for the log reduction in the number of *E. coli* organisms versus cooking temperature is given in Figure 5. Each line in this figure represents a specific precooking treatment. There is a threshold in the response of the model to the cooking temperature. For example, below approximately 50 ^{o}C there is typically an insignificant log reduction in the number of *E. coli* organisms. At temperatures of approximately 50 ^{o}C to 58 ^{o}C there is a small log reduction (of 0 to 3 logs) due to cooking depending on the precooking treatment. Above 58 ^{o}C, there is more sensitivity of the log reduction in *E. coli* organisms to cooking temperature for all the precooking treatments.

5.2 Application of CART to the Slaughter Module to Address High Exposure Scenarios

This section discusses how CART can identify high exposure scenarios and corresponding sensitive inputs based upon application to the slaughter module. This ability of CART is a key advantage when compared to conventional sensitivity analysis methods, as such insights may significantly inform risk management decision making.

Four out of twelve inputs are selected in the regression tree shown in Figure 6. These inputs include chilling effect, number of organisms, washing effect, and contaminated surface of meat trims. The highest average combo bin contamination corresponds with leaf 9 in which there are initially more than approximately of 128 organisms and the chilling effect is greater than 2.2 logs.

The combo bin contamination has the greatest sensitivity to both chilling effect and the number of organisms. The chilling effect appears in the first node of the tree. However, although the number of organisms was not selected until the second level of the tree, this input discriminates the mean response of 3,641 from another branch with much lower mean responses.

Thus, the partitioning of data with respect to chilling effect accounts for a wide range of variation in the response.

The total deviance for the simulated output values in the slaughter module is 3.2×10^{9}. The regression tree shown in Figure 6 reduces the total deviance by approximately 53 percent. The cumulative contributions of each input to the reduction of the total deviance are given in Table 6. The initial number of organisms produces the single largest cumulative contribution to reduction in deviance, although this input is not selected at the first splitting node. Chilling effect is selected at the first splitting node but has a rank of second based upon its cumulative contribution.

The rankings of sensitive inputs based upon CART are summarized in Table 7, along with a comparison to rankings obtained with other methods, including sample and rank regression and sample and rank correlation. Because the model is nonlinear, linear regression using sample values produces a relatively low coefficient of determination. This implies that the rankings from both sample regression and sample correlation may not be reliable. The rankings from these two approaches differ from those of the rank-based regression and correlation. These latter two methods assume monotonicity. The general qualitative agreement regarding the ranking of the two or three most sensitive inputs based upon CART, rank regression, and rank (Spearman) correlation suggest that CART is capable of producing appropriate rankings. The differences among these three methods for lower ranked inputs are typically not of practical significance, since the first two inputs have a much stronger sensitivity than any of the others. Thus, disagreement regarding the rankings of relatively insensitive inputs is not of great consequence.

Because CART is a nonparametric method, it does not assume any specific functional relationship between values of model response and inputs. Hence, sensitivity analysis results based on CART are not undermined when the model has characteristics such as nonlinearity and interactions, while results based on conventional sensitivity analysis methods are sensitive to these assumptions.

Figure 6 also specifies the conditions corresponding to high average contamination scenarios, such as for leaves 6 and 9. The classification rules for these leaves indicate that high values of contamination are mostly controlled by chilling effect and the initial number of organisms on carcasses. Furthermore, by comparing similar classification rules that result in different average contamination levels, one can gain insight into effective measures for reducing average contamination levels. For example, a comparison of the classification rules for leaves 5 and 6 indicate that for carcasses that have initial contamination levels greater than 2,014 organisms controlling the chilling process can lead to comparatively low combo bin contamination. Growth or decline in *E. coli O157:H7* on the surface of carcasses is largely the function of time and temperature. Fluctuation in chiller temperature or failure to adequately chill carcasses may enable growth (FSIS, 2001).

The results from the CART analysis suggest that the process of chilling carcasses is an area for critical control. Even with sufficient decontamination prior to the chilling process, such as washing carcasses, uncontrolled conditions during the chilling process (e.g., uncontrolled chilling time and temperature) leads to substantial contamination in combo bins. CART suggests that a cut-off value of 2.2 logs for growth of organisms during chilling is a critical value beyond which high contamination is expected. This finding also suggests that there may be critical limits associated with chilling time and temperature below which growth can be controlled below the

critical value of 2.2 logs. However, chilling time and temperature are not modeled in the slaughter module. Thus, an implication of the results is that it is beneficial to disaggregate the model and incorporate more detail regarding specific factors that influence the chilling effect in order to estimate the corresponding critical limits.

The ability to identify complex interactions and inform the search for critical limits is an advantage of CART versus methods such as regression analysis. Although it is possible in regression analysis to account for interactions via introduction of addition terms in the model, the results will depend upon how well the regression model mimics the behavior of the MFSPR being simulated. Errors in the regression model structure could translate into misleading insights. Regression analysis may provide conditional insight with respect to high (or low) exposure scenarios. For example, one may use the regression model to estimate the value for a selected input corresponding to high value of the model response, while conditioning the values of other inputs at some point estimates (e.g., mean values). This critical value cannot be generalized as it is conditional to selected values of other inputs.

6.0 Discussion and Conclusions

CART is a powerful method for addressing both continuous and categorical inputs. This is an advantage compared with conventional sensitivity analysis methods. For example, correlation analysis fails to provide accurate insight with respect to sensitivity when categorical inputs are included in the model. Furthermore, CART does not assume any functional relationship between values of output and inputs. These characteristics are key advantages of CART when applied to MFSPR models with typical features such as non-linearity, interactions, threshold and saturation points, and categorical inputs.

Two approaches were presented for inferring sensitivity using CART: (1) visualization of the tree; and (2) contribution of each input to reduction in impurity. These approaches provided approximately similar rankings. In one case, the top ranked input as identified by inspection of the tree was found to have the second largest contribution to reduction in total deviance, but was also identified as substantially more sensitive than the next ranked input.

The ability of CART to incorporate interaction between sensitive inputs was demonstrated. This ability is an advantage when the number of model inputs is substantially high and there are complex interactions that cannot easily be captured by a regression model. Moreover, CART provides detailed insights regarding interaction effects particularly with regard to combinations of specific ranges of values of inputs that produce a particular outcome of interest. Such insight may be directly relevant to identifying risk management alternatives.

CART responds appropriately to the existence of thresholds in the model. This is a key advantage of CART compared to conventional sensitivity analysis methods. A threshold introduces non-linearity in the model response, which typically undermines sensitivity analysis results based on methods that assume a specific functional relationship. Moreover, CART provides insight regarding threshold values since these will typically be selected as cut-off values in a split. However, the selected cut-off value should often be verified using results from other sensitivity analysis methods, including graphical methods such as scatter plots, in order to understand whether the cut-off is a threshold or saturation point.

CART provides explicit insight into the ranges of values of sensitive inputs associated with the worst, or best, outcomes. Other conventional sensitivity analysis methods are not readily able to provide such an insight. For example, the classification rules for the highest mean exposure identify the combination of ranges of sensitive inputs that is associated with the highest

model output. The identification of highest exposure classes is important toward identifying situations to avoid via more effective control of important inputs. Similarly, by identifying the classification rules for the lowest exposure outcomes, one may develop detailed insight regarding critical limits for achieving desired risk management outcomes. Thus, this capability of CART can substantially facilitate the process of identification of important inputs and critical limits from the stand point of risk management.

7.0 Acknowledgment

The work reported here is based on work supported by Cooperative Agreement No. 58-0111-0-005 between the U.S. Department of Agriculture, Office of the Chief Economist, Office of the Risk Assessment and Cost Benefit Analysis (USDA/OCE/ORACBA) and the department of Civil Engineering at North Carolina State University. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Agriculture.

8.0 References

Anderson, E.L., Hattis, D.H., 1999. When and how can you specify a probability distribution when you don't know much? Risk Analysis, 19(1): 43-68.

Breiman, L., Friedman, J.H, Olshen, R.A., and Stone, C.J. (1984). Classification and regression trees. Belmont, Calif.: Wadsworth International Group

CDC, 1982. Epidemic notes and reports: Isolation of *E. coli O157:H7* from sporadic cases of hemorrhagic colitis. United States Morbidity and Mortality Weekly Report, 31(43): 580, 585.

CFSAN, 2001. Draft risk assessment on the public health impact of *Vibrio parahaemolyticus* in raw molluscan shellfish. Prepared by the *Vibrio parahaemolyticus* risk assessment task force (Center for Food Safety and Applied Nutrition) for Food and Drug Administration (FDA), USDA, Washington, DC.

CFSAN, 2003. Quantitative assessment of relative risk to public health from foodborne *Listeria monocytogenes* among selected categories of ready-to-eat foods. Prepared by Center for Food Safety and Applied Nutrition, Food and Drugs Administration and Food Safety and Inspection Service, U. S. Department of Agriculture, Washington DC, September 2003, http://www.foodsafety.gov/~dms/lmr2-toc.html.

Cullen, A.C., Frey, H.C., 1999. Probabilistic techniques in exposure assessment. Plenum Press: New York.

Frey, H.C., 2002. Introduction to special section on sensitivity analysis and summary of NCSU/USDA workshop on sensitivity analysis. Risk Analysis, 22(3): 539-545.

Frey, H.C., Mokhtari, A., Tanwir, D., 2003. Evaluation of selected sensitivity analysis methods based upon applications to two food safety process risk models. Prepared by North

Carolina State University for office of risk assessment and cost-benefit analysis, U.S. Department of Agriculture, Washington, DC, June 2003, www.ce.ncsu.edu/risk.

Frey, H.C., Mokhtari, A., Zheng, J., 2004. Recommended practice regarding selection, application, and interpretation of sensitivity analysis methods applied to food safety process risk models. Prepared by North Carolina State University for office of risk assessment and cost-benefit analysis, U.S. Department of Agriculture, Washington DC , January 2004, www.ce.ncsu.edu/risk.

Frey, H.C., Patil, R., 2002. Identification and review of sensitivity analysis methods. Risk Analysis, 22(3): 553-577.

FSIS, 2001. Draft risk assessment of the public health impact of *E. coli O157:H7* in ground beef. Prepared by the *E. coli O157:H7* Risk Assessment Team for the Food Safety and Inspection Service and United States Department of Agriculture, Washington, DC. http://www.fsis.usda.gov/OPPDE/rdad/FRPubs/00-023N/00-023NReport.pdf.

Hallmark, S.L., Guensler, R., Fomunung, I., 2002. Characterizing on-road variables that affect passenger vehicle model operation. Transportation Research Part D: Transport and Environment, 7(2): 81-98.

Helton, J. C., 1993. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. Reliability Engineering and System Safety, 42(2-3): 327-367.

Helton, J.C., Davis, F.J., 2002. Illustration of sampling-based methods for uncertainty and sensitivity analysis. Risk Analysis, 22(3): 591-622.

Hofer, E., 1999. Sensitivity analysis in the context of uncertainty analysis for computationally intensive models. Computer Physics Communications, 117(1-2): 21-34.

Hulebak, K.L, Schlosser, W., 2002. Hazard analysis and critical control point (HACCP) history and conceptual overview. Risk Analysis, 22(3): 553-578.

Jaykus L.A., 1996. The application of quantitative risk assessment to microbial food safety risks. Critical Reviews in Microbiology; 22: 279-93.

Kass, G.V., 1980. An explanatory technique for investigating large quantities of categorical data. Applied statistics, 29(2): 119-127.

Kleijnen, J.P.C., Helton J.C., 1999. Statistical analyses of scatterplots to identify important factors in large-scale simulations, 1: Review and comparison of techniques. Reliability Engineering and System Safety, 65(2): 147-185.

Loh W., Shih Y., 1997. Split selection methods for classification trees. Statistica Sinica, 7: 815-840.

Mead, P.S., Laurence, S., Vance, D., McCaig, L.F., Bresee, J.S., Shapiro, C., Griffin, P.M., Tauxe, R.V., 1999. Food-related illness and death in the United States. Emerging Infectious Diseases, 5(5): 607-625.

Mishra S., Deeds, N.E., RamaRao, B.S., 2003. Application of classification trees in the sensitivity analysis of probabilistic model results. Reliability Engineering and System Safety, 79(2): 123-129.

Morgan, J.N., Sonquist, J.A., 1963. Problem in the analysis of survey data, and a proposal. Journal of the American Statistical Association, 58(302): 415-434.

Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics, 33(2): 161-174.

Murphy, B.L., 1998. Dealing with uncertainty in risk assessment. Human and Ecological Risk Assessment, 4(3): 685-699.

Nauta, M.J., 2000. Separation of uncertainty and variability in quantitative microbial risk assessment models. International Journal of Food Microbiology, 57(1-2): 9-18.

Quinlan, J.R., 1993. Programs for machine learning, San Mateo, CA: Morgan Kaufmann.

Roberts, C.A., Washington, S., Leonard, J., 1999. Forecasting dynamic vehicular activity on freeways: bridging the gap between travel demand and emerging emissions models. Presented at the 78th Annual Meeting of the Transportation Research Board, Washington, D.C.

Rose, J.B., Sobsey, M.D., 1993. Quantitative risk assessment for viral contamination of shellfish and coastal waters. Journal of Food Protection, 56(12): 1043-1050

Saltelli A., Chan, K., Scott, M., 2000. Sensitivity analysis, probability and statistics series. John Wiley & Sons: New York.

Saltelli, A., Tarantola, S., Chan, K., 1999. A quantitative, model independent method for global sensitivity analysis of model output. Technometrics, 41(1): 39-56.

SERAT, 1998. Salmonella Enteritidis risk assessment for shell eggs and egg products: Final Report. Prepared by Salmonella Enteritidis Risk Assessment Team for The Food Safety and Inspection Service (FSIS), USDA, Washington, DC.

Seward, S., 2000. Application of HACCP in food service. Irish Journal of Agricultural and Food Research, 39(2): 221-227.

SPLUS, 2000. Guide to statistics, Volume I and II, data analysis product division, MathSoft Inc.: Seattle, WA.

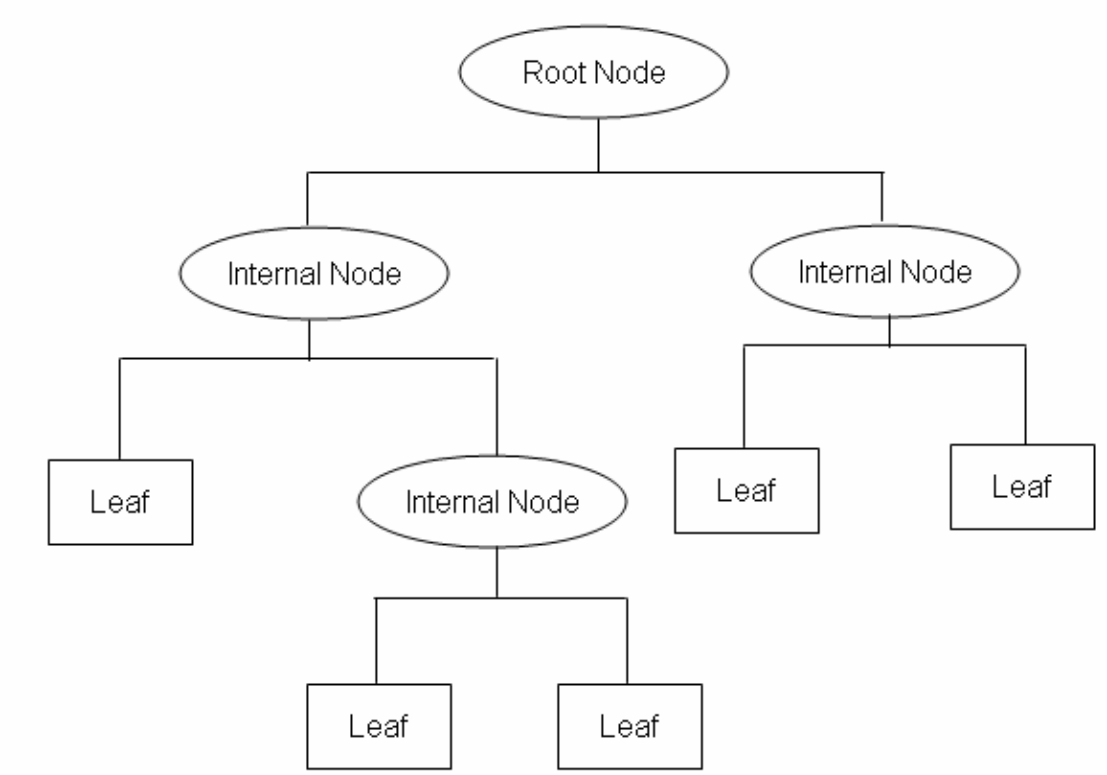


Figure 1. Schematic Diagram of a Classification and Regression Tree Illustrating Root Node, Internal Nodes, and Terminal Nodes (Leaves).

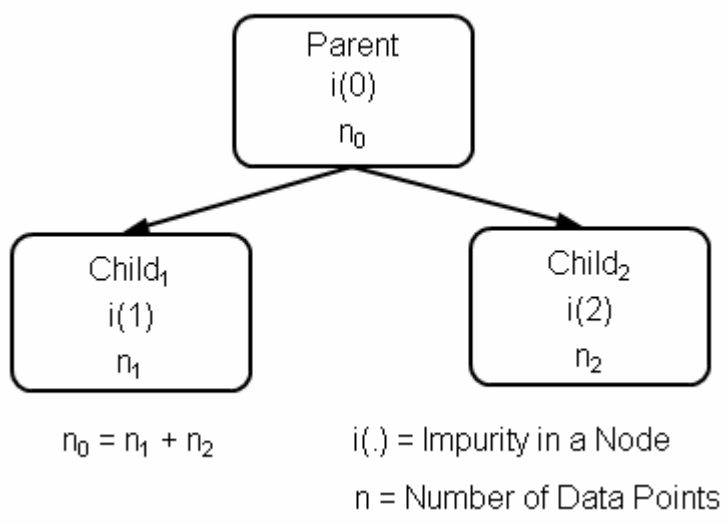


Figure 2. Schematic Diagram of a Split with n_0 Data Points.

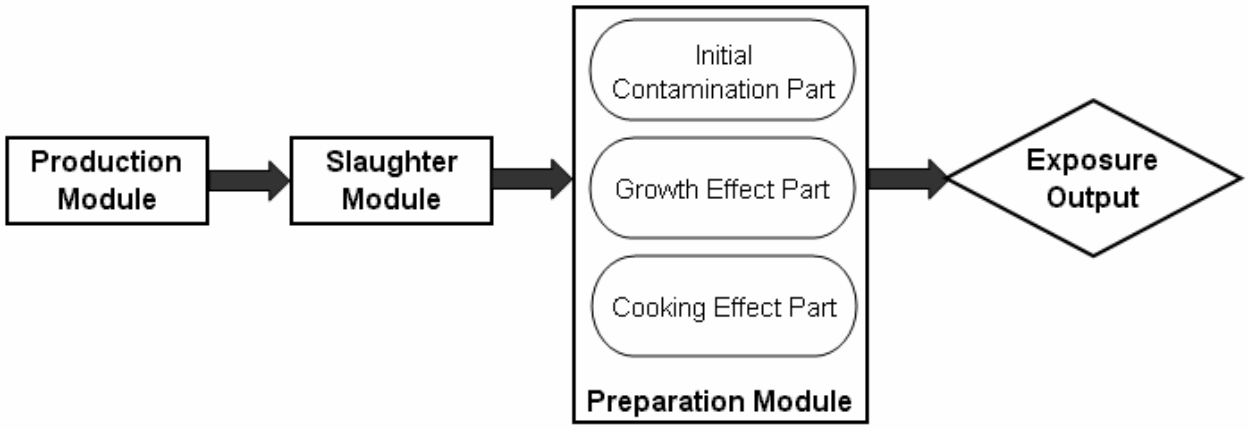


Figure 3. Schematic Diagram of the Connection Between the Exposure and Different Modules and Parts of the *E. coli* Model.

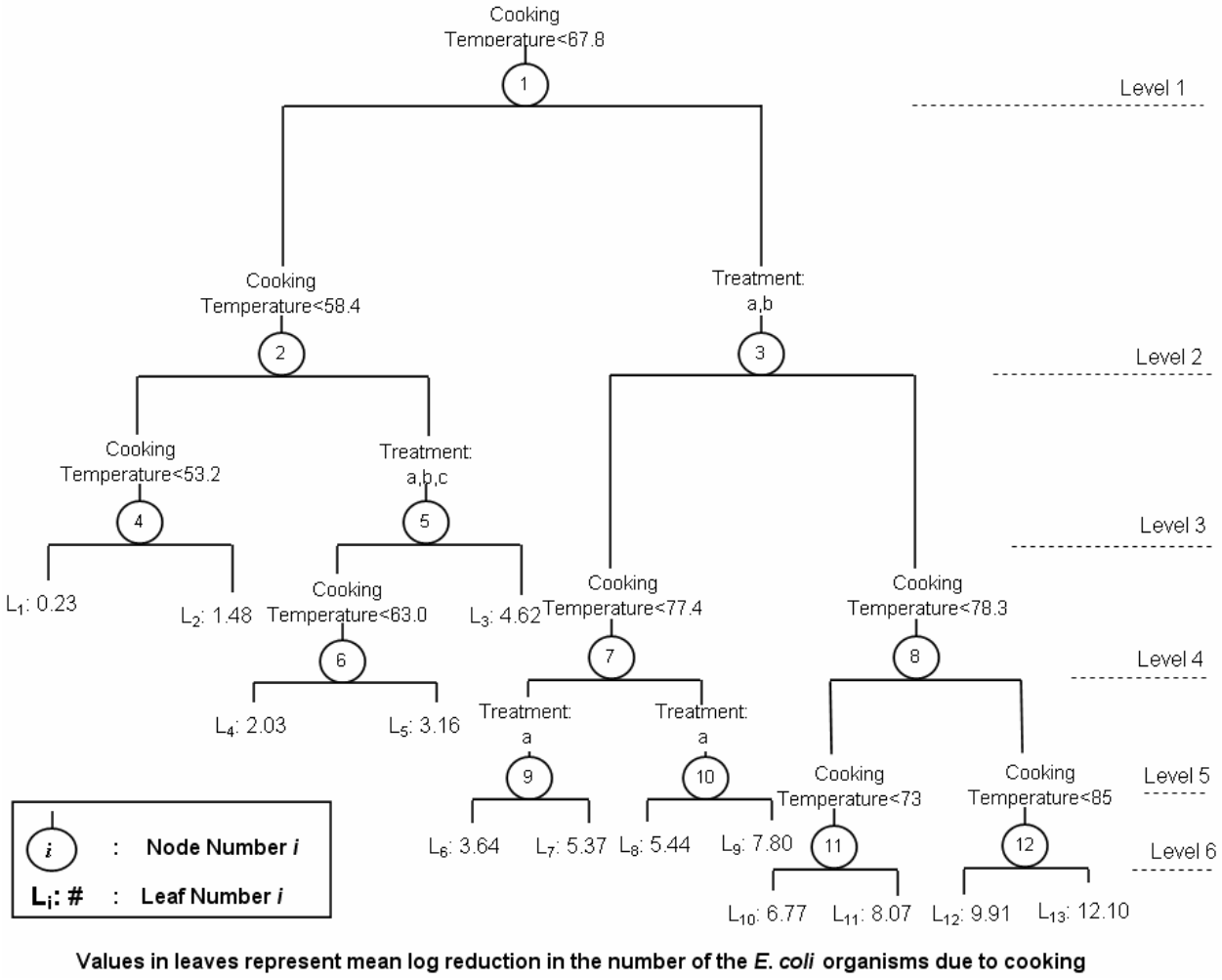


Figure 4. The Regression Tree for the Cooking Effect Part.

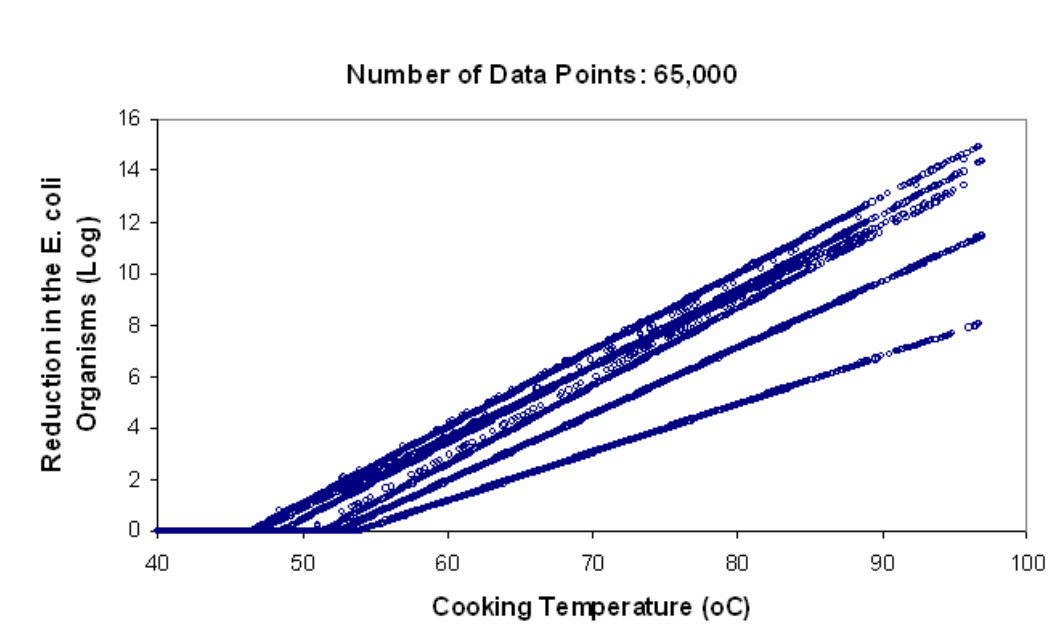


Figure 5. Scatter Plot for the Log Reduction in the Number of *E. coli* Organisms versus the Cooking Temperature at Home.

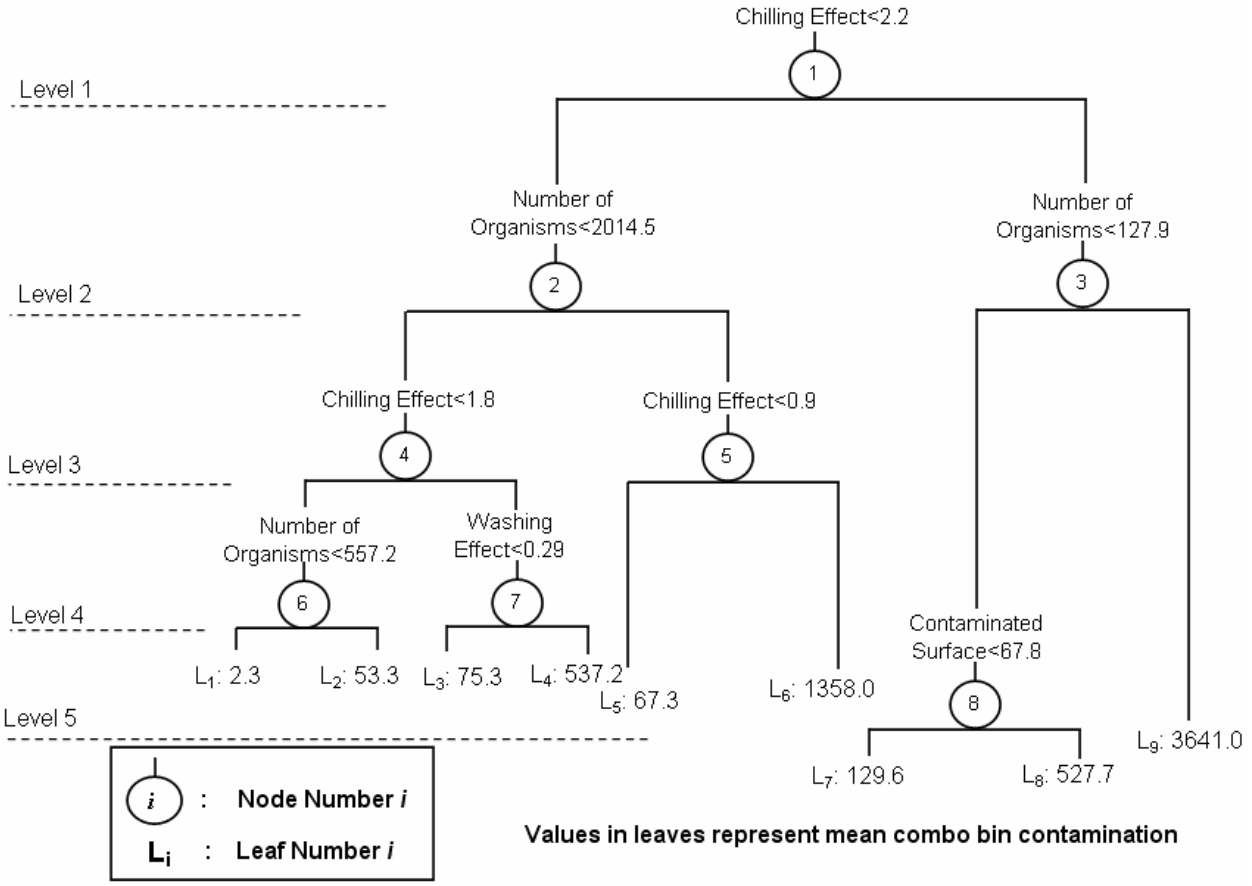


Figure 6. The Regression Tree in the Slaughter Module.

Table 1. Input Variables and the Corresponding Mean Value and 95 Percent Probability Range in the Cooking Effect Part

Input	Mean Value	95% Probability Range	Comment
Precooking Treatment	---	---	Categorical input; refer to Table 3 for the list of levels.
Cooking Place	---	---	Categorical input; holds "*Home*" and "*Away from Home*" values.
Cooking Temperature (^{o}C)	68.9	(44.10, 87.60)	Continuous input

Table 2. Input Variables and the Corresponding Mean Value and 95 Percent Probability Range in the Slaughter Module

Input	Mean Value	95% Probability Range
Total Number of Combo Bin for Each Carcass	4	(2, 6)
Total Number of Infected Animals	4	(0, 14)
Total Number of Contaminated Animals	5	(0, 22)
Probability of Positive Cases at both Steps of Dehiding and Evisceration (%)	39.7	(0, 96.7)
Number of Positive Cases at both Steps of Dehiding and Evisceration	--- (1)	--- (1)
Number of Positive Cases at Evisceration	--- (2)	--- (2)
Chilling Effect (logs)	-0.9	(-3.0, 2.1)
Number of Organisms	118.8	(0, 1100)
Trim/Vacuum/Washing Efficiency (%)	24.6	(0, 60)
Number of Organisms Added During Evisceration	1.2	(0, 2.45)
Washing Effect (%)	12.4	(0, 45)
Contaminated Surface Area (cm^2)	419.20	(0, 2590)

(1) Discrete input which holds values of 0, 1, and 2.
(2) Discrete input which holds values of 0, 1, 2, and 3.

Table 3. Specification of Precooking Treatments in the Cooking Effect Part

Pre-Cooking Treatment	Specification
a	-18°C for 8 days
b	3°C for 9 hours
c	-18°C for 8 days, 21°C for 4 hours
d	15°C for 9 hours
e	3°C for 9 hours, 21°C for 4 hours
f	-18°C for 8 days, 30°C for 4 hours
g	15°C for 9 hours, 21°C for 4 hours
h	3°C for 9 hours, 30°C for 4 hours
i	15°C for 9 hours, 30°C for 4 hours

Source: FSIS, 2001

Table 4. Reduction in Deviance Associated with Selected Inputs in the Regression Tree Generated in the Cooking Effect Part

Tree Level [(1)]	**Reduction in the Total Deviance Corresponding to Each Input** [(2)]	
	Precooking Treatment	**Cooking Temperature**
Level 1		380,170 (1)
Level 2	71,180 (3)	53,308 (2)
Level 3	18,258 (5)	3,638 (4)
Level 4		5,030 (6) 20,312 (7) 32,083 (8)
Level 5	7,128 (9) 8,877 (10)	
Level 6		3,734 (11) 5,152 (12)
Sum	105,443	503,427
Percent of Contribution	17.3	82.7
Rank	2	1

(1) Tree Levels are defined in Figure 4

(2) Numbers in parenthesis indicate the node number (See Figure 4)

Table 5. Comparison of the Results for Stepwise Regression Analysis and CART in the Cooking Effect Part

Input	Stepwise Regression Analysis				CART
	Partial R^2 Value	Model R^2 Value	F Value	Rank	Rank
Cooking Temperature (CT)	0.60	0.60	2,650	1	1
Precooking Treatment (PT)	0.16	0.76	1,630	2	2
CT×PT (1)	0.05	0.81	120		

(1) Interaction effect between cooking temperature and precooking treatment

Table 6. Reduction in Deviance Associated with Selected Inputs in the Regression Tree Generated in the Co-mingled Analysis of Variability and Uncertainty in the Slaughter Module

Tree Level [1]	**Reduction in the Total Deviance Corresponding to Each Input** [2]			
	Chilling Effect	**Initial Number of Organisms**	**Washing Effect**	**Contaminated Surface**
Level 1	357×10^6 (1)			
Level 2		46.4×10^6 (2) $1{,}057.6\times10^6$ (3)		
Level 3	22.7×10^6 (4) 154.5×10^6 (5)			
Level 4		5.3×10^6 (6)	45.6×10^6 (7)	
Level 5				14.6×10^6 (8)
Sum	503.2×10^6	$1{,}109.3\times10^6$	45.6×10^6	14.6×10^6
Percent Contribution	30.1	66.3	2.7	0.9
Rank	2	1	3	4

(1) Tree Levels are defined in Figure 6

(2) Numbers in parenthesis indicate the node number (See Figure 6)

Table 7. Comparison of the Results for Stepwise Regression Analysis, Correlation Analysis, and CART in the Slaughter Module

Inputs [1]	Stepwise Regression (Rank Regression)		Stepwise Regression (Sample Regression)		Pearson (Sample) Correlation Analysis		Spearman (Rank) Correlation Analysis		CART
	Partial R^2 Value	Rank	Partial R^2 Value	Rank	Coefficient	Rank	Coefficient	Rank	Rank[2]
CH_{eff}	0.640	1	0.080	2	0.145	2	0.830	1	1 (2)
N_{org}	0.120	2	0.126	1	0.172	1	0.730	2	2 (1)
W_{eff}	0.070	3	0.006	4	0.110	3	0.620	4	3
P_{both}	0.004	4	0.002	6	0.035	7	0.435	6	---
TNC	0.003	5	0.001	7	0.105	4	0.684	3	---
N_{evis}	0.001	6	0.004	5	0.065	5	0.152	7	---
TVW	0.001	7	0.040	3	0.060	6	0.590	5	---
CSA	--- [3]	--- [3]	--- [3]	--- [3]	--- [3]	--- [3]	--- [3]	--- [3]	4

(1) CH_{eff} = Chilling Effect; CSA = Contaminated Surface Area; N_{evis} = Number of *E. coli O157:H7* organisms Added During the Evisceration Step; N_{org} = Number of Organisms; P_{both} = Probability of Positive Contaminated Cases at Both Steps of Dehiding and Evisceration; TNC = Total Number of Contaminated Animals; TVW = Trim/Vacuum/Washing Efficiency; W_{eff} = Washing Effect;

(2) Rank in parenthesis is based upon the contribution of the input to the reduction of total deviance.

(3) Not selected as a statistically significant input.

PART V

TWO-DIMENSIONAL SENSITIVITY ANALYSIS: APPLICATION OF ANALYSIS OF VARIANCE TO THE *E. COLI* FOOD SAFETY PROCESS RISK MODEL

In Preparation to be Submitted to:

Risk Analysis

Abstract

The application of sensitivity analysis to a microbial food safety process risk (MFSPR) model with a two-dimensional probabilistic framework is demonstrated. MFSPR models typically have characteristics such as nonlinearity, thresholds, interactions, and categorical inputs that make the process of identifying controllable inputs and key sources of uncertainty a challenging task. The growth estimation part of the *E. coli O157:H7* food safety process risk model in ground beef is selected as a testbed. Analysis of variance (ANOVA) is selected as a sensitivity analysis method due to its ability to address the typical characteristics of MFSPR models. Sensitivity analysis results identified groups of inputs with comparable importance including: the most important input; inputs of secondary importance; inputs of minor importance; and unimportant inputs. Although wide ranges of uncertainty associated with the rank of the inputs are identified, the importance of the inputs is found to be substantially robust to uncertainty. This quantification helps a decision maker to confidently select inputs that matter the most for prospective risk management decisions. Bootstrap simulation is demonstrated as an approach for quantifying uncertainty in F values as sensitivity indices. Following the identification of the inputs related to the storage of ground beef servings at home as the most important inputs, risk management recommendations are suggested in the form of a consumer advisory for better handling and storage practices.

Key Words: Sensitivity Analysis, Microbial Food Safety Process Risk Models, Two-Dimensional Probabilistic Framework, Variability, Uncertainty, ANOVA

1.0 Introduction

The objective of this paper is to demonstrate how sensitivity analysis should be applied to risk assessment models that have two-dimensional probabilistic frameworks, and how the results from such an analysis should be presented and interpreted with a focus on microbial food safety process risk (MFSPR) models as a case study. Although sensitivity analysis has been applied to two-dimensional probabilistic frameworks in other scientific disciplines,[1, 2] we believe this is the first such systematic application to MFSPR models. In this paper particular attention is given to distinguishing between variability and uncertainty in sensitivity analysis and its implication in providing more robust results and confidence in insights regarding key model inputs.

The process of identifying key controllable inputs and key sources of uncertainty using sensitivity analysis is challenged by typical characteristics of MFSPR models such as nonlinearity, thresholds, interactions, and categorical inputs.[3] Commonly used sensitivity analysis methods, such as correlation analysis, can be misleading when applied to complex MFSPR models.

The following key methodological questions are addressed:

- How does uncertainty in model inputs affect the ranking of key sources of variability?
- Does co-mingling variability and uncertainty into a single dimension provide different insight regarding key sources of variability than two-dimensional simulation?
- How can one discriminate the importance between closely ranked inputs?
- What are the limitations of commonly used sensitivity analysis methods when applied to MFSPR models?

- What are the implications of sensitivity analysis for refinement of the analysis and for risk management?

Food safety risk assessment modeling is discussed. Motivations for sensitivity analysis of MFSPR models are explained. The rationale for focusing on a selected sensitivity analysis method and MFSPR model as a testbed is given.

1.1 Food Safety Risk Assessment Modeling

In the field of food safety, quantitative risk assessment is widely used to provide insight with respect to risk management strategies.[4, 5] There has been increasing development and use of MFSPR models for risk assessment of food-borne diseases.[6] Examples include *Salmonella* in eggs,[7] *E. coli O157:H7* in ground beef,[8] *Vibrio parahaemolytics* in raw Molluscan Shellfish,[9] and *Listeria monocytogenes* in ready-to-eat (RTE) foods.[10]

MFSPR models generally have probabilistic frameworks, are able to quantify uncertainties in the population risk, include many parameters and inputs that have uncertainty and/or variability, have complex or highly non-linear interactions between inputs, have threshold or saturation points in the model response, and use different types of inputs (e.g., continuous versus categorical). A threshold is a value in an input domain below which a model output does not respond to changes in the input. Similarly, a saturation point is an input value above which the model output does not respond. Parameters and inputs of MFSPR models are generally quantified by using probability distributions that represent variability, uncertainty, or both. Variability represents true heterogeneity in a population that cannot be further reduced. Uncertainty represents lack of perfect knowledge and can be reduced by further measurements.[11-17]

1.2 Need for Sensitivity Analysis

Sensitivity analysis is recognized as an essential component of analyses for complex models.[1] Sensitivity analysis highlights the inputs that have the greatest influence on the results of a model; therefore, it provides useful insights for model builders and users.[18] Sensitivity analysis provides insight regarding which model input contributes the most to uncertainty, variability, or both, for a particular model output. Knowledge of key sources of uncertainty is useful in prioritizing additional data collection or research.

Knowledge of key sources of variability is useful in identifying control measures. Sensitivity analysis is an aid in developing priorities for risk mitigation and management strategies. Sensitivity analysis can be used to prioritize potential critical control points (CCP) in MFSPR models and identifies corresponding critical limits. CCPs are points, steps, or procedures in the process of bringing food from farm to the table at which direct control can be applied, and a food safety hazard can be prevented, eliminated, or reduced to an acceptable level.[19, 20] A critical limit (CL) is a criterion that must be met for each preventive measure associated with a CCP.

1.3 Selection of the Sensitivity Analysis Method and the MFSPR Model for the Case Studies

A sensitivity analysis method applicable to MFSPR models should be able to: deal with simultaneous variation in inputs; address specific model characteristics such as non-linearity, interactions, and thresholds; handle alternative input types (continuous vs categorical); discriminate among sensitive inputs; and provide quantitative measures of sensitivity.[3] Frey et al. presented a comparison of the capabilities of many sensitivity analysis methods with regard to these characteristics.[21] A key conclusion was that methods conventionally used for sensitivity analysis and available in commercial statistical software packages (e.g., correlation analysis)

may not provide robust results when applied to MFSPR models. However, analysis of variance (ANOVA) was identified as a promising method that can deal with nonlinearity, interaction, thresholds, and both categorical and continuous inputs. Thus, ANOVA is selected for application to the case studies in this paper.

ANOVA identifies whether mean output values change with changing values for inputs.[22] ANOVA decomposes the total variability in the output values into the main effect of each input, the interactive effect of pair-wise combinations of inputs, and higher order interactions among inputs.[23] A main effect is the direct effect of an independent input on the output. An interaction effect is the joint effect of two or more independent inputs on the response variable.[24] ANOVA includes a capability to assess thresholds or saturation points in the output. ANOVA can handle alternative types of inputs such as continuous and categorical inputs.

The *E. coli O157:H7* food safety process risk model in ground beef is selected as a testbed.[8] *E. coli O157:H7* represents an important public policy issue. Ground beef servings contaminated with *E. coli O157:H7* organisms are estimated to be responsible for some 73,500 cases of infection, 2,150 hospitalizations, and 61 deaths in the United States each year.[25] The model has a two-dimensional probabilistic framework and typical characteristics of MFSPR models.

ANOVA is briefly described. The *E. coli* model is discussed and specific probabilistic analysis scenarios are defined. Results based on the application of ANOVA are presented. Uncertainty in ANOVA results is quantified. Comparisons are provided between results from ANOVA and those from more commonly used sample (Pearson) and rank (Spearman) correlation analyses.[26, 27] Finally, insights and findings from the analyses are discussed in answer to the key questions.

2.0 Overview of ANOVA

ANOVA is a model independent sensitivity analysis method that evaluates whether there is any statistically significant association between an output and one or more inputs. ANOVA does not assume any specific functional relationship between the output and inputs. However, ANOVA assumes that the response at each value of the inputs is normally distributed with the same variance.[28-32] Nonetheless, ANOVA is generally robust to deviations from this assumption.[33, 34]

An input is referred to as a "factor" and specific ranges of values for each factor are defined as factor "levels". A "treatment" is a specific combination of levels for different factors. An output is referred to as a "response variable," and a "contrast" is a linear combination of two or more factor level means.[22] For example, a contrast can be used to evaluate the change in the mean growth of pathogens when the storage temperature varies between high and low levels for a specific storage time. Contrasts typically provide insight regarding specific model characteristics such as threshold and saturation points.

Categorical factors are easily treated as levels. Continuous factors can be partitioned to mutually exclusive and exhaustive subintervals in order to create levels.[35] The optimal definition of levels for a continuous factor is often a matter of judgment and some experimentation may be required to determine an appropriate division.[21, 35]

ANOVA uses the F test to determine whether a significant difference exists among mean responses for main effects or interactions between factors. The relative magnitude of F values can be used to rank the factors in sensitivity analysis.[36] The higher the F value, the more sensitive the response variable is to the factor. Therefore, factors with higher F values are given higher rankings. An alternative but related definition of sensitivity can be introduced based on the partitioning of the total sum of squares (TSS) of the model response (Y) on all of the main

effects and interaction effects to estimate the relative partial sum of squares (*RPSS*).[37, 38] However, the F value is preferred as it accounts for not only the sum of squares but also the degrees of freedom associated with each factor.

3.0 Overview of the *E. coli* Food Safety Process Risk Model

The U.S. Department of Agriculture conducted a farm-to-table risk assessment to evaluate the public health impact from *E. coli O157:H7* in ground beef .[8] An overview of the model is given in Figure 1. The *E. coli O157:H7* risk assessment includes hazard identification, exposure assessment, hazard characterization, and risk characterization steps. The exposure assessment step consists of three major modules: (1) production; (2) slaughter; and (3) preparation. The preparation module consists of three parts: (1) growth estimation; (2) cooking effect; and (3) serving contamination. The exposure assessment is based upon a probabilistic approach for modeling the prevalence and the concentration of the *E. coli* pathogen in live cattle, carcasses, beef trim, and a single serving of cooked ground beef. The model is implemented in Microsoft Excel using @Risk as an add-in for defining probability distributions for inputs.

The *E. coli* model has a modular framework. Thus, each key activity is modeled in a separate module. For example, there are separate modules for estimation of infection of live cattle in farms and contamination of meat trimings in slaughter plants. Each module may have inputs that are calculated based upon a predecessor module and its outputs may be input to a successor module. In some cases, the output of a module is binned into predefined intervals prior to transferring to successor module. For example, in the slaughter module the contaminant concentration in beef trimings is estimated as a continuous output, but subsequently is binned by 0.5 log increments from 0 to 8 logs.[8] Because of binning, there is a loss of one-to-one correspondence between outputs of one module and the inputs of its predecessors. Therefore,

binning is an obstacle to global sensitivity analysis of the entire model. Creating one-to-one correspondence in this case would require substantial restructuring and recoding of the model. Thus, ANOVA is applied to the individual modules.

The growth estimation part in the preparation module is selected for the case studies. This part has a two-dimensional probabilistic simulation framework. Storage temperature and storage time at retail (Stage 1), transportation (Stage 2), and home (Stage 3) influence the *E. coli O157:H7* levels in ground beef servings. The output is the growth of *E. coli O157:H7* organisms in ground beef servings prior to cooking.

Table 1 summarizes 13 inputs including their mean values and 95 percent probability ranges based on the sampled values of inputs in a co-mingled analysis of variability and uncertainty. Co-mingled analysis of variability and uncertainty is explained in Section 4.2. Table 1 also gives the type of variability and uncertainty distributions for each input.[8] Uncertainty is incorporated in the parameters of the variability distributions. For example, storage temperature at Stage 1 has an empirical distribution based on the available data. In addition to modeling variability in storage temperature as an empirical distribution, uncertainty regarding the actual frequency of each temperature is modeled using a beta distribution as:[8]

$$CF_i = BetaINV\left(\sum_{k=1}^{i} n_k, n_T - \sum_{k=1}^{i} n_k + 1\right) \quad (1)$$

where,

CF_i	=	Cumulative frequency at value i of the empirical distribution
$BetaINV$	=	Inverse of a beta distribution
n_i	=	Number of available data at the i^{th} value of storage temperature
n_T	=	Total number of available survey data

$\sum_{k=1}^{i} n_k$ = Parameter α of the beta distribution

$n_T - \sum_{k=1}^{i} n_k + 1$ = Parameter β of the beta distribution

4.0 Probabilistic Analysis Scenarios

This section explains the probabilistic scenarios defined for sensitivity analysis. Two scenarios are introduced: (1) variability analysis for different realizations of uncertainty; and (2) co-mingled analysis of variability and uncertainty. Procedures for application of sensitivity analysis to these scenarios and specific insights with respect to sensitivity that can be obtained from each of these approaches are briefly discussed.

4.1 Variability Analysis for Different Uncertainty Realizations

In variability analysis for different uncertainty realizations, the objective of the analysis is to distinguish between variability and uncertainty. As illustrated in Figure 2, the focus of sensitivity analysis in this approach is to identify the key variability inputs for each realization of uncertainty.[1, 2] A realization refers to one model simulation based upon one randomly sampled uncertainty value for each probabilistic input. In this situation, sensitivity analysis provides insight regarding whether the identification of the key sources of variability is robust with respect to uncertainty. Each dataset in Figure 2 includes randomly generated values (e.g., *m* values) from variability distributions of each model input for a given uncertainty realization and the corresponding model output values. Sensitivity analysis is applied separately to each uncertainty realization. Thus, for each realization, the key sources of variability, critical limits, or both, are identified. This process is repeated *n* times to arrive at different sensitivity rankings for variability inputs, where *n* refers to the number of uncertainty realizations. For this probabilistic

scenario, values of 100 and 650 were selected for n and m, respectively, based upon the storage limit in each Excel sheet.

To the extent that the sensitivity analyses yield similar results about the rank ordering of key inputs regardless of uncertainty, an analyst or decision maker will have greater confidence that the results of the analysis are robust to uncertainty. If the ranking of key inputs changes substantially from one realization of uncertainty to another, the identification of key inputs would be uncertain. Additional data collection or research may reduce this ambiguity.

4.2 Co-mingled One-Dimensional Variability and Uncertainty Analysis

There may be situations in which it is preferred to incorporate but not distinguish variability and uncertainty. For example, during the process of model-building, the analyst may want to estimate the widest range of values that might be assigned to each model input for purposes of verifying the model and evaluating the robustness of the model to large perturbations in its inputs. As a preliminary step in prioritizing data collection or the development of distributions for model inputs, an analyst may wish to assess the key sources of variation regardless of whether they represent variability or uncertainty. In some cases, an analyst may make a judgment that it is difficult to separate variability from uncertainty [(6)], or that a two-dimensional probabilistic approach is impractical or not necessary based upon the assessment objectives.

Figure 3 illustrates sensitivity analysis in the context of co-mingled variability and uncertainty. There are n datasets generated in n uncertainty realizations of the model. Each dataset has m rows of randomly generated samples from variability distributions and the corresponding output values. Prior to application of sensitivity analysis, these datasets are appended together to form a single dataset with $n \times m$ rows. Sensitivity analysis is then applied to

the co-mingled dataset and a set of rankings is obtained. For the co-mingled analysis in the growth estimation part values of 100 and 650 were selected for n and m, respectively.

5.0 Application of ANOVA

This section presents results from application of ANOVA to each of the probabilistic scenarios defined in Section 4. For each probabilistic scenario, results from ANOVA are compared with results from Pearson and Spearman correlation analyses. Application of ANOVA to identify a saturation point in the model response and a methodology for quantifying ambiguity in ranks in a one-dimensional analysis is demonstrated. Ranking of the important factors based on ANOVA also is verified.

Inputs in the growth estimation part are continuous, and hence, must be partitioned into levels.[35] Frey et al. demonstrated three approaches for defining factor levels for continuous inputs based on: (1) evenly spaced intervals; (2) evenly spaced percentiles; and (3) visual inspection of the cumulative distribution function (CDF) for each input.[21] For the first approach, each input domain is classified into equal ranges. For the second approach, the CDF of the generated values for an input in a probabilistic simulation is used for defining factor levels at evenly spaced percentiles. For the third approach, boundaries for each factor level are defined corresponding to percentiles of the CDF that are associated with a substantial change in the slope. Figure 4 illustrates an example for defining levels based on the visualization of the CDF for maximum density. Three levels are defined for this factor based on values less than 6.5, between 6.5 and 8.5, and greater than 8.5 corresponding to 20^{th} and 80^{th} percentiles of the CDF.

A key consideration in defining levels is the number of data points within factor levels, which affects the power of statistical tests.[39] There is a trade-off between selecting a larger number of factor levels, which can produce more highly resolved insights regarding sensitivity

and getting statistically significant results. There is also a trade-off between the desired number of iterations (e.g., in a Monte Carlo simulation) that are used to populate factor levels and the computational time.

Table 2 summarizes the levels defined for factors in the case studies. These levels are used in both probabilistic scenarios. Levels are mostly defined based on visual inspection of the CDF for each factor. Each CDF is prepared based on the generated values for a factor in the co-mingled analysis of variability and uncertainty, since this probabilistic approach gives the widest range of variation for each factor. For storage times at Stages 1 and 3, levels are defined at equal intervals.

5.1 Variability Analysis for Different Uncertainty Realizations

Table 3 summarizes the results from variability analysis for different uncertainty realizations. Mean, minimum, and maximum estimated F values from all uncertainty realizations for each factor are given. The percentage of the uncertainty realizations that produced a statistically significant F value for each factor is quantified. Furthermore, factors are ranked based on the relative magnitude of F values. The mean rank and the range of ranks are presented. For each uncertainty realization, interactions between storage temperature and storage time at all three stages are evaluated.

The factors can be classified into four categories based on their importance in uncertainty realizations: (1) Group 1: the most important factor; (2) Group 2: factors with secondary importance with ranks between 2 and 4; (3) Group 3: factors with minor importance with ranks 5 and 6; and (4) Group 4: factors that are unimportant with ranks between 7 and 13. Factors in the latter group typically were not identified as statistically significant in most of uncertainty

realizations. Factors are classified into these four groups based on their dominant range of ranks in uncertainty realizations.

Figure 5 depicts the probabilities associated with selection of factors in each of the four importance groups in uncertainty realizations. Based on the dominant importance groups for factors, *Time3* is the most important factor as it is ranked 1 in 72 percent of uncertainty realizations. A group of three factors (i.e., *Temp3, Time1,* and *Temp1*) are within Group 2 with mean ranks between 2.7 and 3.1. Factors in Group 3 include *LP3* and *LP1* with mean ranks of 6.0 and 6.2, respectively. Group 4 contains seven factors including *GT1, GT3, GT2, Temp2, LP2, MD,* and *Time2* with mean ranks between 11.0 and 12.3.

The interaction effect between storage temperature and storage time at Stage 2 is unimportant. Since factors related to Stage 2 mostly are identified as unimportant, the transportation step does not require additional refinement. The interaction effect between storage time and storage temperature at home has higher importance than interaction between similar factors at retail stores. The mean F values estimated for these two interaction effects differ by a ratio of approximately 2. Although the interaction effects are not considered in the overall ranking in Table 3, their relative importance can be compared with main effects using their mean F values similar to an approach presented by Rose et al.[38] Factors contributing to the two statistically significant interaction effects have relatively large mean F values and are categorized in the first two importance groups. The interaction effect between *Time3* and *Temp3* has a mean F value approximately the same as that of *Temp3*. Thus, the interaction has comparable importance with *Temp3*.

Although the F values calculated for each main or interaction effect indicate the statistical significance of corresponding effect, the coefficient of determination (R^2) indicates whether the

selected effects adequately capture variability in the response variable, and hence, can be used as a diagnostic for the results of ANOVA.[22] Generally, a high R^2 value implies that results are not compromised by incomplete specification of effects or by inappropriate definition of the levels for a factor. Figure 6 shows CDF graphs for R^2 values estimated in different uncertainty realizations for two ANOVA models. For an ANOVA model based only on the main effects, the 95 percent probability range of R^2 values is between 0.45 and 0.72 with a mean value of 0.61. For an ANOVA model based on all of the main effects and selected interactions as shown in Table 3, the 95 percent probability range of R^2 values is between 0.77 and 0.92 with a mean value of 0.85. Therefore, interaction effects between storage time and storage temperature in Stages 1 to 3 accounted for approximately 24 percent of the total variability in the response variable. Because the ANOVA model with interaction terms accounts for a substantial amount of variability in the response, ranks based on the F values are deemed to be reliable.

Pearson and Spearman correlation analyses are selected as conventional sensitivity analysis methods for comparison with ANOVA. Inputs are ranked based on the absolute values of statistically significant correlation coefficients. Table 4 summarizes mean ranks, range of ranks, and the percentage of the uncertainty realizations that produced a statistically significant correlation coefficient. These methods are not able to provide insight with respect to interactions between inputs. Similar to ANOVA results, inputs can be grouped into four categories based on their importance. The probabilities associated with selection of inputs in each of the four importance groups are shown in Figure 7. Inputs within these categories are typically different from those based on the ANOVA results. Some sensitive inputs (e.g., *GT3* and *LP3*) via both Pearson and Spearman correlation analyses were not sensitive based on ANOVA. Among statistically significant inputs, there is not a clear agreement between the ranks based on the

ANOVA results and those from Pearson and Spearman correlation analyses. For example, while ANOVA selected *Time3* as the most important factor, *Temp3* and *Time1* are selected as the most important inputs based on Pearson and Spearman correlation analyses, respectively. However, similar to ANOVA, Pearson and Spearman correlation analyses identified inputs related to Stage 2 as unimportant.

5.2 Co-Mingled Analysis of Variability and Uncertainty

Table 5 summarizes results when variability and uncertainty are co-mingled into one dimension. The main effects of all factors and interaction effects between storage temperature and storage time at all three stages are evaluated. For each factor the number of times that two-dimensional sensitivity analysis results provided the same ranking as the co-mingled analysis is quantified. The number of times that each factor fell into the same importance group based on the two-dimensional analysis results also are quantified. Table 5 also includes rankings based on Pearson and Spearman correlation analyses.

Typically, the F values for main and interaction effects are substantially larger than those given in Table 3 for the two-dimensional analysis. Substantially larger sample sizes within levels defined for factors in the co-mingled analysis resulted in larger F values implying higher confidence on statistically significant factors and interaction effects.

Although co-mingled analysis typically provides different ranks compared to the two-dimensional analysis, these two probabilistic approaches provide similar results with respect to identification of groups of factors with comparable importance. Except for the most important factor (*Time3*), there is typically low probability that rankings based on the co-mingled analysis match those from two-dimensional analysis. For example, *Temp1* has rank of two in the co-mingled analysis. The probability that *Temp1* has the same rank in the two-dimensional analysis

is only 17 percent. In contrast, factors are categorized into the same importance group as those for the two-dimensional analysis with substantially high probability. For example, there is 85 percent probability from the two-dimensional analysis that *Temp1* fits in the group of factors with secondary importance, thus implying that in 85 percent of uncertainty realizations, *Temp1* has a rank between 2 and 4.

There is a statistically significant interaction between storage time and storage temperature at Stages 1 and 3. The relative magnitude of F values indicates that interaction effect between these two factors at Stage 3 has higher importance. Similar to the two-dimensional analysis, there is no statistically significant interaction between storage time and storage temperature at Stage 2.

Rankings based on Pearson and Spearman correlation analyses typically are different from ANOVA results, except for inputs related to Stage 2. All three methods identified no statistically significant effects for inputs related to Stage 2. Groups of inputs with comparable importance based on Pearson and Spearman correlation analysis results are different from those identified by ANOVA. For example, the most important inputs based on Pearson and Spearman correlation analysis are *Temp3* and *Time1*, respectively, which are different from *Time3* selected by ANOVA. Similarly, there is lack of agreement for inputs selected in secondary and minor importance groups based on these methods.

When performing one-dimensional probabilistic analysis, a key question is regarding how large the differences in F values between two factors must be in order to clearly discriminate their importance. A method for answering this question is illustrated. The range of sampling distributions of F values provides insight regarding how large the ratio of two F values must be in order for the ranks of the corresponding factors to be substantially different.

Empirical bootstrap simulation is used to generate sampling distributions of uncertainty for F values.[40] In the empirical bootstrap approach, an alternative randomized version of the original Monte Carlo simulation is obtained by sampling with replacement from the original set of random values. ANOVA is applied to each of the bootstrap samples to produce a distribution of F values. Summary of the results from 200 bootstrap simulations are given in Table 6. The summary table is comprised of mean F values, 95% probability range of F values, and coefficients of variation. The percentage of the bootstrap simulations that produced a statistically significant F value for each factor is quantified. The arithmetic mean of 200 ranks and the range of ranks for each factor also are quantified.

There is a substantial range of uncertainty associated with the estimates of F value. For example, *Temp1* is estimated to have a mean rank of 2.9. The F value for this factor has a mean of 1445 and 95 percent probability range of 873 to 2040, or approximately plus or minus 40 percent of the mean. *Temp3* has a mean rank of 3.7, a mean F value of 1130, and a 95 percent probability range of 650 to 1670, or approximately plus or minus 50 percent. The sampling distributions of the F values for these two factors do not have statistically significant correlation. Thus, the overlap in the confidence intervals of *Temp1* and *Temp3* indicates a possibility that the rank order between these two factors can reverse, even though on average the F value for *Time1* is larger than that for *Temp3* by a factor of 1.3. This reversal in ranks happened in 32 percent of bootstrap simulations. In contrast, *Temp3* has a statistically significantly larger F value than the factor with the fifth highest average rank, which is *LP3*. The 95 percent probability ranges for F values of these two factors do not overlap. Therefore, although there is some ambiguity regarding which of the factors may be the third most important; it is clear that *Temp1,* and *Temp3* are more important than *LP3*.

The coefficient of variation is approximately 0.30 or less for F values greater than approximately 200. For F values smaller than approximately 20, the coefficient of variation ranges from approximately 0.3 to 1.85. These results suggest that the coefficient of variation may be relatively constant for F values that are statistically significant and substantially large. Since a 95 percent probability range might typically be enclosed by plus or minus two standard deviations for a symmetric sampling distribution, one might infer that differences in F values of approximately 60 percent or more imply a clear discrimination in rank order. This value is specific to the example provided here and should not be used to make general quantitative judgments regarding differences between F values obtained with different sample sizes or models. However, the technique used here can be applied to other case studies. The case study results suggest that two factors having similar F values are typically of comparable importance.

ANOVA provides insight regarding special model characteristics such as non-linearity, thresholds and saturation points, and interactions by comparison of mean output values at different factor levels. Table 7 summarizes an example in which a set of contrasts is used to infer an interaction effect between storage temperature and storage time at Stage 3 and possible saturation points. When storage temperature in Stage 3 is at the first level (i.e., between 7.5^{o}C and 11^{o}C), storage time influences growth of *E. coli* in ground beef until the tenth day, since statistically significant contrasts are estimated for the difference in pathogen growth for consecutive days through the tenth. After the tenth day there is no significant difference in the estimated growth, indicating that maximum population density is achieved by the tenth day. When storage temperature is at the second level (i.e., between 11^{o}C and 14.5^{o}C), the contrasts indicate that maximum population density is reached in only five days. When the storage temperature increases to the third, fourth and fifth levels, maximum population density is

reached in four, three, and two days, respectively. This pattern indicates an interaction effect between storage time and temperature and the corresponding saturation points for growth.

Proper definition of factor levels for storage time or storage temperature can help identifying thresholds with respect to these factors which can be directly used as CLs in the risk management recommendations. For example, knowing that growth is zero in the beginning of the storage process, one may look for two consecutive time intervals that growth starts to show statistically significant difference following precursor time intervals with no significant difference in growth. The storage time level at which growth starts to show statistically significant change is indeed a threshold. Using that cut off as a recommended storage time can ensure prevention of growth of the *E. coli O157:H7* organisms in ground beef servings. Results in Table 7 do not show this trend as storage time intervals are substantially large (i.e., one day) and growth does happen within a day. When storage temperature is at the first level (i.e., between 7.5 and 11^{o}C), there is statistically significant difference in growth between the first and the second day. This implies that perhaps refining factor levels for storage time into intervals with a smaller period (e.g., half a day or 6 hours) can help identifying such a trend in growth. However, this refinement is not done for the case studies due to the lack of sufficient sample size.

5.3 Verification of ANOVA Results

Comparison of ANOVA and correlation analysis typically showed different insights about the inputs that matter the most. A key question is regarding which of these methods provides more appropriate and accurate insights with respect to the ranks of inputs. An approach based on the conditional analysis of the model is used for verification of the ANOVA results.[41] The verification is done for the ANOVA results based on the co-mingled analysis of variability and uncertainty.

Conditional complementary cumulative distribution functions (CCDF) of the output are used to verify the ANOVA results. In order to use CCDFs, a subset of important factors (i.e., factors in the first three importance groups) was selected. The validity of rankings was studied by holding the selected factors at their mean values and performing *N* realizations. If by holding one or more factors at their mean values there is a decrease in the spread of CCDF graph, then it can be concluded that selected factors are in fact important. One would expect a systematic decrease in the spread of the CCDF graph when more important factors are conditioned at their mean values.

Figure 8 presents the conditional CCDF graphs for two cases: (a) based on conditioning the top six important factors using ANOVA results; (b) based on conditioning the top six important inputs using Pearson correlation analysis results. In each case, "0 fixed" represents the situation in which all inputs are varied based on their probability distributions, while "6 fixed" represents the situation in which all inputs except those with ranks 1 to 6 are varied based on their probability distributions. For the latter case, the first six inputs are conditioned at their mean values. Similar cases are shown when the number of "fixed" inputs varies from 1 to 5. Figure 8(a) shows a systematic decrease in the spread of the CCDF when each additional important factor is held at its mean value. This observation indicates that ANOVA identified the top six important factors correctly in the co-mingled analysis of variability and uncertainty. In contrast, Figure 8(b) shows that conditioning on mean values of *GT3* and *LP3* does not substantially decrease the spread of the CCDF graph. The spreads of CCDFs when conditioning *GT3* and *LP3* to their mean values are approximately 0.78 and 0.76 logs compared to 0.81 logs for conditioning *Time3*. Although *Temp1* and *Time1* are ranked as 5 and 6 based on Pearson correlation analysis, respectively, contribution of these inputs to the change in the spread of

CCDF is substantially larger than that of *GT3* and *LP3*. The spreads of CCDF when conditioning *Temp1* and *Time1* following the top four inputs at their mean values are approximately 0.58 and 0.22 logs, respectively. Comparison of the conditional CCDFs in Figures 8(a) and 8(b) for the "1 fixed" cases indicates that the most important input selected by ANOVA (*Time3*) has more influence on the change in the spread of CCDF than the one selected based on Pearson correlation analysis (*Temp3*).

Similar assessment is done but not shown here based on Spearman correlation analysis results. Conditioning the inputs at their mean values following the order they are ranked using Spearman correlation analysis also did not show systematic decrease in the spread of CCDFs. Thus, ANOVA provided more appropriate and accurate insights with respect to the ranks of inputs.

6.0 Discussion and Conclusions

This paper has demonstrated the capabilities of ANOVA as a sensitivity analysis method for application to MFSPR models. ANOVA showed as a promising sensitivity analysis method considering its abilities to address typical characteristics of MFSPR models such as thresholds and saturation points, non-linearity, and interactions. Key methodological questions raised in the introduction are addressed in the following.

The importance of an input can vary from substantially high to low due to uncertainty in inputs in a two-dimensional probabilistic framework. Some inputs were consistently ranked high, whereas others were consistently ranked low. The information about range of ranks and associated probabilities is useful to a decision maker. By quantifying uncertainty in the rank of each input, groups of model inputs with similar importance are identified. For the four groups of model inputs identified, inputs have a clear difference with respect to sensitivity in each group

compared to those in other groups. Although there is a wide range of uncertainty associated with the rank of each input based on ANOVA, the importance group associated with each of the inputs is substantially robust to uncertainty. For example, *Temp1* has a wide range of uncertainty in ranks between 1 and 8. However, a decision maker can confidently conclude that this input is of secondary importance as *Temp1* was selected in Group 2 in 85 percent of uncertainty realizations.

A comparison of one-dimensional and two-dimensional results implied that the top ranked input was the same for each of the probabilistic simulation approaches. Furthermore, these two approaches identified approximately the same groups of inputs with comparable importance. Thus, it may be possible to develop a robust insight regarding key inputs using a simplified one-dimensional probabilistic simulation approach if resources do not permit a two dimensional approach. Alternatively, as a part of model development, sensitivity analysis can initially be performed in a one-dimensional analysis to identify priorities for data collection due to the computationally intensiveness of a two-dimensional sensitivity analysis. However, although the results obtained in the specific case studies in this paper may be indicative, they are not definitive. Thus, where resources permit, a two-dimensional sensitivity analysis is generally preferred unless a one-dimensional analysis is appropriately justified.

Significant differences between F values indicate that ranks associated with the corresponding factors are clearly discriminated. In contrast, there is comparable sensitivity to factors that have similar F values. Uncertainty in point estimates of F values from one-dimensional simulations should be considered when making judgment regarding sensitivity with comparisons of the F values of two or more factors. Bootstrap simulation was demonstrated as an approach for quantifying uncertainty in F values based upon re-sampling from a one-dimensional

simulation. For the case scenarios presented in this paper and selected sample size, the range of uncertainty in statistically significant F values that were substantially large was found to be approximately plus or minus 60 percent or less. This implies that differences in F values of 60 percent or more were associated with clear differences in rank order between factors.

Insights from Pearson and Spearman correlation analyses typically differed from those of ANOVA. Verification of the relative importance of sensitive inputs obtained based upon ANOVA and Pearson correlation was done by comparison of model results when sensitive inputs are set to their mean values in a progression. The rankings based on ANOVA were found to be more reasonable than those based on correlation analysis. Pearson and Spearman correlation analysis methods are typically built-in features of commonly used software tools such as @Risk^{TM} and Crystal Ball^{TM} and are referred to here as "push-button" methods because of their ease of implementation. Under pressure of deadlines and faced with constraints of an existing model, an analyst may prefer to use these methods rather than attempt to implement more refined techniques such as ANOVA. However, a key question is whether such methods are adequate for the intended case study. If an analyst chooses to proceed with a push-button method whose theoretical basis differs from the key characteristic of the model to which it is applied, then the accuracy and robustness of the results cannot be assured. In the long-run, it will be helpful to the food safety risk community if add-ins can be developed, either as shareware or commercially, that incorporate sensitivity analysis methods such as ANOVA that are more compatible with the characteristics of MFSPR models.

Insights from sensitivity analysis implied the importance of the inputs related to Stage 3 (i.e., home) with respect to their influence on the growth of *E. coli O157:H7* organisms. Identifying practically controllable inputs such as storage time and storage temperature based on

the analysis results suggests the possible risk management strategies. However, applying direct control on the consumer side cannot be deemed as an effective risk management approach because of the possible lack of compliance. Increasing a consumer's knowledge and awareness regarding risk free ground beef handling and storage practices via consumer advisory programs can effectively decrease the risk of exposure to *E. coli* organisms. Ground beef is highly perishable and must be handled and stored properly to avoid spoilage. Insights from sensitivity analysis indicate that even in a refrigeration temperature around 7^{o}C, ground beef servings can reach the saturation point in growth within few days. Thus, if ground beef servings are not going to be consumed within few days, they should be frozen as growth does not happen at freezing temperature. Defrosting ground beef servings should be done in the refrigerator the day before cooking rather than at room temperature as insights from sensitivity analysis showed significant interaction between storage temperature and storage time. High values of growth are possible within a short amount of time at high temperature. Sensitivity analysis results showed that growth reaches the saturation point in only a day when ground beef servings are held at room temperature (21.5-25^{o}C). This could even happen in a fraction of a day in higher temperatures.

Although the presented case studies focused on identification of key sources of variability followed by quantification of uncertainty in the analysis results, a probabilistic scenario can be arranged in order to identify key contributors to uncertainty in the model output conditional on different values for quantities that are subject to variability. Results of such an analysis will provide insight regarding how key sources of uncertainty for exposure or risk may change for different combinations of values for variable inputs. For example, the key sources of uncertainty may be different in 90th percentile of the output variability distribution compared to those for 50th percentile. If the ranking of the key sources of uncertainty substantially changes for different

variability percentiles, decision-makers could focus on the key sources of uncertainty for the percentiles of most interest. For example, key sources of uncertainty can be identified for the upper 5 percent of the population exposed to the risk. Knowledge of key sources of uncertainty for the most exposed or at risk portion of the population can be used to prioritize additional data collection or research that could reduce uncertainty. The assessment can be revised based upon new information, and a decision could be made at a later time based upon the reduced uncertainties.

7.0 Acknowledgment

The work reported here is based on work supported by Cooperative Agreement No. 58-0111-0-005 between the U.S. Department of Agriculture, Office of the Chief Economist, Office of the Risk Assessment and Cost Benefit Analysis (USDA/OCE/ORACBA) and the Department of Civil, Construction, and Environmental Engineering at North Carolina State University. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Agriculture.

8.0 References

1. Helton, J. C. 1997. Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty. *Journal of Statistical Computation and Simulation, 57(1), 3-76.*
2. Frey, H. C., & Rhodes, D. S. (1996). Characterizing, simulating, and analyzing variability and uncertainty: An illustration of methods using an air toxics emissions example. *Human and Ecological Risk Assessment, 2(4), 762-797.*
3. Frey, H. C. (2002). Introduction to special section on sensitivity analysis and summary of NCSU/USDA workshop on sensitivity analysis. *Risk Analysis, 22(3), 539-545.*
4. Rose, J. B., & Sobsey, M. D. (1993). Quantitative risk assessment for viral contamination of shellfish and coastal waters. *Journal of Food Protection, 56(12), 1043-1050.*
5. Jaykus L. A., (1996). The application of quantitative risk assessment to microbial food safety risks. *Critical Reviews in Microbiology, 22, 279-93.*
6. Nauta, M. J. (2000). Separation of uncertainty and variability in quantitative microbial risk assessment models. *International Journal of Food Microbiology, 57(1), 9-18.*
7. SERAT, (1998). *Salmonella Enteritidis risk assessment for shell eggs and egg products: Final report.* Prepared by the Salmonella Enteritidis Risk Assessment Team (SERAT) for the Food Safety and Inspection Service (FSIS), U.S. Department of Agriculture, Washington, D.C.
8. FSIS, (2001). *Draft risk assessment of the public health impact of Escherichia Coli O157:H7 in ground beef.* Prepared for the Food Safety and Inspection Service and United States Department of Agriculture, Washington, DC.
9. CFSAN, (2001). *Draft risk assessment on the public health impact of Vibrio parahaemolyticus in raw molluscan shellfish*. Prepared by the *Vibrio parahaemolyticus*

Risk Assessment Task Force (Center for Food Safety and Applied Nutrition) for Food and Drug Administration (FDA), USDA, Washington, DC.

10. CFSAN, (2003). Quantitative assessment of relative risk to public health from foodborne Listeria monocytogenes among selected categories of ready-to-eat foods. Prepared by Center for Food Safety and Nutrition, Food and Drugs Administration, USDA, Washington, DC. http://www.foodsafety.gov/~dms/lmr2-toc.html

11. Chernoff, H., & Moses, L. E. (1959). *Elementary Decision Theory.* New York, NY: John Wiley and Sons.

12. Kaplan, S., & Garrick, B. J. (1981). On the quantitative definition of risk. *Risk Analysis, 1(1), 11-27.*

13. Pate-Cornell, M. E. (2002). Risk and uncertainty analysis in government safety decisions. *Risk analysis, 22(3), 633-646.*

14. Apostolakis, G. E. (1989). Uncertainty in probabilistic risk assessment. *Nuclear Engineering and Design, 115, 173-179.*

15. Murphy, B. L. (1998). Dealing with uncertainty in risk assessment. *Human and Ecological Risk Assessment, 4(3), 685-699.*

16. Anderson , E. L., & Hattis D. (1999). Uncertainty and variability. *Risk Analysis, 19(1), 47-49.*

17. Cullen, A. C., & Frey, H. C. (1999). *Probabilistic Techniques in Exposure Assessment.* Plenum Press: New York.

18. McCarthy, M. A., Burgman, M. A., & Ferson, S. (1995). Sensitivity analysis for models of population viability. *Biological Conservation, 73(1), 93-100.*

19. Hulebak, K. L, & Schlosser, W. (2002). Hazard analysis and critical control point (HACCP) history and conceptual overview. *Risk Analysis, 22(3), 553-578.*

20. Seward, S. (2000). Application of HACCP in food service. Irish Journal of Agricultural and Food Research, 39(2), 221-227.

21. Frey, H. C., Mokhtari, A., & Zheng, J. (2004). Recommended practice regarding selection, application, and interpretation of sensitivity analysis methods applied to food safety process risk models. Prepared by North Carolina State University for Office of Risk Assessment and Cost-Benefit Analysis, U.S. Department of Agriculture, Washington, DC. www.ce.ncsu.edu/risk.

22. Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied Linear Statistical Models*. Fourth Edition. McGraw-Hill: Chicago, IL.

23. Krishnaiah, P. R (1981). *Analysis of Variance*. Elsevier: New York.

24. Rutherford, A. (2001). *Introducing ANOVA and ANCOVA: A GLM Approach.* Sage Publications: Thousand Oaks, CA.

25. Mead, P. S., Laurence, S., Vance, D., McCaig, L. F., Bresee, J. S., Shapiro, C., Griffin, P. M., & Tauxe, R. V. (1999). Food-related illness and death in the United States. *Emerging Infectious Diseases, 5(5), 607-625.*

26. Edwards, A. L. (1976). *An Introduction to Linear Regression and Correlation.* W. H. Freeman: San Francisco, CA.

27. Gibbons, J. D. (1985). *Nonparametric Statistical Inference,* 2nd Edition. Marcel Dekker: New York.

28. Edwards, A. F. (1979). Multiple Regression and the Analysis of Variance and Covariance. W. H. Freeman: San Francisco, CA.

29. Hicks, C. R. (1973). *Fundamental Concepts in the Design of Experiments.* Holt, Rinehart, and Winston: New York.

30. Krishnaiah, P. R (1981). *Analysis of Variance*. Elsevier: New York.

31. Peterson, R. G. (1986). *Design and Analysis of Experiments,* New York: Marcel Dekker.

32. Scheffé, H. (1959). *The Analysis of Variance.* New York: John Wiley and Sons.

33. Lindman, H. R (1974). *Analysis of Variance in Complex Experimental Designs.* W. H. Freeman & Co.: San Francisco, CA.

34. Conover, W. J. (1980). *Practical Non-Parametric Statistics.* 2nd Edition. New York: John Wiley and Sons Inc.

35. Kleijnen, J. P., & Helton, J. C. (1999). Statistical analyses of scatterplots to identify important factors in large-scale simulations, 1: Review and comparison of techniques. *Reliability Engineering and System Safety, 65(2), 147-185.*

36. Carlucci, A., Napilitano, F., Girolami, A., & Monteleone, E. (1999). Methodological approach to evaluate the effects of age at slaughter and storage temperature and time on sensory profile of lamb meat. *Meat Science, 52(4), 391-395.*

37. Gardner, R. H., & Trabalka, J.R. (1985). *Methods of uncertainty analysis for a global carbon dioxide model.* DOE/OR/21400-4, US Department of Energy, Washington, DC.

38. Rose, K. A., Smith, E. P., Gardner, R. H., Brenkert, A. L., & Bartell, S. M. (1991). Parameter sensitivities, Monte Carlo filtering, and model forecasting under uncertainty. *Journal of Forecasting, 10(1), 117-133.*

39. Giesbrecht, F., & Gumpertz, M. (1996). Planning, Construction, and Statistical Analysis of Comparative Experiments. John Wiley and Sons: New York.

40. Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* Chapman and Hall, New York.

41. Lu, Y. & Mohanty, S. (2002). Sensitivity analysis of a complex, proposed geologic waste disposal system using the Fourier Amplitude Sensitivity Test Method. *Journal of Reliability Engineering and System Safety, 72(3), 275-291.*

Table 1. Input Variables and Corresponding Mean Values, 95% Probability Ranges, and Distribution Types in the Growth Estimation Part of the *E. coli* Model (FSIS, 2001)

Factor	Mean Value (1)	95% Probability Range (1)	Distribution Type		Unit
			Variability	Uncertainty (2)	
Storage Temperature, Retail (Temp1)	8.7	(7.7, 13.1)	EDF (3)	Beta	^{o}C
Storage Temperature, Transportation (Temp2)	9.4	(7.7, 15.5)	EDF (3)	Beta	^{o}C
Storage Temperature, Home (Temp3)	9.1	(7.7, 19)	EDF (3)	Beta	^{o}C
Storage Time, Retail (Time1)	23.4	(0.5, 92.6)	Exponential	Uniform	hr
Storage Time, Transportation (Time2)	1.0	(0, 2.0)	EDF (3)	Beta	hr
Storage Time, Home (Time3)	25.2	(0.6, 99.0)	Exponential	Uniform	hr
Maximum Density (MD)	7.5	(5.4, 9.6)	Triangle (4)	Uniform	log
Lag Period, Retail (LP1)	73.5	(19.6, 136.1)	Exponential	Normal	hr
Lag Period, Transportation (LP2)	64.9	(12.5, 131.5)	Exponential	Normal	hr
Lag Period, Home (LP3)	72.0	(8.2, 137.9)	Exponential	Normal	hr
Generation Time, Retail (GT1)	9.9	(2.8, 15.9)	Exponential	Normal	hr
Generation Time, Transportation (GT2)	8.7	(1.9, 15.5)	Exponential	Normal	hr
Generation Time, Home (GT3)	9.8	(1.3, 16.0)	Exponential	Normal	hr

(1) Values are based on the co-mingled analysis of variability and uncertainty
(2) Uncertainty distributions are defined for parameters of variability distributions
(3) Empirical distribution function based on FSIS data. Beta distribution is used to define the corresponding cumulative frequency at each temperature.
(4) Uncertainty is defined for the most likely value of the triangular distribution.

Table 2. Levels Defined for Factors in the Growth Estimation Part of the *E. coli* Model

Factor [(1)]	**Number of levels**	**Levels and Corresponding Percentiles** [(2)]
Temp1	5	7.5-11, 11-14.5, 14.5-18, 18-21.5, >21.5 [(3)]
Temp2	3	7.5-13.5, 13.5-19.5, >19.5 [(3)]
Temp3	5	7.5-11, 11-14.5, 14.5-18, 18-21.5, >21.5 [(3)]
Time1	12	0-24, 24-48, …, 264-288, >288 [(3)]
Time2	2	0-3.5, >3.5
Time3	12	0-24, 24-48, …, 264-288, >288 [(3)]
MD	3	{<6.5,6.5-8.5, >8.5} {20^{th}, 80^{th}} Percentiles
LP1	4	{<50, 50-65, 65-95, >95} {20^{th}, 50^{th}, 80^{th}} Percentiles
LP2	4	{<35, 35-55, 55-90, >90} {20^{th}, 50^{th}, 80^{th}} Percentiles
LP3	4	{<45, 45-65, 65-95, >95} {20^{th}, 50^{th}, 80^{th}} Percentiles
GT1	4	{<7, 7-9.5, 9.5-12.5, >12.5} {20^{th}, 50^{th}, 80^{th}} Percentiles
GT2	4	{<4.5, 4.5-8, 8-12, >12} {20^{th}, 50^{th}, 80^{th}} Percentiles
GT3	4	{<6.5, 6.5-9.5, 9.5-13, >13} {20^{th}, 50^{th}, 80^{th}} Percentiles

(1) The abbreviations used for factors in this table are the same as those defined in Table 1.
(2) The ranges that define each factor level and the percentiles of the CDF corresponding to the break point between factor levels are given.
(3) For this factor equal intervals are used as levels.

Table 3. Summary of the ANOVA Results for Two-Dimensional Variability Simulation for Different Uncertainty Realizations

Factor [(1)]	**Mean F Value**	**[*min, max*] Range of F values**	**Frequency** [(2)]	**Mean Rank** [(3)]	**Range of Rank**
Temp1	103	(2.0, 1233)	98	3.1	1-8
Temp2	1.0	(0.0, 15)	10	11.2	5-13
Temp3	119	(16.9, 689)	100	2.7	1-4
Time1	99	(6.8, 495)	100	2.8	1-4
Time2	0.3	(0.1, 2.7)	0	12.3	10-13
Time3	188	(32, 1449)	100	1.5	1-4
MD	1.1	(0.1, 5.9)	6	11.7	5-13
LP1	4.8	(0.6, 18)	77	6.2	4-11
LP2	1.1	(0.1, 7.4)	6	11.8	5-12
LP3	4.6	(0.3, 14)	87	6.0	4-12
GT1	2.0	(0.1, 47)	14	11.0	5-13
GT2	1.0	(0.1, 4.4)	10	11.5	5-13
GT3	1.4	(0.1, 9.4)	11	11.1	5-13
Time1×Temp1	66	(10.5, 340)	100		
Time2×Temp2	0.7	(0.1, 4.5)	2		
Time3×Temp3	123	(46, 840)	100		

(1) The abbreviations used for factors in this table are the same as those defined in Table 1.
(2) The percentage of the uncertainty realizations for which the F values were statistically significant.
(3) Arithmetic average of 100 ranks for each input.

Table 4. Summary of Sample (Pearson) and Rank (Spearman) Correlation Analyses Results for Two-Dimensional Variability Simulation for Different Uncertainty Realizations

Inputs [1]	Sample (Pearson) Correlation Analysis			Rank (Spearman) Correlation Analysis		
	Frequency [2]	Mean Rank	Range of Rank	Frequency [2]	Mean Rank	Range of Rank
Temp1	93	4.4	1-12	97	6.9	3-12
Temp2	2	11.0	8-13	7	11.1	9-13
Temp3	100	1.8	1-6	100	5.1	3-8
Time1	91	4.4	1-10	100	1.7	1-5
Time2	2	10.7	7-13	6	10.9	7-13
Time3	100	3.8	1-8	100	2.0	1-6
MD	3	10.8	9-13	1	10.9	9-13
LP1	81	7.0	3-13	99	4.8	2-8
LP2	5	10.5	6-13	8	10.8	7-13
LP3	99	4.9	2-111	100	3.3	2-8
GT1	80	6.8	2-13	94	7.3	4-12
GT2	1	11.3	9-13	4	11.2	10-13
GT3	99	3.7	2-9	100	5.2	3-8

(1) The abbreviations used for factors in this table are the same as those defined in Table 1.
(2) The percentage of the uncertainty realizations for which the F values were statistically significant.

Table 5. Summary of the ANOVA Results for Co-Mingled Analysis of Variability and Uncertainty ($R^2 = 0.82$)

Factor[(1)]	F Value	Rank	Equal Rank in 2D (%) [(2)]	Group in 2D [(3)]	Equal Group in 2D (%) [(3)]	Rank SCA[(4)]	Rank RCA [(5)]
Temp1	1,130	2	17	2	85	5	7
Temp2	0.1	13	7	4	91	13	12
Temp3	1,035	3	25	2	97	1	4
Time1	550	4	31	2	86	6	1
Time2	0.3	12	7	4	100	12	13
Time3	5,350	1	72	1	72	2	2
MD	34	9	11	4	92	9	13
LP1	261	5	38	3	66	7	6
LP2	0.7	11	10	4	95	10	11
LP3	200	6	32	3	73	4	3
GT1	52	7	16	4	84	8	8
GT2	1.0	10	17	4	85	11	10
GT3	45	8	12	4	92	3	5
Temp1 × Time1	1,190						
Temp2 × Time2	0.04						
Temp3 × Time3	2,270						

(1) The abbreviations used for factors in this table are the same as those defined in Table 1.
(2) Percent of times that rank in 2D analysis equals the rank based on co-mingled analysis.
(3) Group 1: most important input (i.e., rank 1); Group 2: secondary importance inputs (i.e., ranks between 2 and 4); Group 3: minor importance inputs (i.e., ranks between 5 and 6); Group 4: unimportant inputs (i.e., ranks between 7 and 13).
(4) Rank based on sample (Pearson) correlation analysis.
(5) Rank based on rank (Spearman) correlation analysis.

Table 6. Summary of the ANOVA Results for 200 Bootstrap Simulations for F value Statistics

Factor [(1)]	**Mean F Value**	**95% Probability Range**	**SD/Mean** [(2)]	**Frequency**[(3)] **(%)**	**Mean Rank**	**Range of Rank**
Temp1	1445	(873, 2040)	0.21	100	2.9	2-4
Temp2	2.7	(0.3, 7.8)	0.83	33	10.9	10-13
Temp3	1130	(650, 1670)	0.32	100	3.7	2-4
Time1	1550	(1275, 1820)	0.10	100	2.5	2-4
Time2	0.7	(0.0, 3.6)	1.75	2	12.5	10-13
Time3	6060	(5250, 6860)	0.05	100	1.0	1 [(4)]
MD	29	(14, 45)	0.41	100	8.9	8-9
LP1	273	(236, 315)	0.07	100	5.8	5-6
LP2	1.7	(0.3, 4.3)	0.68	22	11.4	10-13
LP3	325	(295, 379)	0.07	100	5.0	5-6
GT1	75	(52, 97)	0.26	100	7.0	7-8
GT2	1.9	(0.3, 5.3)	0.70	25	11.3	10-13
GT3	43	(27, 55)	0.29	100	8.1	7-9
Temp1 × Time1	1920	(1660, 2150)	0.07	100		
Temp2 × Time2	0.7	(0.3, 3.5)	1.85	2		
Temp3 × Time3	3605	(3270, 4060)	0.06	100		

(1) The abbreviations used for factors in this table are the same as those defined in Table 1.

(2) The ratio of standard deviation (SD) to mean F values for each factor (also referred to as coefficient of variation).

(3) The percentage of the bootstrap simulations for which the F values were statistically significant

Time3 consistently had rank 1 in bootstrap simulations.

Table 7. Evaluation of ANOVA Contrasts for the Growth Estimation Regarding the Interactions Between Storage Temperature and Storage Time at Stage 3

Contrast	Estimate (1)	F Value	Pr>F	Significant(2)
T [7.5-11°C], Time 1st and 2nd days	0.005	112	<0.0001	Yes
T [7.5-11°C], Time 2nd and 3rd days	0.026	1280	<0.0001	Yes
T [7.5-11°C], Time 3rd and 4th days	0.049	1720	<0.0001	Yes
T [7.5-11°C], Time 4th and 5th days	0.069	1280	<0.0001	Yes
T [7.5-11°C], Time 5th and 6th days	0.074	610	<0.0001	Yes
T [7.5-11°C], Time 6th and 7th days	0.103	455	<0.0001	Yes
T [7.5-11°C], Time 7th and 8th days	0.042	30.8	<0.0001	Yes
T [7.5-11°C], Time 8th and 9th days	0.031	6.8	0.008	Yes
T [7.5-11°C], Time 9th and 10th days	0.108	38.8	<0.0001	Yes
T [7.5-11°C], Time 10th and 11th days	------	0.16	0.8	No
T [11-14.5°C], Time 1st and 2nd days	0.116	4060	<0.0001	Yes
T [11-14.5°C], Time 2nd and 3rd days	0.211	5290	<0.0001	Yes
T [11-14.5°C], Time 3rd and 4th days	0.218	2170	<0.0001	Yes
T [11-14.5°C], Time 4th and 5th days	0.119	240	<0.0001	Yes
T [11-14.5°C], Time 5th and 6th days	0.087	48.4	<0.0001	Yes
T [11-14.5°C], Time 6th and 7th days	------	0.9	0.6	No
T [18-21.5°C], Time 1st and 2nd days	0.55	2630	<0.0001	Yes
T [18-21.5°C], Time 2nd and 3rd days	-----	0.1	0.4	No
T [21.5-25°C], Time 1st and 2nd days	0.503	6270	<0.0001	Yes
T [21.5-25°C], Time 2nd and 3rd days	-----	2.4	0.09	No

(1) The *Estimate* column represents the estimate of the difference in the growth of the *E. coli* organisms in two consecutive days.

(2) Contrasts with *P* values less than 0.05 are statistically significant.

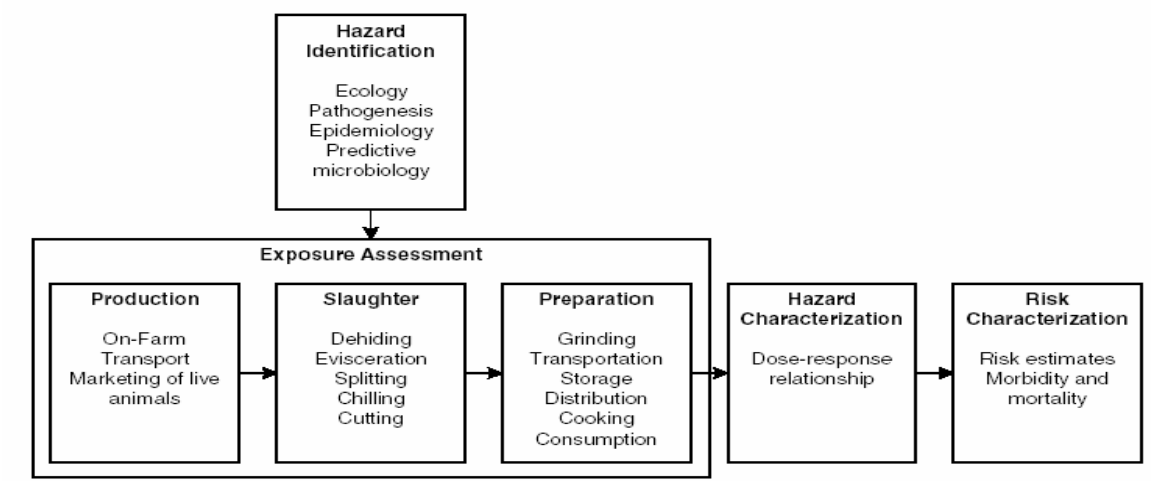


Figure 1. Schematic Diagram of the Farm-to-Table Risk Assessment Model for *E. coli O157:H7* in Ground Beef. Source: FSIS, 2001.

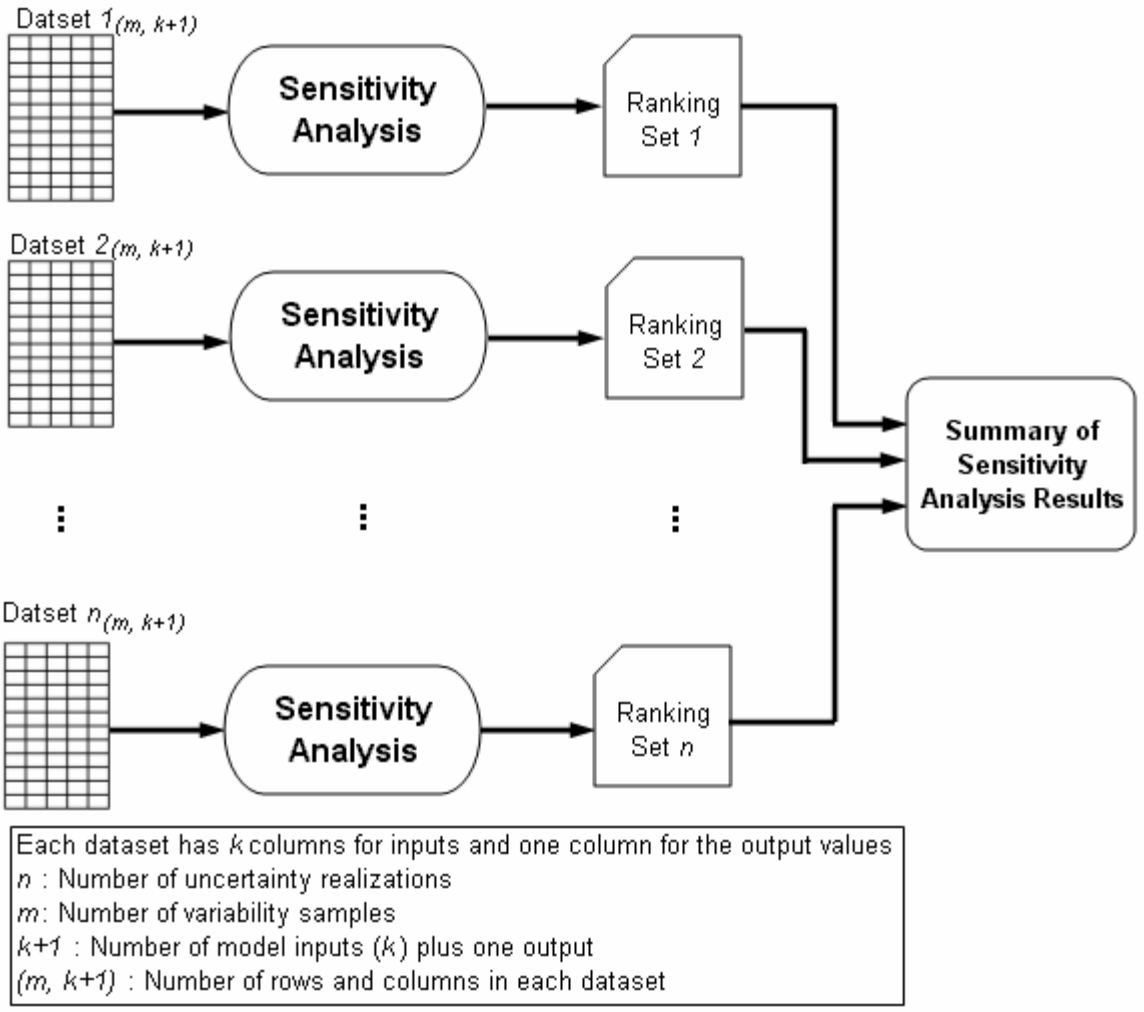


Figure 2. Case Study Scenario for Application of Sensitivity Analysis to Variability Analysis for Different Uncertainty Realizations of a Model.

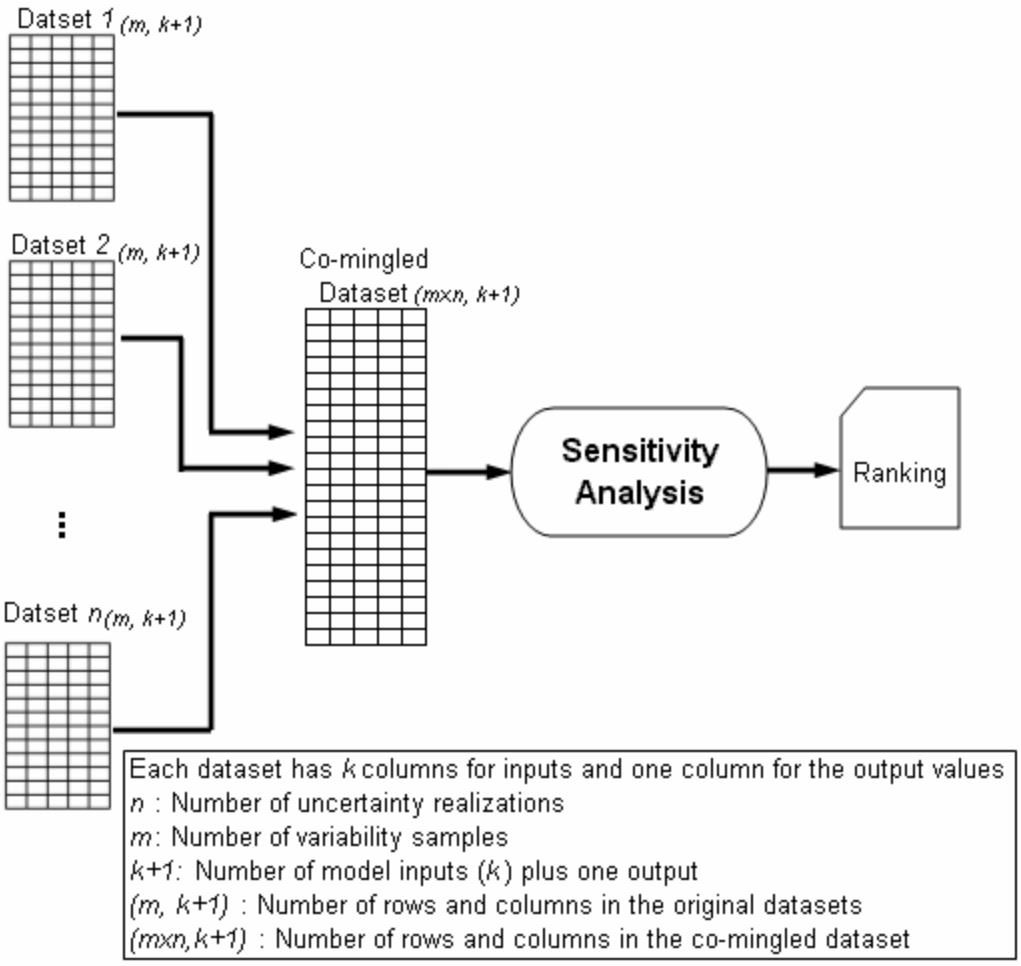


Figure 3. Case Study Scenario for Application of Sensitivity Analysis to the Co-mingled Variability and Uncertainty Analysis of a Model.

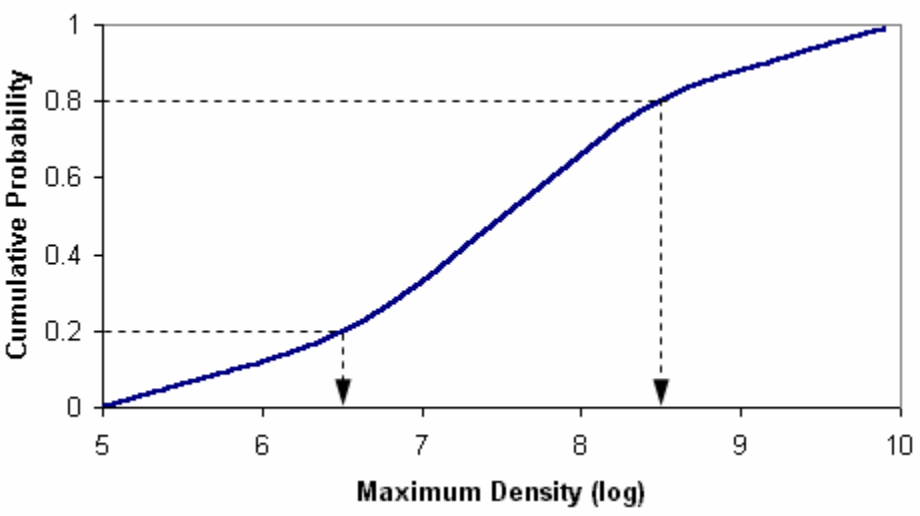


Figure 4. Cumulative Probability Function for Maximum Density.

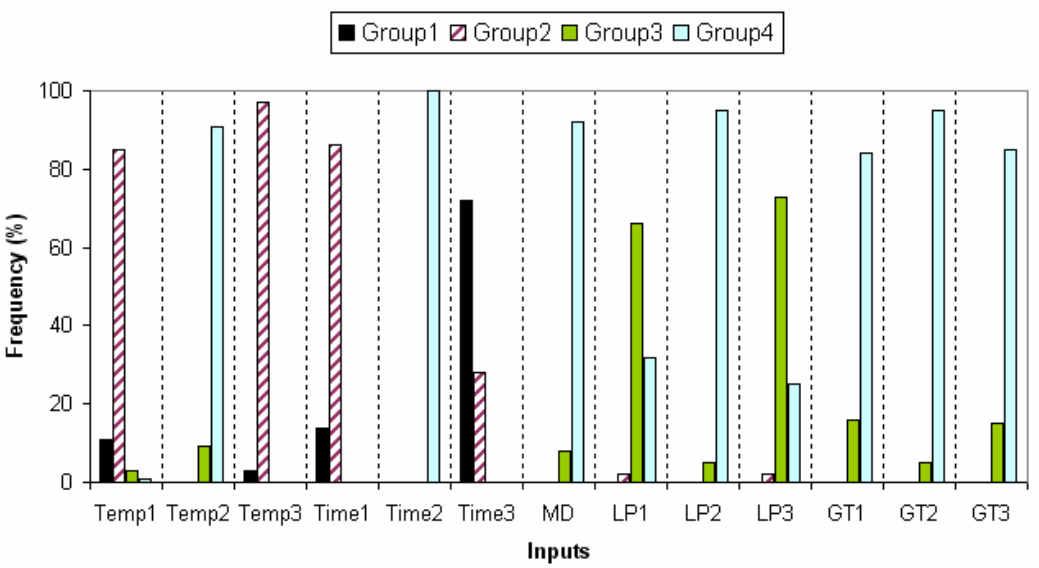


Figure 5. Frequencies Associated with Identification of Each Factor in an Importance Group Based on ANOVA. Group 1: The Most Important Factor; Group 2: Factors with Secondary Importance; Group 3: Factors with Minor Importance; and Group 4: Unimportant Factors.

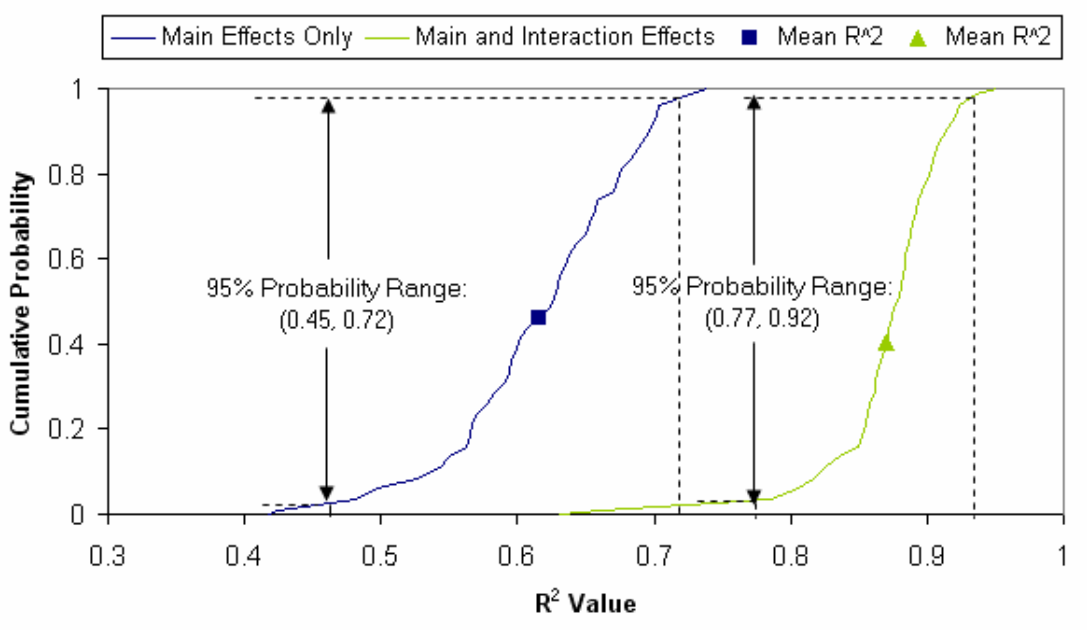


Figure 6. Cumulative Probability Distribution of R^2 Values in the Two-Dimensional Analysis for a Model including Main Effects and a Model Including Main and Interaction Effects.

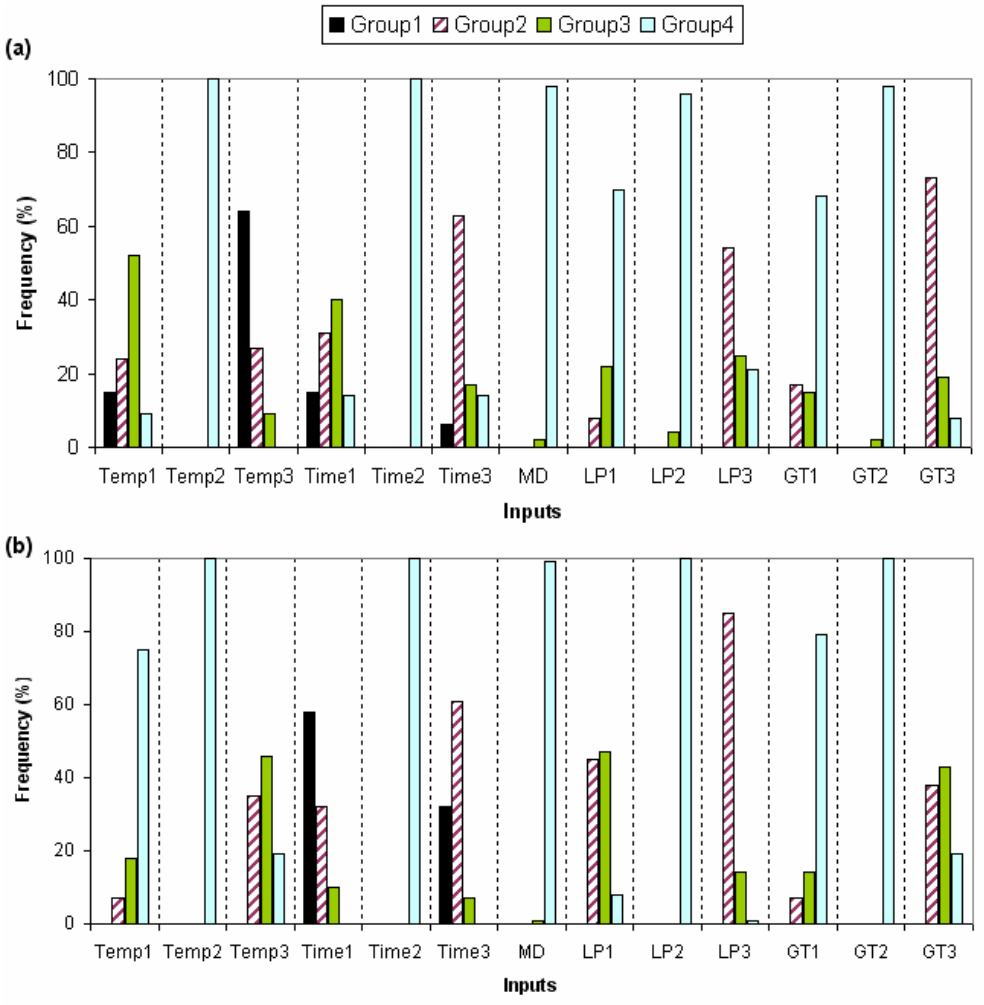


Figure 7. Frequencies Associated with Identification of Each Input in an Importance Group Based on: (a) Sample Correlation; and (b) Rank Correlation. Group 1: The Most Important Input; Group 2: Inputs with Secondary Importance; Group 3: Inputs with Minor Importance; and Group 4: Unimportant Inputs.

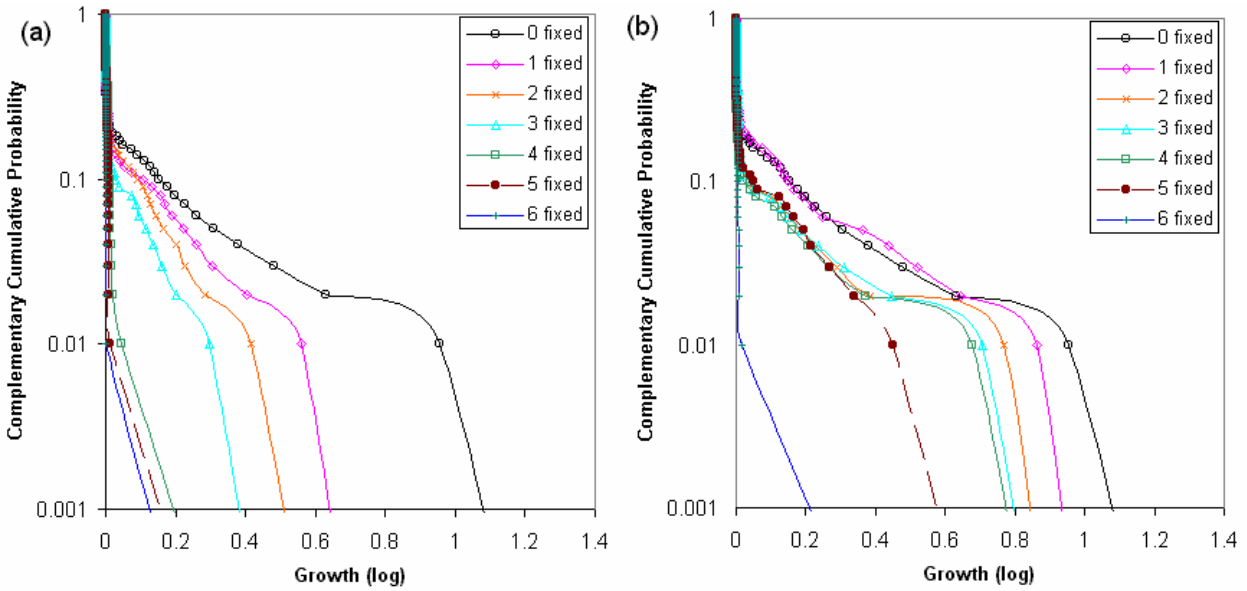


Figure 8. Conditional Complementary Cumulative Distribution Functions: (a) Inputs are Successively Conditioned Based on ANOVA Results; and (b) Inputs are Successively Conditioned Based on Pearson Correlation Analysis Results. The Output is Conditioned by Holding 0, 1, 2, 3, 4, 5, and 6 Most Important Inputs at Their Mean Values.

PART VI

EVALUATION OF SAMPLING-BASED METHODS FOR SENSITIVITY ANALYSIS: CASE STUDY FOR THE *E. COLI* FOOD SAFETY PROCESS RISK MODEL

In preparation to be submitted to:

Human and Ecological Risk Assessment

Abstract

Sensitivity analysis methods can help in identifying key sources of variability and uncertainty that could be priorities for additional data collection or research. Because microbial food safety process risk (MFSPR) models are typically nonlinear, with interactions and threshold effects, a key objective of this work was to evaluate and compare selected sampling-based sensitivity analysis methods based upon application to an example model. Pearson correlation analysis, Spearman correlation analysis, sample standardized regression analysis, rank regression analysis, sample stepwise regression analysis, and rank stepwise regression analysis were applied to an *E. coli O157:H7* microbial food safety process risk model with a focus on the slaughter module. Selected methods were applied to a probabilistic simulation that distinguished variability and uncertainty in inputs in order to identify key sources of variability with quantification of uncertainty in those estimates. Selected methods typically identified the same inputs as the most important one and those with no importance. However, there were differences in the rankings of inputs with secondary and minor importance. Pearson and Spearman correlation analyses, which do not account for interactions between inputs, typically produced comparable results to each other but often differed from the four regression-based methods. Selected methods are evaluated based on the characteristics of an ideal sensitivity analysis method for application to MFSPR models.

Key Words: Sampling-Based Sensitivity Analysis Methods, Microbial Food Safety Process Risk Models, Variability, Uncertainty

1.0 Introduction

Sensitivity analysis is becoming increasingly recognized as an integral part of model development and a tool for helping to answer key policy questions (McCarthy et al., 1995; Helton, 1997). Sensitivity analysis can be used for: quantifying the impact of changes in input values on model output (Cullen and Frey, 1999); evaluating how the variation in the output of a model can be apportioned, qualitatively or quantitatively, among model inputs (Saltelli et al., 2002); and identifying factors contributing to particular outcomes (i.e., worst or best) of interest (Mishra et al., 2003). Sensitivity analysis can provide insights regarding the key sources of variability and uncertainty among model inputs, and hence, prioritize additional data collection and research (Cullen and Frey, 1999).

Microbial food safety process risk (MFSPR) models are examples in which the process of identifying key controllable inputs and key sources of variability and uncertainty is challenged by typical characteristics of those models such as non-linearity, thresholds, interactions, and use of both continuous and categorical inputs (Frey, 2002). Thus, there is a need to evaluate and recommend methods for sensitivity analysis that can reliably be used with such models. Patil and Frey (2003) evaluated multiple sensitivity analysis methods based upon applications to a draft *Vibrio parahaemolyticus* risk assessment model developed by the Food and Drug Administration (FDA) (CFSAN, 2001). This paper evaluates in more detail a smaller number of sensitivity analysis methods using a different risk assessment model as a testbed, and differs from earlier work in accounting for two-dimensional simulation of variability and uncertainty.

This paper evaluates the capability of several selected sensitivity analysis methods to address characteristics associated with MFSPR models. The questions that this paper addresses are:

- What are the advantages and disadvantages of the selected methods?
- How does uncertainty in model inputs affect ranks of key sources of variability?
- What are the key similarities and differences of results based on the selected sensitivity analysis methods?
- What are the implications of sensitivity analysis for refinement of the analysis and for risk management?

2.0 Methodology

The methodology for systematic evaluation of selected sensitivity analysis methods consists of choice of an example MFSPR model to use as a testbed, defining case scenarios, specification of the types of simulations to apply using the model, choice of sensitivity analysis methods, application of the methods, and comparison of results.

2.1 Choice of a MFSPR Model

MFSPR models have been widely used in the field of microbial food safety to provide insight with respect to risk management strategies (Rose and Sobsey, 1993; Jaykus, 1996). Examples include *Salmonella* in eggs (SERAT, 1998), *E. coli O157:H7* in ground beef (FSIS, 2001), *Vibrio parahaemolytics* in raw Molluscan Shellfish (CFSAN, 2001), and *Listeria monocytogenes* in ready-to-eat (RTE) foods (CFSAN, 2003).

The *E. coli O157:H7* food safety process risk model in ground beef is selected as a testbed. *E. coli O157:H7* represents an important public policy issue. Ground beef servings contaminated with *E. coli O157:H7* organisms are estimated to be responsible for some 73,500 cases of infection, 2,150 hospitalizations, and 61 deaths in the United States each year (Mead el al., 1999). The *E. coli* model has a two-dimensional probabilistic framework and typical characteristics of MFSPR models.

2.2 Definition of Case Scenarios for Example Case Studies

The *E. coli* model is implemented in Microsoft Excel and uses @Risk (software add-in) for defining probability distributions for inputs. The *E. coli* model has a modular framework. Each key activity or process such as production, slaughter, and preparation is modeled in a separate module (FSIS, 2001). Each module may have inputs that are calculated based upon a predecessor module and its outputs may be fed to a successor module. Typically, the output of a module is binned into predefined intervals prior to transferring to a successor module. This modeling characteristic leads to a loss of one-to-one correspondence between outputs of one module and the inputs of its predecessors. However, creating one-to-one correspondence in this case would require substantial restructuring and recoding of the model. Therefore, for the case studies, selected sensitivity analysis methods are applied to the individual modules.

The slaughter module is selected as a testbed for evaluation of selected sensitivity analysis methods. This module has a two-dimensional probabilistic simulation framework. The slaughter module estimates the occurrence and extent of *E. coli* contamination as live cattle transition to carcasses, meat trim, and aggregates of meat trims in 2000-pound combo bins destined for commercial ground beef production. The slaughter module includes seven steps: (1) arrival of live cattle at the slaughter plant; (2) dehiding; (3) decontamination following dehiding; (4) evisceration; (5) final washing; (6) chilling; and (7) carcass fabrication (FSIS, 2001). The output of interest in the slaughter module is the contamination in combo bins.

Table 1 summarizes 12 inputs including their mean values and 95 percent probability ranges. These values are estimated based on the sampled values from probability distributions of these inputs in a Monte Carlo simulation that co-mingles variability and uncertainty. Co-mingling variability and uncertainty in a single dimension provides the widest range of variation for each input (Frey et al., 2004). Among the inputs, chilling effect (CH_{eff}), washing effect (W_{eff}),

and trim/vacuum/washing efficiency (TVW) represent potentially controllable sources of variability. Table 1 also gives the type of variability and uncertainty distributions for each input (FSIS, 2001). Uncertainty is incorporated in the parameters of the variability distributions. For example, number of combo bin that each carcass contributes has a triangular distribution representing variability in the input. The most likely value of the variability distribution is uncertain and quantified using a uniform distribution.

2.3 Probabilistic Analysis

Alternative approaches can be used to distinguish between variability and uncertainty in inputs and their contribution to the output variation (Frey and Rhodes, 1996; Cullen and Frey, 1999; Frey et al., 2004). The choice of an appropriate approach should be made taking into account the assessment objectives, the data quality objectives, the availability of data, and the importance of the assessment.

Variability analysis for different uncertainty realizations is selected as the probabilistic approach in the slaughter module. As illustrated in Figure 1, the focus of sensitivity analysis is to identify the key variability inputs for each realization of uncertainty (Frey and Rhodes, 1996; Helton, 1997). A realization refers to one model simulation based upon one randomly sampled value for each probabilistic input. In this situation, sensitivity analysis provides insight regarding the robustness of identified key sources of variability with respect to uncertainty. Sensitivity analysis is applied separately to each uncertainty realization. For this probabilistic scenario, values of 100 and 650 were selected for the number of uncertainty realizations and variability iterations, respectively, based upon the storage limit in each Excel sheet. Thus, there are 100 alternative sensitivity rankings of key sources of variability. Differences in rankings among 100

uncertainty realizations indicate uncertainty regarding the identification of key sources of variability.

2.4 Selected Sensitivity Analysis Methods

With respect to MFSPR models, preferred sensitivity analysis capabilities include ability to: consider simultaneous variation in inputs; address specific model characteristics such as non-linearity, interactions, and thresholds; handle alternative input types (e.g., continuous and categorical); discriminate among sensitive inputs; and provide quantitative measure of sensitivity (Frey, 2002).

Frey et al. (2004) presented a comparison of the capabilities of many sensitivity analysis methods with regard to these characteristics. Although many approaches to sensitivity analysis exist, sampling-based methods are potentially suitable because of: conceptual simplicity and ease of implementation; availability of variety of sensitivity analysis procedures; and effectiveness as a model verification procedure (Helton and Davis, 2002). Sampling-based methods have been widely used for sensitivity analysis in different scientific disciplines (Whiting et al., 1993; Gwo et al., 1996; Kolev and Hofer, 1996; Hofer, 1999). Such methods involve generation and exploration of a mapping from model inputs domain to the model simulation results.

Among various available sampling-based methods, sample (Pearson) correlation analysis (PCA), rank (Spearman) correlation analysis (SCA), sample standardized regression analysis (SRA), rank regression analysis (RRA), sample stepwise regression analysis (SSRA), and rank stepwise regression analysis (RSRA) are selected for evaluation with respect to their behavior and effectiveness when applied to a MFSPR model. Each of these methods is briefly described.

2.4.1 Correlation Analysis

Correlation analysis can be performed based on the sampled values of inputs (PCA) or rank transformed values of inputs (SCA). PCA evaluates the strength of linear association between paired input and output values, while SCA is a measure of the strength of the monotonic relationship between two variables and can account for monotonic nonlinear relationships (Gibbons 1985, Siegel and Castellan 1988, and Kendall 1990). Both Pearson and Spearman correlation coefficients can range from -1 to +1. A value of zero represents a lack of correlation (Edwards, 1976). Correlation analysis measures the effect of one input (i.e., X_i) at a time on the response variable (Y) (Helton and Davis, 2002).

Inputs can be ranked based on the absolute values of correlation coefficients. Pearson coefficients are inaccurate for nonlinear models, and Spearman coefficients are inaccurate for non-monotonic models. Neither Pearson nor Spearman coefficients can capture complex dependencies such as thresholds or directly deal with interactions among inputs.

2.4.2 Regression Analysis

There are many variations of regression analysis (Iman et al., 1985; Neter et al., 1996; Sen and Srivastava, 1990). Regression analysis can be used to assess the combined effects of multiple inputs on an output (Neter et al., 1996). This study focused on evaluation of standardized least squares regression since this is a commonly used method. A regression model is defined as:

$$\hat{y} = b_o + \sum_{i=1}^{n} b_i T_i \qquad (1)$$

where $b_0, b_1, \ldots, b_n$ are regression coefficients. The term T_i can be an input term (X_i), an interaction term ($X_i \times X_j$), or any higher order term. Typically, the data for each input and output are standardized to remove the effects of scale (Neter et. al., 1996).

Similar to correlation analysis, regression analysis can be performed based on the sampled values of inputs (SRA) or rank transformed values of inputs (RRA). SRA involves fitting a regression model to a dataset that includes input values and corresponding output values, while RRA is based upon ranks for the values of inputs and the output. RRA is especially useful when there is high amount of variance or noise in the data or if the model is non-linear but monotonic (Saltelli, 2000).

For a linear regression model, a standardized regression coefficient provides a measure of importance for each input (Devore and Peck, 1996; Neter et al., 1996). Similarly, rank regression coefficients, which are independent of scale, can be used to rank the inputs. However, rank regression coefficients cannot be transformed back to obtain sensitivities in terms of the original values of an input (Saltelli, et al., 2000).

An alternative but related definition of importance is based on the partitioning of the total sum of squares (*TSS*) of the output to each of the regression terms to estimate the relative partial sum of squares (*RPSS*) (Gardner and Trabalka, 1985; Rose et al., 1991). This approach is used as a measure of importance when interaction terms are considered in the regression model. The *RPSS* for term T_i is calculated as:

$$RPSS_{T_i} = 100 \times \frac{(RSS - RSS_{-T_i})}{TSS} \tag{2}$$

where,

RSS = Regression sum of squares for the full model

RSS_{-Ti} = Regression sum of squares for the model with term T_i missing

TSS = Total sum of squares

The adequacy of the regression model can be assessed using the coefficient of multiple determination, R^2, which is a measure of the amount of variance in the output explained by a

regression model (Draper and Smith, 1981). In rank regression analysis, a high R^2 value indicates a monotonic relationship.

2.4.3 Stepwise Regression Analysis

When there are many inputs to consider, regression models such as Equation (2) that contain all inputs typically are unwieldy and unnecessary. Stepwise regression analysis is a search procedure that sequentially identifies the subset of inputs that should be included in the regression model (Neter et al., 1996). Inputs are added one at a time until no additional significant input can be identified (Helton and Davis, 2000).

The importance of inputs is determined by the order in which they enter the regression model, the size of the regression coefficient for each input, and the incremental change in the coefficient of determination, known as partial R^2 values (Helton and Davis, 2002). When the inputs are uncorrelated, the partial R^2 value is the fraction of the total uncertainty in the output that can be accounted for by the input being added at a given step (Saltelli et al., 2000).

3.0 Comparison of Sampling-Based Sensitivity Analysis Methods

This section presents results from comparison of selected sampling-based sensitivity analysis methods. Methods are compared based on the similarity of the results for the key sources of variability and associated ranges of uncertainty. The macro feature in SAS© was implemented for automation of the application of these methods to the two-dimensional simulation of variability and uncertainty and summarization of the results.

Later in a section for discussion and conclusions, methods are evaluated based on their capabilities in addressing characteristics such as considering simultaneous variation in inputs, considering combined effect of inputs, addressing interactions, applicability in a two-

dimensional framework, ease of implementation, providing measure of statistical significance, and discriminating the important inputs.

3.1 Correlation Analysis

Table 2 summarizes the percentage of the uncertainty realizations that produced a statistically significant correlation coefficient, mean ranks, and range of ranks based on PCA and SCA in 100 uncertainty realizations. Table 2 identifies one of four importance groups with which each input is associated: (1) Group 1: the most important input; (2) Group 2: inputs with secondary importance with ranks between 2 and 5; (3) Group 3: inputs with minor importance with ranks between 6 and 8; and (4) Group 4: inputs that are unimportant with ranks between 9 and 12. Inputs within each group have a clear difference with respect to sensitivity to those in other groups. Mean Pearson correlation coefficients are 0.69, 0.31, 0.13, and 0.03 for Groups 1 to 4, respectively, with a typically ±10 percent uncertainty range, while mean Spearman correlation coefficients are 0.90, 0.77, 0.54, and 0.14 for Groups 1 to 4, respectively, with a typically ±3 percent uncertainty range. The dominant importance group for each input is identified and the corresponding frequency is given.

Based on PCA results, there is typically substantial range of uncertainty in the rank of each input. The importance of an input can vary from substantially high to low (i.e., unimportant) due to uncertainty. However, the dominant importance group associated with each input is typically robust to uncertainty. For example, CH_{eff} has a range of ranks between 1 and 12. However, CH_{eff} is dominantly selected in Group 1 in 68 percent of uncertainty realizations and there is only a 7 percent probability that CH_{eff} is assigned to Group 3 or 4. In contrast, there is ambiguity with respect to the importance of W_{eff}. Three probabilities associated with identification of W_{eff} in Groups 2 to 4 range from 29 to 37 percent. Thus, the actual importance

of W_{eff} may vary between Groups 2, 3, and 4. Figure 2(a) shows the probabilities associated with selection of the inputs in each of the importance groups.

SCA typically provides similar results to those from PCA with respect to identification of groups of inputs with comparable importance. For example, SCA identifies CH_{eff} as the most important input in 84 percent of uncertainty realizations with a mean rank of 1.4. However, SCA provides more robust results with respect to uncertainty, because: (1) the probability that each input is identified as statistically significant in uncertainty realizations typically is higher; (2) the ranges of uncertainty in ranks associated with inputs are narrower; and (3) the probability associated with the dominant importance group for each input is substantially higher. For example, SCA identifies a statistically significant correlation coefficient for TNCB in 94 percent of uncertainty realizations, while this probability is only 30 percent for PCA. The CH_{eff} rank varies between 1 and 4 in SCA compared to 1 and 12 in PCA. N_{org} is selected within Group 2 with a probability of 84 percent compared to a probability of 57 percent in PCA. The probabilities associated with identification of other inputs in each of the four importance groups are shown in Figure 2(b).

A hypothesis is tested to evaluate whether there is a statistically significant difference between estimated mean ranks for each input based on PCA and SCA. Figure 3 shows a summary of the results including mean ranks and the corresponding 95 percent confidence intervals. If 95 percent confidence intervals overlap, corresponding mean ranks do not significantly differ. Typically, PCA and SCA provide comparable results for inputs with high importance. For example, the confidence intervals based on PCA and SCA overlap for CH_{eff}, N_{org}, and TNC. Thus, results based on the two methods are comparable for these three inputs. However, PCA and SCA typically do not provide comparable results for inputs that have mean

ranks greater than 6. The confidence intervals for those inputs do not overlap indicating a statistically significant difference for the mean ranks.

The main findings based on the comparison of PCA and SCA are as follows: (1) both methods normally provided similar results with respect to the identification of groups of inputs with comparable importance; (2) the mean ranks for inputs selected in Groups 1 and 2 based on the two methods are statistically comparable; and (3) SCA provided more robust insights with respect to the key sources of variability.

3.2 Regression Analysis

Table 3 summarizes results based on SRA. Four interaction terms are considered in the regression model including $CH_{eff} \times TVW$, $N_{org} \times TVW$, $N_{org} \times W_{eff}$, and $N_{evisc} \times W_{eff}$. Similar to correlation analysis, four importance groups are defined. The most important regression term based on SRA is CH_{eff} with a mean rank of 3.9. N_{org}, TVW, and interaction terms including $CH_{eff} \times TVW$, $N_{org} \times TVW$, and $N_{org} \times W_{eff}$ are dominantly selected in Group2 with probabilities between 43 and 61 percent. No regression term is dominantly selected in Group 3. Because regression terms in Group 4 were typically identified to have no statistically significant effect in most of the uncertainty realizations, the regression terms selected in Groups 1 and 2 account for the most of the variability in the output that is captured by the regression model.

There is a wide range of uncertainty associated with rank of each regression term causing the importance of an input vary from substantially high to low (i.e., unimportant). The dominant importance group for each regression term is typically robust to uncertainty. However, for some regression terms there is ambiguity in their actual importance due to comparable probabilities associated with each of the four importance groups. These probabilities are shown in Figure 4(a). For example, the interaction between CH_{eff} and TVW has a range of ranks between 1 and 16.

However, this interaction is selected in Group 2 in 61 percent of uncertainty realizations and there is only an 18 percent probability that $CH_{eff} \times TVW$ is assigned to Groups 3 and 4. In contrast, there is ambiguity with respect to the importance of CSA as probabilities associated with selection of CSA in Groups 2 to 4 range from 28 to 41. Due to the comparable probabilities, the actual importance of CSA may vary between Groups 2, 3, and 4.

Similarity between SRA and RRA results is with respect to the identification of groups of inputs with comparable importance. RRA results summarized in Table 4 indicate that CH_{eff} is identified as the most important input in 40 percent of uncertainty realizations with a mean rank of 3.7. However, RRA provides different importance groups associated with TNC and CSA. These two inputs were mostly selected in Group 2. RRA presents more robust insights with respect to the key sources of variability. The dominant importance group associated with each regression term is typically robust to uncertainty. Except for CH_{eff} that has comparable probabilities associated with Groups 1 and 2, other regression terms have clearly higher probabilities associated with their dominant importance groups. The associated probabilities with the importance groups are shown in Figure 4(b).For example, P_{both} is dominantly selected in Group 4 with a probability of 86 percent. There is only a 14 percent probability that P_{both} has a higher importance.

A hypothesis is checked to evaluate whether estimated mean ranks for each regression term based on the two methods are significantly different. Results from this analysis including mean ranks and corresponding 95 percent confidence intervals for the regression terms are shown in Figure 5. An overlap in confidence intervals indicates that the corresponding mean ranks are statistically comparable. Typically, mean ranks estimated based on SRA and RRA for

the regression terms are statistically comparable as the confidence intervals overlap. The only exceptions are the mean ranks for TNC and CSA with no overlap in the confidence intervals.

Comparison of the R^2 values estimated based on SRA and RRA may provide insight about the functional association between the model response and the inputs. For a model that has a non-linear but monotonic functional structure, the use of rank transformation can improve the amount of variability in the output that is captured by the regression model. Figure 6 shows the cumulative distribution functions (CDF) for the 100 R^2 values obtained in the two-dimensional simulation based on SRA and RRA. The R^2 values for RRA varied between 0.72 and 0.89 with an average of 0.78. The R^2 values for SRA vary between 0.18 and 0.80 with an average of 0.62. Rank transformation of data improved the amount of variability in the output that is captured by the regression model. Thus, the slaughter module has a non-linear but monotonic functional form.

The main findings based on the comparison of SRA and RRA are: (1) both methods normally provided similar results with respect to the identification of groups of inputs with comparable importance; (2) the mean ranks for the regression terms are typically statistically comparable; and (3) RRA provided more robust insights with respect to the key sources of variability; and (4) the slaughter module has a non-linear but monotonic functional form.

3.3 Stepwise Regression Analysis

The summary of results based on SSRA and RSRA is given in Table 5. The percentage of uncertainty realizations that each term is selected in the regression model is quantified. Mean partial R^2 and corresponding 95 percent probability range also are given. For each regression term, the mean rank represents an arithmetic mean of ranks in 100 uncertainty realizations based

on the order in which the term entered the regression model. Table 5 also provides the importance group associated with each regression term.

SSRA selects CH_{eff} as the most important input. This input is selected in the regression model in 97 percent of uncertainty realizations. In average, selection of CH_{eff} in the regression model is associated with approximately 20 percent of the output variability that is accounted for by the regression model. CH_{eff} has a statistically significant interaction with TVW that accounts for approximately 13 percent of the captured output variability. This interaction effect is selected in all the uncertainty realizations. Typically, CH_{eff}, N_{org}, N_{evisc}, and the four interaction terms account for most of the output variability captured by the regression model. Other inputs normally have minor contribution to this amount.

For each regression term, wide ranges of uncertainty associated with ranks are identified. Each regression term may have substantially high to low importance. However, the dominant importance group for each term has a probability that is typically clearly higher than those associated with other importance groups. These probabilities are shown in Figure 7(a). For example, the rank for TNCB varies between 5 and 15 in uncertainty realizations. However, there is a 96 percent probability that TNCB is selected in Group 4, while there is only a 4 percent probability that TNCB has a higher importance. There is uncertainty with respect to the dominant importance group of N_{org}, W_{eff}, and $N_{org} \times$ TVW. For example, three probabilities associated with identification of N_{org} in Groups 2 to 4 range from 25 to 36 percent. Thus, the actual importance of N_{org} may vary between Groups 2, 3, and 4.

RSRA similarly identified CH_{eff} as the most important input with a probability of 46 percent. CH_{eff} accounts for 34 percent of the output variability in average and was selected in all uncertainty realizations. CH_{eff} has a statistically significant interaction with TVW that in average

accounts for 31 percent of the output variability. Considering the mean partial R^2 values for CH_{eff}, N_{org}, CSA, TNCB, $CH_{eff} \times TVW$, $N_{org} \times TVW$, and $N_{org} \times Weff$, these regression terms account for a substantial amount of the output variability that is captured by the regression model.

There is uncertainty with respect to the importance groups for N_{org}, W_{eff}, CSA, $CH_{eff} \times TVW$, and $N_{org} \times TVW$ based on RSRA. For these regression terms, there are comparable probabilities associated with selection of the terms in each of the four importance groups. These probabilities are shown in Figure 7(b). For example, W_{eff} is selected in Groups 2 to 4 with probabilities range between 31 and 37 percent. Thus, the actual importance of W_{eff} may vary Groups 2, 3, and 4.

Figure 8 shows the mean ranks and corresponding 95 percent confidence intervals for the regression terms based on SSRA and RSRA. Similar to correlation and regression analyses, the insight from this figure is used to evaluate the hypothesis testing whether the mean ranks for each regression term is significantly different based on SSRA and RSRA. Typically, the mean ranks based on the two methods are statistically comparable for the regression terms due to the overlap in their confidence intervals.

The main findings based on the comparison of SSRA and RSRA are: (1) there is not a complete agreement between results of SSRA and RSRA with respect to identification of groups of inputs with comparable importance; (2) the mean ranks for the regression terms are typically statistically comparable; and (3) a few regression terms account for substantial amount of the output variability captured by the regression modelThe summary of results based on SSRA and RSRA is given in Table 5. The percentage of uncertainty realizations that each term is selected in the regression model is quantified. Mean partial R^2 and corresponding 95 percent probability

range also are given. For each regression term, the mean rank represents an arithmetic mean of ranks in 100 uncertainty realizations based on the order in which the term entered the regression model. Table 5 also provides the importance group associated with each regression term.

SSRA selects CH_{eff} as the most important input. This input is selected in the regression model in 97 percent of uncertainty realizations. In average, selection of CH_{eff} in the regression model is associated with approximately 20 percent of the output variability that is accounted for by the regression model. CH_{eff} has a statistically significant interaction with TVW that accounts for approximately 13 percent of the captured output variability. This interaction effect is selected in all the uncertainty realizations. Typically, CH_{eff}, N_{org}, N_{evisc}, and the four interaction terms account for most of the output variability captured by the regression model. Other inputs normally have minor contribution to this amount.

For each regression term, wide ranges of uncertainty associated with ranks are identified. Each regression term may have substantially high to low importance. However, the dominant importance group for each term has a probability that is typically clearly higher than those associated with other importance groups. These probabilities are shown in Figure 7(a). For example, the rank for TNCB varies between 5 and 15 in uncertainty realizations. However, there is a 96 percent probability that TNCB is selected in Group 4, while there is only a 4 percent probability that TNCB has a higher importance. There is uncertainty with respect to the dominant importance group of N_{org}, W_{eff}, and $N_{org} \times$ TVW. For example, three probabilities associated with identification of N_{org} in Groups 2 to 4 range from 25 to 36 percent. Thus, the actual importance of N_{org} may vary between Groups 2, 3, and 4.

RSRA similarly identified CH_{eff} as the most important input with a probability of 46 percent. CH_{eff} accounts for 34 percent of the output variability in average and was selected in all

uncertainty realizations. CH_{eff} has a statistically significant interaction with TVW that in average accounts for 31 percent of the output variability. Considering the mean partial R^2 values for CH_{eff}, N_{org}, CSA, TNCB, CH_{eff} × TVW, N_{org} × TVW, and N_{org} × Weff, these regression terms account for a substantial amount of the output variability that is captured by the regression model.

There is uncertainty with respect to the importance groups for N_{org}, W_{eff}, CSA, CH_{eff} × TVW, and N_{org} × TVW based on RSRA. For these regression terms, there are comparable probabilities associated with selection of the terms in each of the four importance groups. These probabilities are shown in Figure 7(b). For example, W_{eff} is selected in Groups 2 to 4 with probabilities range between 31 and 37 percent. Thus, the actual importance of W_{eff} may vary Groups 2, 3, and 4.

Figure 8 shows the mean ranks and corresponding 95 percent confidence intervals for the regression terms based on SSRA and RSRA. Similar to correlation and regression analyses, the insight from this figure is used to evaluate the hypothesis testing whether the mean ranks for each regression term is significantly different based on SSRA and RSRA. Typically, the mean ranks based on the two methods are statistically comparable for the regression terms due to the overlap in their confidence intervals.

The main findings based on the comparison of SSRA and RSRA are: (1) there is not a complete agreement between results of SSRA and RSRA with respect to identification of groups of inputs with comparable importance; (2) the mean ranks for the regression terms are typically statistically comparable; and (3) a few regression terms account for substantial amount of the output variability captured by the regression model.

3.4 Comparison of Results from Selected Sampling-Based Sensitivity Analysis Methods

Selected sampling-based methods typically provided similar results with respect to identification of the input with highest importance and those with no importance. All methods identified CH_{eff} as the most important input. TNCB, P_{both}, N_{both}, NPE, and N_{evisc} were identified in the group of unimportant inputs by all methods. There is some ambiguity in relative importance of inputs within secondary and minor importance groups. Selected sampling-based methods typically provide different insights with respect to inputs in these two groups.

In general, PCA and SCA tend to have different results in comparison with the regression-based methods. Figure 8 shows the mean ranks and the corresponding 95 percent confidence interval for four inputs including CH_{eff}, N_{org}, TNC and TNI. These inputs are selected as examples to demonstrate an agreement of regression-based methods with respect to providing statistically comparable mean ranks for inputs and their typical difference to the two methods based on correlation. Confidence intervals for the inputs based on the regression-based methods typically overlap indicating a statistical agreement between the mean ranks. In contrast, mean ranks based on PCA and SCA typically have statistically significant difference with the mean ranks based on the regression-based methods with no overlap in the intervals. However, there is an agreement between the results based on PCA and SCA.

4.0 Discussion and Conclusions

Sampling-based methods are usually used for: identifying the variability and/or uncertainty in the model outputs; and determining the important inputs that mostly affect that variability and/or uncertainty. This paper demonstrated the latter using case studies with a complex MFSPR model that has a two-dimensional probabilistic framework with a focus on identifying key sources of variability and quantification of uncertainty in the results. Sampling-

based sensitivity analysis procedures are based on identifying the functional association between values of output and inputs. These methods do not provide robust results when the mapping between output and inputs substantially differs from the one that is the underlying assumption of the particular method. Thus, it is typically recommended that several sensitivity analysis procedures be applied and results be compared. Using this approach there is high probability that important model inputs will be identified by at least one of the procedures.

The key characteristics of an ideal sensitivity analysis method are given in Table 6. These characteristics are inclusive but not exclusive and other features can be added to this comparison based on the specific case study requirement. Each sampling-based sensitivity analysis method is evaluated with respect to addressing these characteristics considering the insights from the case studies. Ideally, a sensitivity analysis method should: respond to the effects of simultaneous variation in all inputs; consider the combined effect of multiple inputs; handle interactions and non-linearities; be applicable in a two-dimensional probabilistic framework; be easy to implement; provide measure of statistical significance of inputs; discriminate the important inputs.

Regression-based methods provided more robust insights with respect to sensitivity as they not only assess the combined effects of multiple inputs on a response variable, but also they identified statistically significant interactions between inputs. The only limitation is for cases that there are many model inputs or the interactions between inputs are more general rather than a particular multiplicative form. In such cases, properly addressing all interaction terms in the regression model becomes more tedious and requires *a priori* knowledge of the functional form of the interaction terms.

The ease of implementation may often constrain the feasibility of a method. A method is typically easier to implement when software tools already exist, especially if they have user-friendly interfaces (i.e., push button methods). Basically, correlation analysis is easy to implement as this method is a typical feature of commonly used statistical software packages such as Crystal Ball^{TM} or @Risk. Of course, ease of implementation will be a function of software availability and programming skill level. However, 'push button' methods such as correlation analysis do not account for typical characteristics of MFSPR models such as interactions. Thus, even though regression-based methods may be more difficult to apply than some of the readily available methods such as correlation analysis, their use may be necessary to capture important characteristics of MFSPR models.

The importance of an input based on selected sampling-based methods varied from substantially high to low due to uncertainty in inputs. Some inputs were consistently ranked high, whereas others were consistently ranked low. The information about range of ranks and associated probabilities is useful to a decision maker. By quantifying uncertainty in the rank of each input, groups of model inputs with similar importance are identified. For the four groups of model inputs identified, inputs have a clear difference with respect to sensitivity in each group compared to those in other groups. Although there is a wide range of uncertainty associated with the rank of each input, the importance group associated with each of the inputs is substantially robust to uncertainty. For example, CH_{eff} has a wide range of uncertainty in ranks between 1 and 12 based on Pearson correlation analysis. However, a decision maker can confidently conclude that this input has the highest importance as CH_{eff} was selected in Group 1 in 68 percent of uncertainty realizations. Typically, rank transforming the data was associated with lower uncertainty in sensitivity analysis results by providing: (1) higher probability for statistically

significant inputs and interaction terms; (2) narrower ranges of uncertainty for ranks of inputs and interaction terms; and (3) higher probability that inputs and interaction terms were selected in a same importance group in different uncertainty realizations.

Three potentially controllable sources of variability in the slaughter module were the first decontamination process after dehiding carcasses (TVW), the second decontamination process after evisceration of carcasses and prior to chilling (W_{eff}), and the chilling process (CH_{eff}). The other sources of variability in the model were mostly some model parameters or variables that cannot be directly controlled. Thus, with respect to the risk management special attention should be devoted to key controllable sources of variability in order to reduce the risk to an acceptable level. Identification of CH_{eff} as the most important input based on all selected methods indicates that the chiller is an actual critical control point in a slaughter module. Growth or decline of *E. coli O157:H7* on the surface of carcasses is largely a function of time and temperature. Fluctuation in chiller temperature or the outright failure to adequately chill carcasses may enable growth. Thus, even with substantial decontamination prior to the chilling step, high contamination may become possible during the chilling process. However, chilling time and temperature are not incorporated in the slaughter module. Thus, an implication of the results is that it is beneficial to disaggregate the model and incorporate more detail regarding specific factors that influence the chilling effect in order to have more clear insights with respect to factors that substantially contribute to the output variability.

Selected sampling-based sensitivity analysis methods typically identified higher importance for the first decontamination step (TVW) compared to the second decontamination step (W_{eff}). TVW is mostly selected in the group of inputs with secondary or minor importance, while W_{eff} is mostly selected in the group of inputs with minor or no importance. Thus, in order

to reduce the contamination in final products of a slaughter plant it is more productive to increase the efficiency of the decontamination process following dehiding animals. Although washing carcasses prior to the chilling process potentially can reduce the contamination, insights from sensitivity analysis indicated that the second decontamination step is not as effective as the first one.

Case studies presented in this paper focused on identification of key sources of variability followed by quantification of uncertainty in the analysis results. However, a probabilistic scenario can be arranged in order to identify key contributors to uncertainty in the model output conditional on different values for quantities that are subject to variability. Based on the insights from such an analysis, one can identify how key sources of uncertainty for exposure or risk may change for different combinations of values for variable inputs. Knowledge of key sources of uncertainty for the most exposed or at risk portion of the population (e.g., upper 5 percent of the population exposed to the risk) can be used to prioritize additional data collection or research that could reduce uncertainty. The assessment can be revised based upon new information, and a decision could be made at a later time based upon the reduced uncertainties.

5.0 Acknowledgment

The work reported here is based on work supported by Cooperative Agreement No. 58-0111-0-005 between the U.S. Department of Agriculture, Office of the Chief Economist, Office of the Risk Assessment and Cost Benefit Analysis (USDA/OCE/ORACBA) and the Department of Civil, Construction, and Environmental Engineering at North Carolina State University. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Agriculture.

6.0 References

CFSAN (Center for Food Safety and Applied Nutrition) 2001. Draft Risk Assessment on the Public Health Impact of *Vibrio parahaemolyticus* in Raw Molluscan Shellfish. Prepared by Food and Drug Administration, and U.S. Department of Agriculture, Washington, DC.

CFSAN (Center for Food Safety and Applied Nutrition) 2003. Quantitative Assessment of Relative Risk to Public Health from Foodborne *Listeria monocytogenes* Among Selected Categories of Ready-to-Eat Foods. Prepared by Center for Food Safety and Nutrition, Food and Drug Administration, and U. S. Department of Agriculture, Washington, DC. http://www.foodsafety.gov/~dms/lmr2-toc.html

Cullen, A. C. and Frey, H. C. 1999. Probabilistic Techniques in Exposure Assessment. Plenum Press: New York.

Devore, J. L. and Peck, R. 1996. Statistics: The Exploration and Analysis of Data. 3rd Edition. Brooks/Cole Publishing Company: London, England.

Draper, N.R. and Smith, H. 1981. Applied Regression Analysis. Second Edition. John Wiley and Sons: New York.

Edwards, A. L. 1976. An Introduction to Linear Regression and Correlation. San Francisco, CA: W. H. Freeman.

Frey, H.C. 2002. Introduction to special section on sensitivity analysis and summary of NCSU/USDA workshop on sensitivity analysis. Risk Analysis, 22(3): 539-545.

Frey, H.C., Mokhtari, A., and Zheng, J. 2004. Recommended practice regarding selection, application, and interpretation of sensitivity analysis methods applied to food safety process risk models. Prepared by North Carolina State University for Office of Risk

Assessment and Cost-Benefit Analysis, U.S. Department of Agriculture. Washington DC. www.ce.ncsu.edu/risk.

Frey, H.C. and Rhodes D.S. 1996. Characterizing, simulating, and analyzing variability and uncertainty: an illustration of methods using an air toxics emissions example. Human and Ecological Risk Assessment, 2(4): 762-797.

FSIS (Food Safety and Inspection Service) 2001. Draft Risk Assessment of the Public Health Impact of Escherichia Coli O157:H7 in Ground Beef. Prepared for the Food Safety and Inspection Service and United States Department of Agriculture, Washington, DC. www.fsis.usda.gov/OPPDE/rdad/FRPubs/00-023N/00-023NReport.pdf

Gardner, R. H. and Trabalka, J. R. 1985. Methods of Uncertainty Analysis for a Global Carbon Dioxide Model. DOE/OR/21400-4, US Department of Energy, Washington, DC.

Gibbons, J. D. 1985. Nonparametric Statistical Inference. Statistics: textbooks and monographs. Marcel Dekker, Inc, New York and Basel, 2nd edition.

Gwo, J. P., Toran, L. E., Morris, M. D., Wilson, G. V. 1996. Subsurface stormflow modeling with sensitivity analysis using a Latin-Hypercube sampling technique. Ground Water, 34(5): 811-818.

Helton, J.C. 1997. Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty. Journal of Statistical Computation and Simulation, 57(1): 3-76.

Helton, J. C. and Davis, F. J. 2002. Illustration of Sampling-Based Methods for Uncertainty and Sensitivity Analysis. Risk Analysis, 22(3): 591-622.

Hofer, E. 1999. Sensitivity analysis in the context of uncertainty analysis for computationally intensive models. Computer Physics Communications, 117(1-2): 21-34.

Iman, R. L., Shortencarier, M. J., and Jhonson, J. D. 1985. A FORTRAN 77 program and users guide for the calculation of partial correlation and standardized regression coefficients. Report No. SAND85-0044, Sandia National Laboratories, Albuquerque, NM.

Jaykus L. A. 1996. The application of quantitative risk assessment to microbial food safety risks. Critical Reviews in Microbiology, 22: 279-93.

Kendall, M. and Stuart, A. 1979. The Advanced Theory of Statistics, 4th Ed. MacMillian Publishing Co.: New York.

Kolev, N. I. and Hofer, E. 1996. Uncertainty and sensitivity analysis of a post-experiment simulation of non-explosive melt-water interaction. Experimental Thermal and Fluid Science, 13(2): 98-116.

Mead, P. S., Laurence, S., Vance, D., McCaig, L. F., Bresee, J. S., Shapiro, C., Griffin, P. M., and Tauxe, R. V. 1999. Food-related illness and death in the United States. Emerging Infectious Diseases, 5(5): 607-625.

Mishra, S., Deeds, N. E., and RamaRao, B. S. 2003. Application of classification trees in the sensitivity analysis of probabilistic model results. Reliability Engineering and System Safety, 79(2): 123-129.

McCarthy, M. A., Burgman, M. A., and Ferson, S. 1995. Sensitivity analysis for models of population viability. Biological Conservation, 73(2): 93-100.

Neter, J., Kutner, M.H., Nachtsheim, C.J., and Wasserman, W. 1996. Applied Linear Statistical Models, Fourth Edition. McGraw-Hill: Chicago, IL.

Patil, S. R. and Frey, H. C. 2003. Comparison of sensitivity analysis methods based upon applications to a food safety risk model. Risk Analysis, Submitted December 19, 2002, accepted July 29, 2003.

Rose, K. A., Smith, E. P., Gardner, R. H., Brenkert, A. L., and Bartell, S. M. 1991. Parameter sensitivities, Monte Carlo filtering, and model forecasting under uncertainty. Journal of Forecasting, 10(1): 117-133.

Rose, J. B. and Sobsey, M. D. 1993. Quantitative risk assessment for viral contamination of shellfish and coastal waters. Journal of Food Protection, 56(12), 1043-1050.

Saltelli A., Chan, K., and Scott, M. 2000. Sensitivity Analysis. Probability and Statistics Series. John Wiley & Sons: New York.

Saltelli, A. 2002. Sensitivity analysis for importance assessment. Risk Analysis, 22(3): 1-12.

Sen, A., and Srivastava, M. 1990. Regression Analysis: Theory, Methods, and Applications. Springer-Verlag: New York.

SERAT 1998. Salmonella Enteritidis Risk Assessment for Shell Eggs and Egg Products: Final Report. Prepared by the Salmonella Enteritidis Risk Assessment Team (SERAT) for the Food Safety and Inspection Service (FSIS), U.S. Department of Agriculture, Washington, D.C.

Siegel, S. and Castellan, N. J. 1988. Nonparametric Statistics for the Behavioral Sciences, 2nd Edition. McGraw-Hill, New York.

Whiting, W. B., Tong, T. M., and Reed, M. E. 1993. Effect of uncertainties in thermodynamic data and model parameters on calculated process performances. Industrial and Engineering Chemistry Research, 32(7): 1367-1371.

Table 1. Input Variables and Corresponding Mean Values, 95% Probability Ranges, and Distribution Types in the Slaughter Module of the *E. coli* Model

Inputs	Mean Value (1)	95% Probability Range (1)	Distribution Type	
			Variability	Uncertainty (2)
Total Number of Combo Bin for Each Carcass (TNCB)	4	(2, 6)	Triangular	Uniform
Total Number of Infected Animals (TNI)	4	(0, 14)	Binomial	Exponential
Total Number of Contaminated Animals (TNC)	5	(0, 22)	Poisson	Beta
Probability of Contaminated Cases at both Steps of Dehiding and Evisceration (P_{both})	39.7	(0, 96.7)	Uniform	Binomial
Number of Contaminated Cases at both Steps of Dehiding and Evisceration (N_{both})	1	(0, 2)	Binomial	Poisson
Number of Contaminated Cases at Evisceration (NCE)	1	(0, 3)	Binomial	Poisson
Chilling Effect in logs (CH_{eff})	-0.9	(-3.0, 2.1)	Normal	Uniform
Number of Organisms (N_{org})	118.8	(0, 1100)	EDF(3)	Beta
Trim/Vacuum/Washing Efficiency (TVW)	24.6	(0, 60)	Triangular	Uniform
Number of Organisms Added During Evisceration (N_{evisc})	1.2	(0, 2.45)	EDF(3)	Beta
Washing Effect (W_{eff})	12.4	(0, 45)	Triangular	Uniform
Contaminated Surface Area in cm^2 (CSA)	419.20	(0, 2590)	Binomial	Uniform

(1) Values are based on the co-mingled analysis of variability and uncertainty
(2) Uncertainty distributions are defined for parameters of variability distributions
(3) Empirical distribution function based on FSIS data. Beta distribution is used to define the corresponding cumulative frequency at each value of the distribution.

Table 2. Summary of Pearson and Spearman Correlation Analyses Results

Inputs (1)	Sample (Pearson) Correlation Analysis					Rank (Spearman) Correlation Analysis				
	Freq. (%) (2)	Mean Rank (3)	Range of Rank	Importance Group (4)	Same Importance Group (%) (5)	Freq. (%) (2)	Mean Rank (3)	Range of Rank	Importance Group (4)	Same Importance Group (%) (5)
TNCB	30	9.3	4-12	4	65	94	10.0	9-12	4	100
TNI	97	5.2	2-11	2	66	100	5.8	3-9	3	52
TNC	97	4.5	1-12	2	80	100	3.9	1-7	2	89
P_{both}	40	9.0	4-12	4	53	100	8.0	6-11	3	94
N_{both}	38	8.7	2-12	4	69	65	11.5	9-12	4	100
NCE	47	8.2	2-12	4	55	84	10.2	8-12	4	94
CH_{eff}	100	1.9	1-12	1	68	100	1.4	1-4	1	84
N_{org}	95	3.7	1-12	2	57	100	3.1	1-6	2	84
TVW	68	7.4	3-12	3	56	100	5.6	1-8	3	66
N_{evisc}	35	8.8	1-12	4	71	82	10.0	7-12	4	98
W_{eff}	75	6.6	1-12	3	37	100	5.0	2-8	3	53
CSA	99	4.6	1-12	2	73	100	3.6	1-7	2	85

(1) The abbreviations used for factors in this table are the same as those defined in Table 1.
(2) The percentage of the uncertainty realizations that produce statistically significant correlation coefficient (*P* value less than 0.05).
(3) Arithmetic mean of ranks for 100 uncertainty realizations.
(4) For each input, the importance group represents the dominant range of ranks in uncertainty realizations (See Figure 2 for corresponding probabilities). Group 1: most important input (i.e., rank 1); Group 2: secondary importance inputs (i.e., ranks between 2 and 5); Group 3: minor importance inputs (i.e., ranks between 6 and 8); Group 4: unimportant inputs (i.e., ranks between 9 and 12).
(5) Percentage of the uncertainty realizations that each input fits within the same importance group defined based on the mean ranks.

Table 3. Summary of Sample Standardized Regression Analysis Results

Inputs [1]	Sample Standardized Regression Analysis				
	Freq. (%) [2]	Mean Rank[3]	Range of Rank	Importance Group [4]	Same Importance Group (%) [5]
TNCB	17	13.6	7-16	4	93
TNI	21	11.9	3-16	4	80
TNC	77	8.4	1-16	4	43
P_{both}	24	12.9	6-16	4	89
N_{both}	58	10.3	2-16	4	69
NCE	53	10.3	3-16	4	65
CH_{eff}	91	3.9	1-16	1	43
N_{org}	79	6.1	1-16	2	45
TVW	79	6.3	1-16	2	46
N_{evisc}	62	9.0	1-15	4	32
W_{eff}	74	9.0	1-15	4	32
CSA	74	8.0	1-16	4	41
$CH_{eff} \times TVW$	81	4.1	1-16	2	61
$N_{org} \times TVW$	80	6.6	1-16	2	43
$N_{org} \times W_{eff}$	86	5.4	1-16	2	45
$N_{evisc} \times W_{eff}$	66	9.2	2-16	4	60

(1) The abbreviations used for factors in this table are the same as those defined in Table 1.

(2) The percentage of the uncertainty realizations that produce statistically significant regression coefficient (*P* value less than 0.05).

(3) Arithmetic mean of ranks for 100 uncertainty realizations.

(4) For each input, the importance group represents the dominant range of ranks in uncertainty realizations (See Figure 5 for corresponding probabilities). Group 1: most important input (i.e., rank 1); Group 2: secondary importance inputs (i.e., ranks between 2 and 5); Group 3: minor importance inputs (i.e., ranks between 6 and 8); Group 4: unimportant inputs (i.e., ranks between 9 and 16).

(5) Percentage of the uncertainty realizations that each input fits within the same importance group defined based on the mean ranks.

Table 4. Summary of Rank Regression Analysis Results

Inputs [1]	Rank Regression Analysis				
	Freq. (%) [2]	Mean Rank[3]	Range of Rank	Importance Group [4]	Same Importance Group (%) [5]
TNCB	21	13.4	9-16	4	100
TNI	28	13.1	7-16	4	94
TNC	86	5.8	2-16	2	58
P_{both}	31	12.4	5-16	4	86
N_{both}	63	11.0	5-16	4	73
NCE	61	10.7	5-16	4	76
CH_{eff}	95	3.7	1-13	1	40
N_{org}	89	6.1	1-15	2	42
TVW	94	5.1	1-15	2	49
N_{evisc}	68	9.8	3-15	4	65
W_{eff}	78	9.8	3-15	4	65
CSA	90	6.2	1-15	2	46
$CH_{eff} \times TVW$	89	5.4	1-15	2	44
$N_{org} \times TVW$	83	6.3	1-15	2	41
$N_{org} \times W_{eff}$	87	6.6	1-15	2	49
$N_{evisc} \times W_{eff}$	69	9.7	2-16	4	59

(1) The abbreviations used for factors in this table are the same as those defined in Table 1.

(2) The percentage of the uncertainty realizations that produce statistically significant regression coefficient (*P* value less than 0.05).

(3) Arithmetic mean of ranks for 100 uncertainty realizations.

(4) For each input, the importance group represents the dominant range of ranks in uncertainty realizations (See Figure 5 for corresponding probabilities). Group 1: most important input (i.e., rank 1); Group 2: secondary importance inputs (i.e., ranks between 2 and 5); Group 3: minor importance inputs (i.e., ranks between 6 and 8); Group 4: unimportant inputs (i.e., ranks between 9 and 16).

(5) Percentage of the uncertainty realizations that each input fits within the same importance group defined based on the mean ranks.

Table 5. Summary of Sample Stepwise Regression Analysis and Rank Stepwise Regression Analysis Results

Inputs[1]	Sample Stepwise Regression Analysis						Rank Stepwise Regression Analysis					
	Freq. [2]	Mean Partial R^2	95% Probability Range of R^2	Mean Rank [3]	Range of Rank	Group [4]	Freq. [2]	Mean Partial R^2	95% Probability Range of R^2	Mean Rank [3]	Range of Rank	Group [4]
TNCB	16	0.002	(0.000, 0.008)	10.3	5-15	4	12	0.000	(0.000, 0.001)	10.3	6-13	4
TNI	44	0.003	(0.000, 0.020)	9.3	3-14	4	25	0.000	(0.000, 0.001)	10.7	6-14	4
TNC	80	0.012	(0.000, 0.095)	6.6	1-12	4	90	0.014	(0.000, 0.050)	6.0	1-12	2
P_{both}	26	0.001	(0.000, 0.005)	10.1	6-15	4	58	0.001	(0.000, 0.006)	7.7	3-14	4
N_{both}	53	0.007	(0.000, 0.037)	7.7	2-13	4	66	0.001	(0.000, 0.003)	8.1	2-13	4
NCE	55	0.006	(0.000, 0.036)	8.3	3-13	4	53	0.001	(0.000, 0.005)	8.6	3-15	4
CH_{eff}	97	0.203	(0.000, 0.690)	2.9	1-13	1	100	0.340	(0.000, 0.720)	3.4	1-13	1
N_{org}	83	0.053	(0.000, 0.420)	5.7	1-14	4	91	0.022	(0.000, 0.120)	6.2	1-13	3
TVW	86	0.011	(0.000, 0.067)	6.7	2-14	4	91	0.004	(0.000, 0.015)	5.7	3-11	2
N_{evisc}	56	0.053	(0.000, 0.480)	7.7	1-14	4	63	0.001	(0.000, 0.006)	8.4	2-14	4
W_{eff}	82	0.008	(0.000, 0.058)	6.7	2-13	4	97	0.004	(0.000, 0.020)	6.3	2-13	2
CSA	84	0.012	(0.000, 0.072)	6.8	1-13	4	97	0.032	(0.000, 0.610)	6.2	1-11	2
$CH_{eff} \times TVW$	96	0.128	(0.000, 0.680)	3.8	1-13	2	100	0.312	(0.000, 0.650)	4.4	1-12	2
$N_{org} \times TVW$	87	0.072	(0.000, 0.390)	5.5	1-13	4	93	0.055	(0.000, 0.590)	6.0	1-13	4
$N_{org} \times W_{eff}$	95	0.110	(0.000, 0.510)	4.3	1-12	2	97	0.080	(0.000, 0.510)	4.4	1-13	2
$N_{evisc} \times W_{eff}$	65	0.044	(0.000, 0.420)	7.1	1-14	4	43	0.001	(0.000, 0.007)	8.1	2-13	4

(1) The abbreviations used for factors in this table are the same as those defined in Table 1.

(2) The percentage of uncertainty realizations that the input or the interaction term is selected for the regression model in stepwise regression analysis.

(3) Arithmetic mean of ranks for 100 uncertainty realizations. In each uncertainty realization, regression terms are ranked based on the order in which they are selected in the regression model.

(4) For each input, the importance group represents the dominant range of ranks in uncertainty realizations (See Figure 7 for corresponding probabilities). Group 1: most important input (i.e., rank 1); Group 2: secondary importance inputs (i.e., ranks between 2 and 5); Group 3: minor importance inputs (i.e., ranks between 6 and 8); Group 4: unimportant inputs (i.e., ranks between 9 and 16).

Table 6. Summary of Key Characteristics of Selected Sampling-Based Sensitivity Analysis Methods

Characteristics	Sensitivity Analysis Methods (1)					
	PCA	**SCA**	**SRA**	**RRA**	**SSRA**	**RSRA**
Simultaneous Variation	Yes	Yes	Yes	Yes	Yes	Yes
Combined Effect of Multiple Inputs	No	No	Yes	Yes	Yes	Yes
Interaction	No	No	Yes	Yes	Yes	Yes
Non-linearity	No	Yes(2)	No	Yes(2)	No	Yes(2)
Two-Dimensional Simulation	Yes	Yes	Yes	Yes	Yes	Yes
Ease of Implementation	Yes	Yes	No	No	No	No
Measure of Statistical Significance	Yes	Yes	Yes	Yes	Yes	Yes
Discrimination of Important Inputs	Yes	Yes	Yes	Yes	Yes	Yes

(1) PCA: Pearson Correlation Analysis; SCA: Spearman Correlation Analysis; SRA: Sample Regression Analysis; RRA: Rank Regression Analysis; SSRA: Sample Stepwise Regression Analysis; RSRA: Rank Stepwise Regression Analysis.

(2) Rank transformation can mitigate the problem of non-linearity in monotonic models.

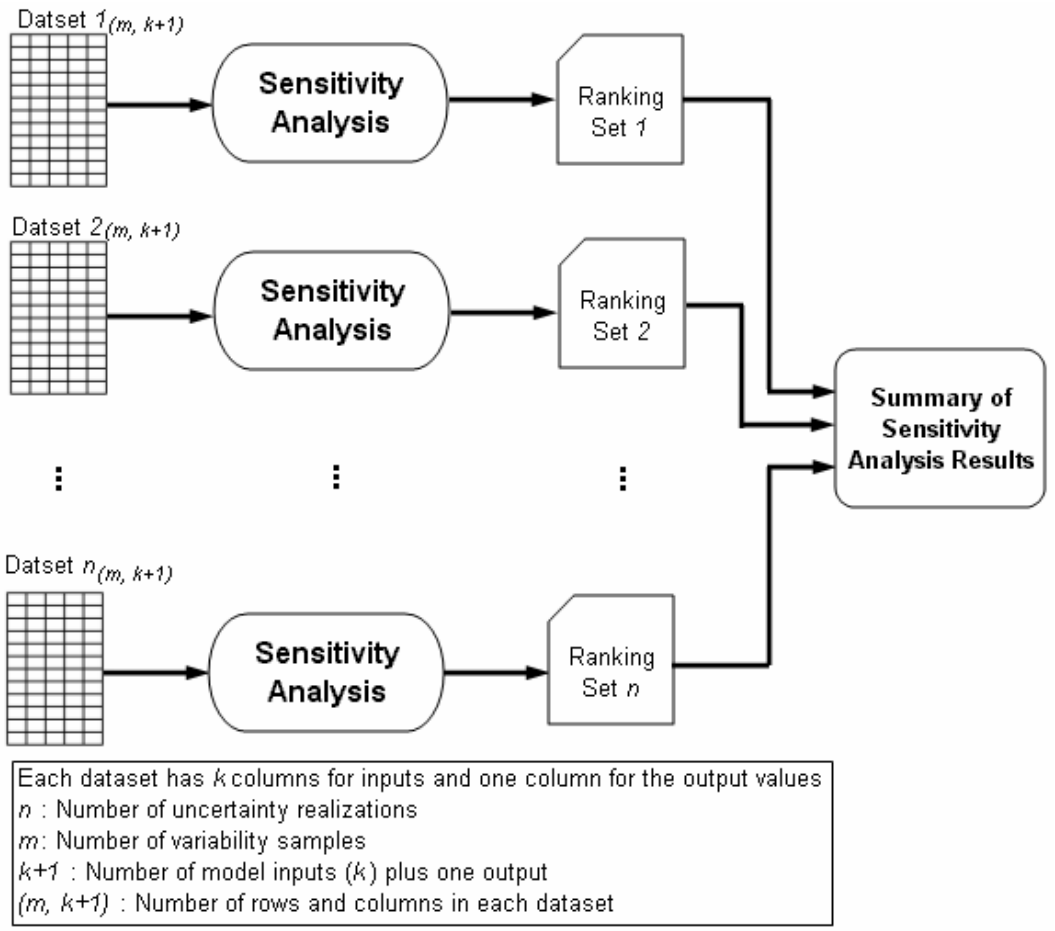


Figure 1. Case Study Scenario for Application of Sensitivity Analysis to Variability Analysis for Different Uncertainty Realizations of a Model.

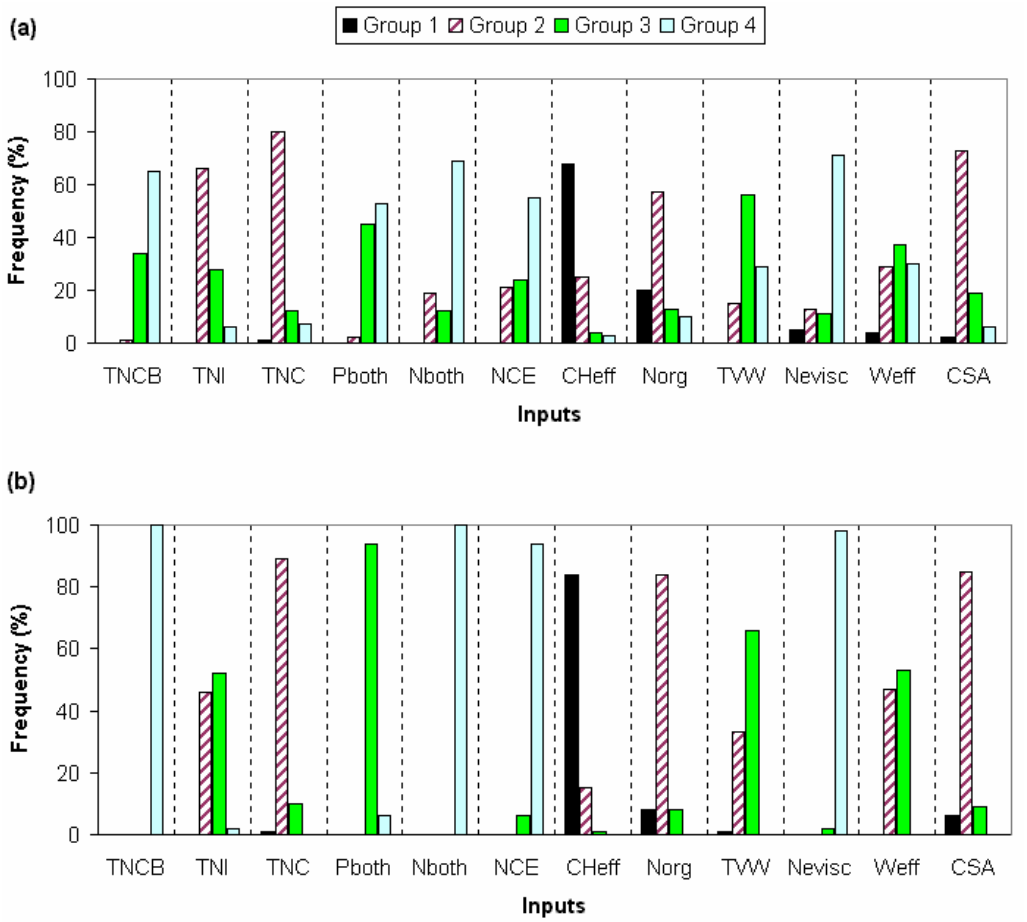


Figure 2. Frequencies Associated with Identification of Each Input in an Importance Group Based on: (a) Pearson Correlation; and (b) Spearman Correlation. Group 1: The Most Important Input; Group 2: Inputs with Secondary Importance; Group 3: Inputs with Minor Importance; and Group 4: Unimportant Inputs.

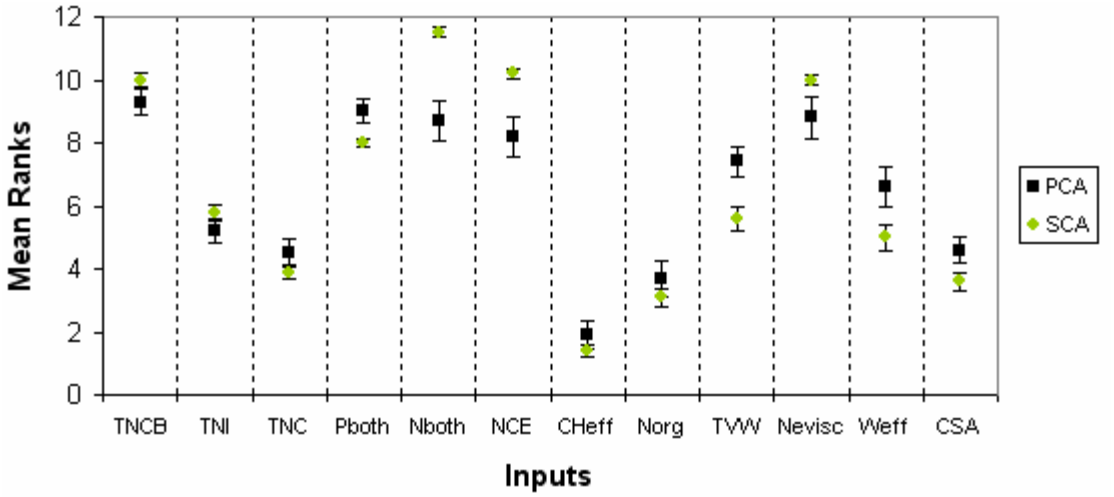


Figure 3. Mean Ranks and Corresponding 95 Percent Confidence Intervals Based on Pearson Correlation Analysis (PCA) and Spearman Correlation Analysis (SCA).

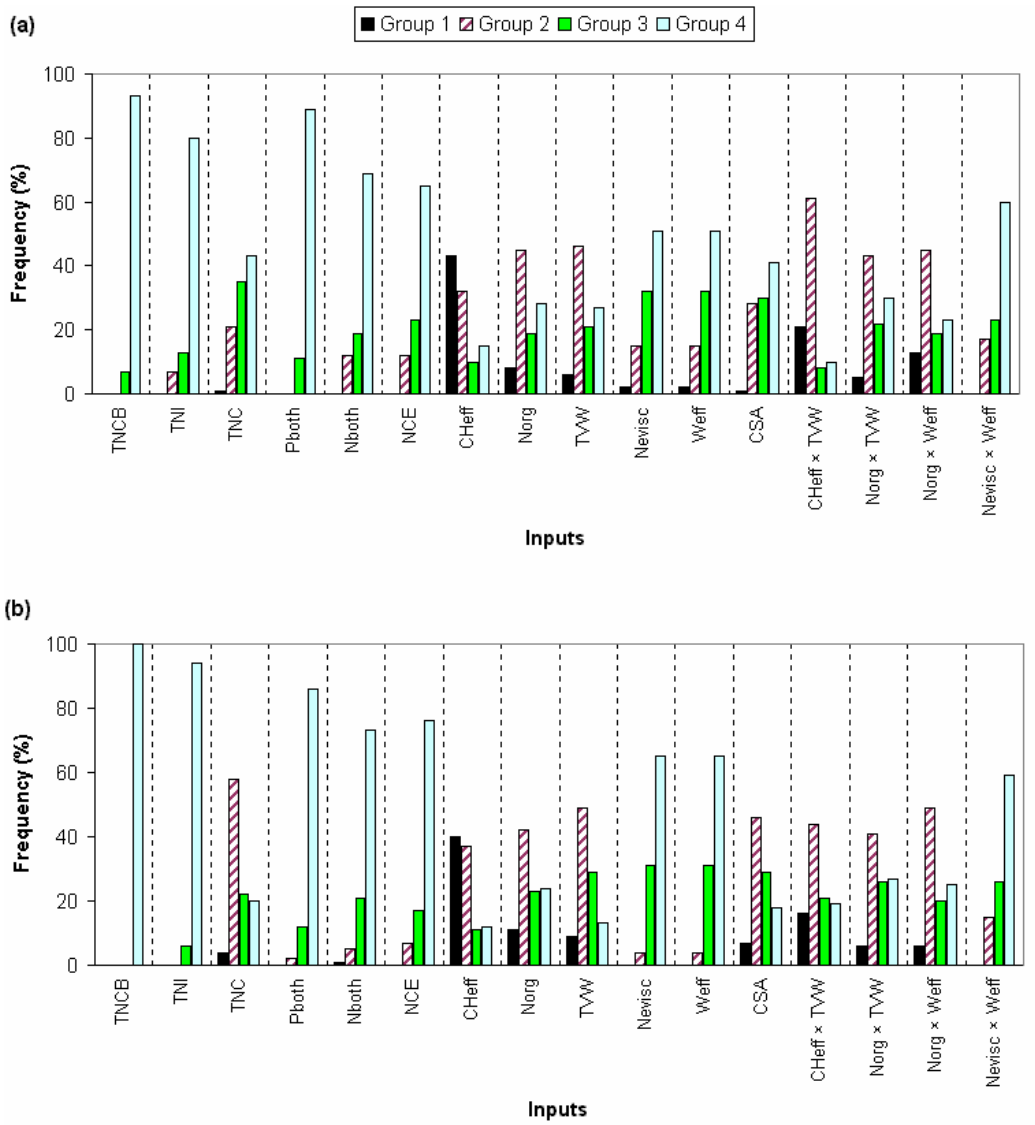


Figure 4. Frequencies Associated with Identification of Each Input in an Importance Group Based on: (a) Sample Regression; and (b) Rank Regression. Group 1: The Most Important Input; Group 2: Inputs with Secondary Importance; Group 3: Inputs with Minor Importance; and Group 4: Unimportant Inputs.

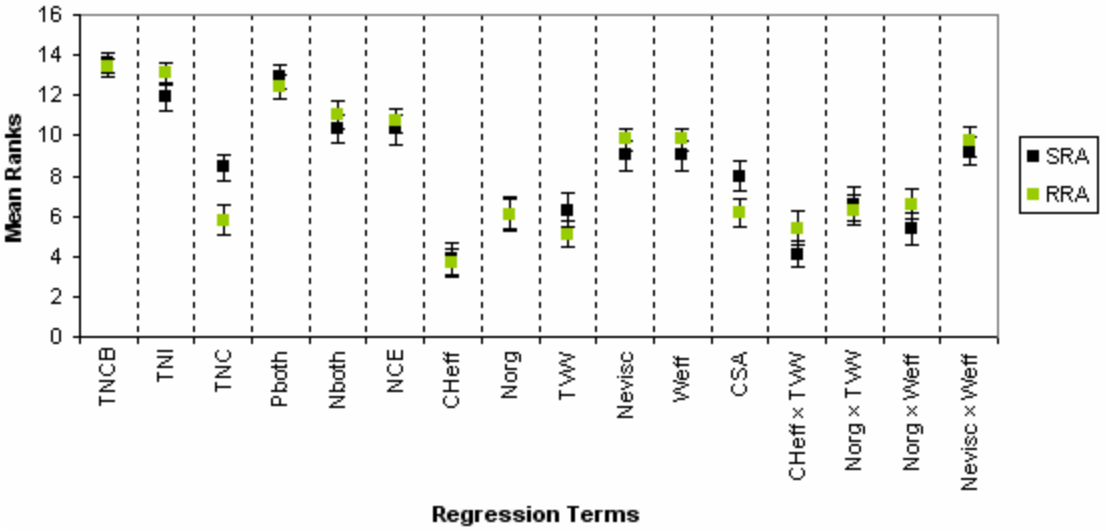


Figure 5. Mean Ranks and Corresponding 95 Percent Confidence Intervals Based on Sample Regression Analysis (SRA) and Rank Regression Analysis (RRA).

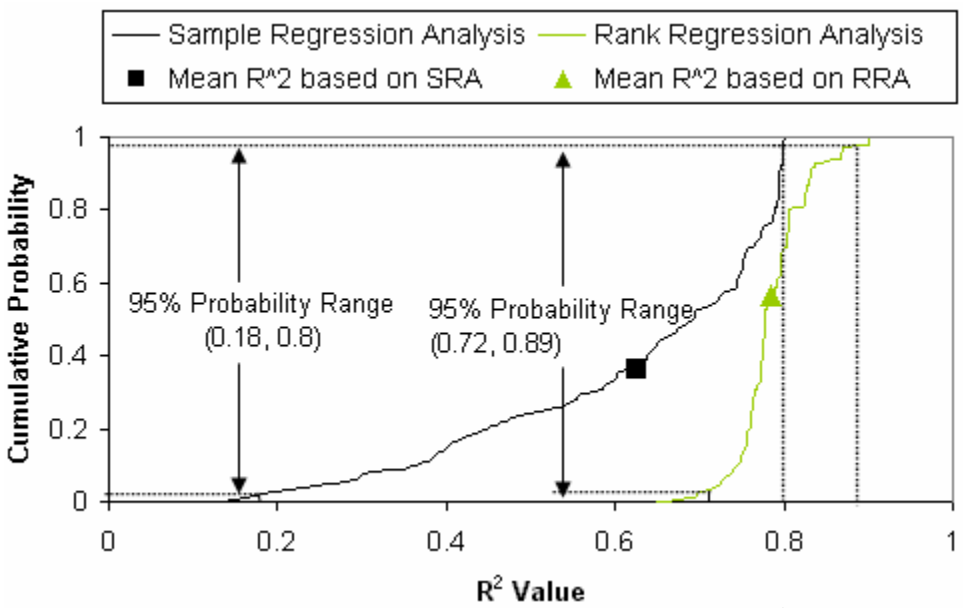


Figure 6. Cumulative Probability for Coefficient of Determination, R^2, in Sample Regression Analysis (SRA) and Rank Regression Analysis (RRA).

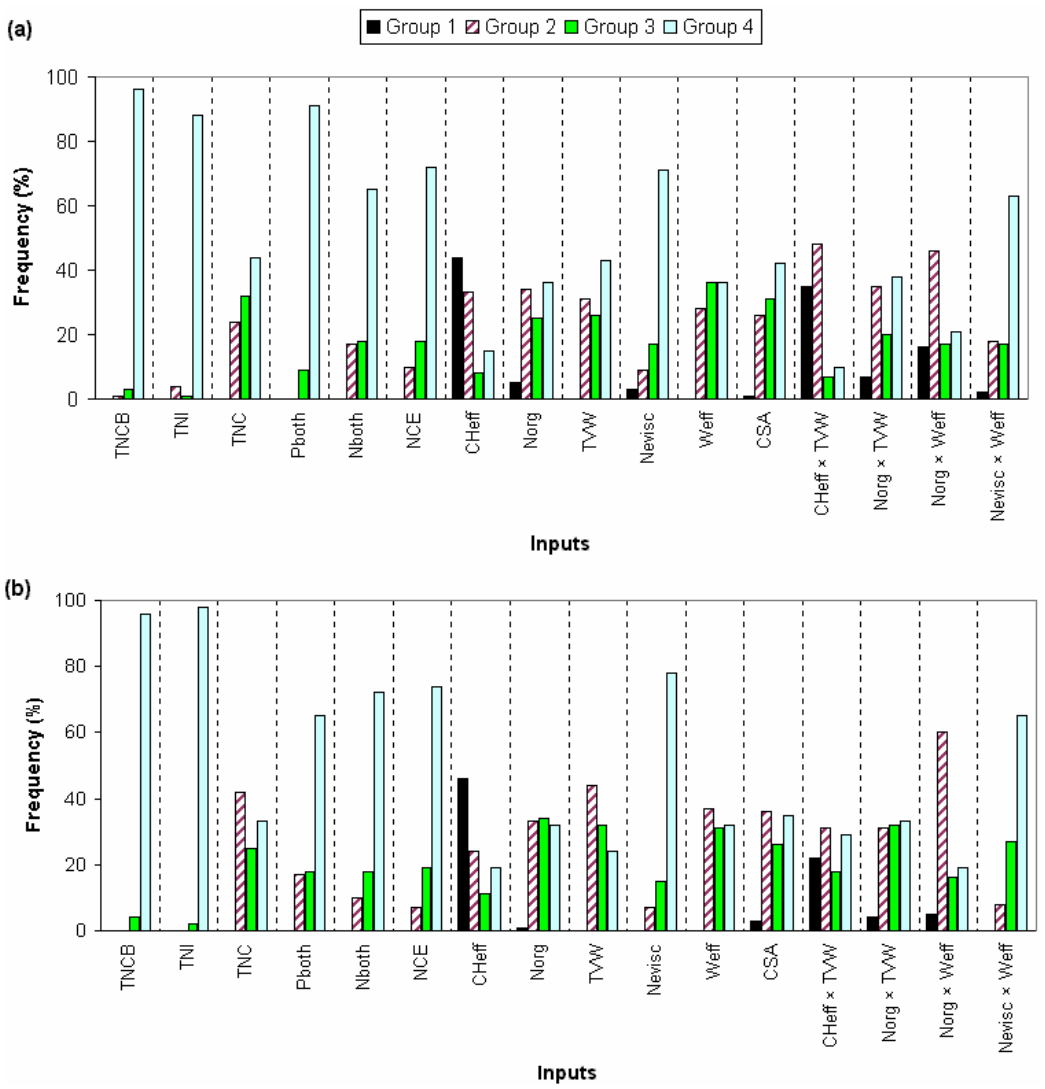


Figure 7. Frequencies Associated with Identification of Each Input in an Importance Group Based on: (a) Sample Stepwise Regression; and (b) Rank Stepwise Regression. Group 1: The Most Important Input; Group 2: Inputs with Secondary Importance; Group 3: Inputs with Minor Importance; and Group 4: Unimportant Inputs.

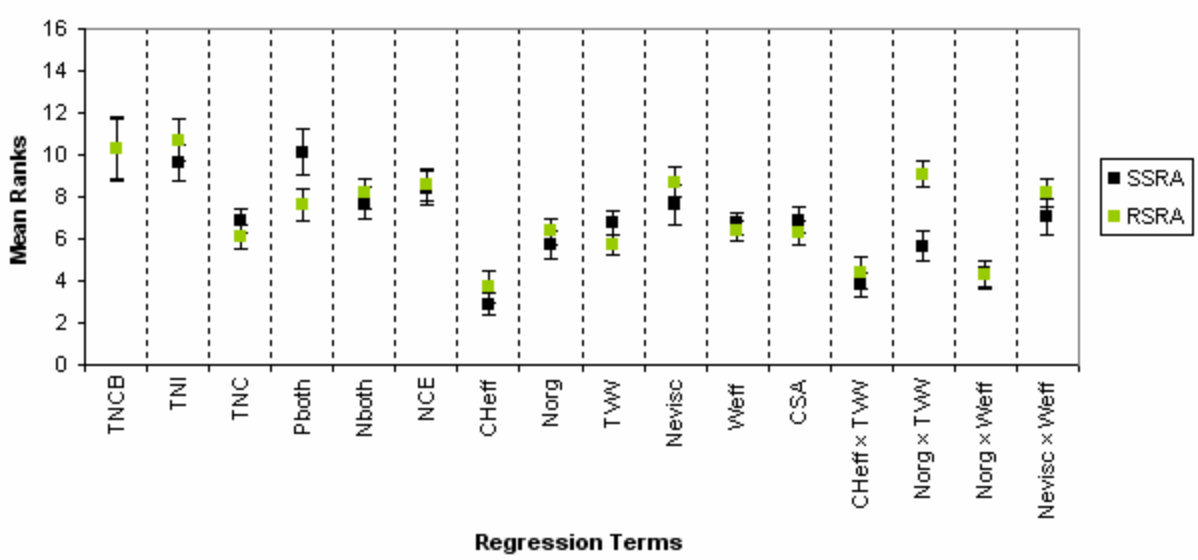


Figure 8. Mean Ranks and Corresponding 95 Percent Confidence Intervals Based on Sample Stepwise Regression Analysis (SSRA) and Rank Stepwise Regression Analysis (RSRA).

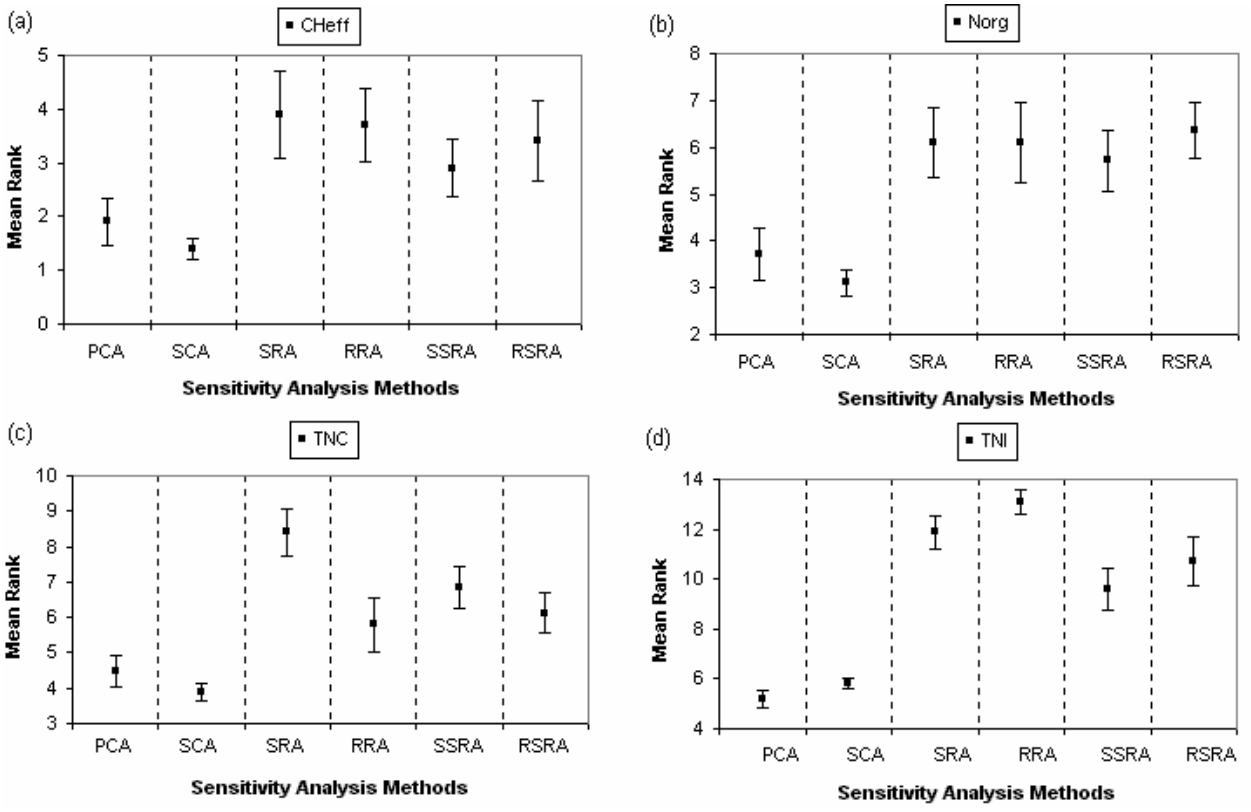


Figure 9. Mean Ranks and Corresponding 95 Percent Confidence Intervals Based on Pearson Correlation (PCA), Spearman Correlation (SCA), Sample Regression (SRA), Rank Regression (RRA), Sample Stepwise Regression (SSRA), and Rank Stepwise Regression (RSRA) Analyses for: (a) CHeff; (b) Norg; (c) TNC; and (d) TNI.

PART VII

CONCLUSIONS AND RECOMMENDATIONS

7.0 Summary and Conclusions

In this part, conclusions drawn from the associated research work and the results are presented. Key questions that are raised in Section 1.1 of Part I are addressed. Suggestions for future research directions related to this dissertation are provided.

7.1 When Should Sensitivity Analysis be Performed?

The usefulness of sensitivity analysis depends on the overall assessment objectives of a model under study. The overall assessment objectives should be clearly laid out before the model is built. Once the objectives are defined, the model must be designed to address these objectives. For example, if the objective is to develop insight into possible risk management strategies, variables that can be controlled and variables related to risk management objectives must be incorporated into the model. Next, the answers sought from application of sensitivity analysis should be clearly listed. The usefulness of sensitivity analysis can then be assessed based on whether the available methods of sensitivity analysis can address the questions under consideration in a manner that is appropriate to the characteristics of the model.

The case studies presented in this dissertation were basically focused on prioritizing critical control points in a model, identifying key sources of variability with quantification of uncertainty in those estimates, and model refinement. Application of sensitivity analysis for model verification and validation is discussed but not demonstrated. However, sensitivity analysis also can be used for conditional analysis of a model. Conditional analysis features "what-if" scenario analysis of a model and can focus on identification of factors contributing to high exposures and risks. In "what-if" scenario analysis, specific goals with respect to the risk mitigation can be modeled. Sensitivity analysis provides a tool to evaluate how these goals can be achieved by identifying key inputs and model assumptions contributing most to the

predefined scenario. Through this approach, the analysis can be framed in a way that is more responsive to the public's concerns and interests, thereby facilitating public review of the analysis. Furthermore, sensitivity analysis can provide explicit insight into the combination of key values and/or ranges of inputs that lead to the worst, or best, outcomes. The identification of worst case scenarios with respect to the exposure or risk is important for identifying possible approaches to mitigate exposure and/or risk.

As an example of a "what-if" scenario, risk managers may be interested in decreasing the mean exposure level by a specific amount. The "what-if" scenario may be in the form "what if we want to decrease the mean exposure level by one log?" Sensitivity analysis allows risk assessors and managers to consider cases where the risk can be reduced by controlling the input domains. For example, in order to constrain the mean growth of the pathogen to an acceptable level with respect to a particular management strategy it may be necessary to hold storage temperature below a specific limit, while holding storage time within a specific range. Deviation of storage temperature and storage time from the identified domains may lead to a substantial increase in the number of pathogen organisms due to growth. Thus, constraining the inputs to specific ranges can keep the risk within acceptable levels.

7.2 What are Key Considerations in Preparation of a Model to Facilitate Sensitivity Analysis?

Part II focused on providing guidance on how to facilitate sensitivity analysis when working with an existing model or a new model, with a focus on food safety process risk models. For an existing model, in order to determine appropriate methods and scope for performing a sensitivity analysis, it is necessary to clearly identify model structure, model inputs, internal inputs used in models associated with binning and aggregation of model output of interest, and the type of probabilistic simulation. For a new model, recommendations that can facilitate

sensitivity analysis involve the modeling environment, design of the probabilistic simulation, modeling strategies, and documentation of models.

7.3 How Can One Select Among Typical Sensitivity Analysis Methods?

The performance of a sensitivity analysis method depends on the model structure. The difficulty of the analysis increases when the model has characteristics such as non-linearity, interaction, non-monotonicity, and non-additivity. A modeler typically does not have *a priori* knowledge about the behavior of the model with respect to these characteristics, and hence, the preferred sensitivity analysis method would be the one that can address all of these characteristics. Moreover, the number and the characteristics of the inputs may affect the selection of a sensitivity analysis method. Considering the dimensionality of the input space, the larger the number of inputs, the more expensive and difficult the analysis, and hence, lower tendency to select sophisticated sensitivity analysis methods. Based on our experience with the case studies, any increase in either the number of inputs and/or their ranges of variation increases the possibility of non-negligible interaction effects between inputs.

The selection of sensitivity analysis methods depends on factors such as objective of the analysis, characteristics of the model under study, amount of detail expected from sensitivity analysis, characteristics of the software used for sensitivity analysis, and available computing resources. All the above reasons make the choice of the sensitivity analysis method a very delicate step. Although a ‘universal recipe’ does not exist, we provided a decision framework summarizing the discussions regarding selection of appropriate sensitivity analysis methods. This decision framework can substantially facilitate the process of selecting a sensitivity analysis method for a practitioner.

7.4 What are the Considerations in Presenting Results of Sensitivity Analysis Methods?

The general principles for presentation of sensitivity analysis results include: (1) clearly identify the target audience and tailor the presentation to that audience; (2) explicitly convey the objectives of the analysis; (3) describe the scenario so that the scope of the analysis is clear; (4) describe the model at an appropriate level of detail for the target audience; (5) describe the rationale for the selected sensitivity analysis methods; and (6) clearly present the results of the sensitivity analysis.

The purpose of communicating the sensitivity analysis results to an audience is to ensure the proper and effective exchange of information and insights from the results of the analysis. The audience may include analysts, stakeholders, media, decision-makers, or some combination of these. The audience may include representatives of government agencies, private companies, non-profit organizations, academic institutions, or other organizations. Furthermore, the audience may include persons with a range of academic backgrounds and experience, including (for example) liberal arts, food science, microbiology, toxicology, sociology, economics, medicine, engineering, law, or others. Thus, there may be substantial variability in the background and perspectives among members of the intended audience.

The methods for presenting and interpreting results may be different for different audiences. Therefore, a basic principle for communicating sensitivity analysis results is that the audience should be clearly identified. For example, when presenting the sensitivity results to an audience with a limited scientific background, extensive information regarding the technical details of the methodology is probably not helpful or could be counter-productive. The time available with such an audience could be better spent conveying the objectives, scenarios, general characteristics of the model, general approach used for sensitivity analysis, and presentation of the sensitivity analysis results in plain language with the use of some supporting

graphics. Qualitative information may be more useful than quantitative information for such audiences. If the objective is to provide insight into priorities for CCP and the selection of critical control limits, then the presentation could convey the relative importance of each possible CCP in either qualitative terms or using a simplified sensitivity index.

If the target audience is primarily other analysts, or persons with a strong technical background, then it is necessary to provide the most important technical details regarding the methods used, the models used, input assumptions, and interpretation of results, in addition to the basic qualitative information regarding objectives and scenarios. Moreover, the results of the analysis must be reproducible. However, jargon should be used with care. Even for a technical audience, it is likely that not everyone is familiar with the same jargon because each person may have a different technical background. Thus, any terminology unique to sensitivity analysis should either be avoided or will need to be clearly defined for the benefit of the audience.

An analyst should ensure that the primary objectives of the sensitivity analysis are clearly communicated to the target audience. As a rhetorical device, it is often helpful to state the objectives in the form of key questions that will be answered by the results of an analysis. For example, one can state that the objective of the analysis is to answer one or more of the following key questions:

- Which controllable model inputs contribute the most to variation in exposure and risk?
- Which of the potential critical control points are the highest priorities in terms of avoiding the highest exposures?
- What critical limits are likely to achieve the specified risk management objectives?

- What are the key sources of uncertainty in the analysis?
- Which uncertainties are the highest priorities for reduction based upon additional data analysis or research?
- How well or appropriately does the model respond to changes in input values?

It is necessary to clearly describe the scenario of the sensitivity analysis when presenting the results to audience since the results may be specific to the particular assumptions made in a scenario. Clear introduction to the scenario will help the audience understand the implications of the results. For example, one may want to specify to the audience whether the insights from the sensitivity results can be generalized.

A description of the model used is helpful for the audience. Summarizing the key model characteristics can help the audience better understand the reasons for application of the selected sensitivity analysis methods and evaluate the credibility of the results. The description of the model should include: (1) model characteristics; and (2) model inputs and the output of interest. Model characteristics include linearity, thresholds, interactions, and information regarding the probabilistic framework of the model. Probability distributions of the inputs should be provided including specification of whether the input represents variability, uncertainty, or both. Model outputs should be clearly presented to the audience.

Presentation of the rationale for selection of the sensitivity analysis method(s) will help establish their credibility, relevance, and validity. For example, when a model is non-linear and non-monotonic, the analyst can justify why ANOVA was selected while other sensitivity analysis methods such as correlation analysis were not used.

7.5 Can Conventional Sensitivity Analysis Methods Provide Robust Insights in Spite of their Apparent Limitations?

Conventionally used sensitivity analysis methods can be broadly classified into local sensitivity analysis methods and "push-button" methods. The local sensitivity analysis methods concentrate on the local impact of the inputs on the model and typically is carried out by computing partial derivatives of the output function with respect to the inputs. Methods classified as "push-button" are typically built-in features of commonly used software tools such as @Risk™ and Crystal Ball™. Under pressure of deadlines and faced with constraints of an existing model, an analyst may prefer to use these methods rather than attempt to implement more refined techniques. However, a key question is whether such methods are adequate for the intended case study.

Local sensitivity analysis typically includes nominal range sensitivity analysis, differential sensitivity analysis, and automatic differentiation. Derivatives of the model output with respect to each input are considered as sensitivity indices. Typically, each sensitivity index is normalized by ratio of standard deviation of the input to the standard deviation of the output. However, these method do not perform properly when the inputs have a dependency structure (e.g., they are correlated) or the model is non-linear.

Push-button methods typically include sample (Pearson) and rank (Spearman) correlation coefficients applied in the context of a one-dimensional probabilistic simulation. These methods are applicable to global sensitivity of linear or monotonic models, respectively. Sample correlations are not a robust method for dealing with thresholds or interactions and may not provide useful insights when applied to categorical inputs. Rank correlations are more robust than sample correlations with regard to thresholds but are not capable of providing insight regarding the existence of a threshold or the value of the threshold. If an analyst chooses to

proceed with a push-button method whose theoretical basis differs from the key characteristic of the model to which it is applied, then the accuracy and robustness of the results cannot be assured.

In the long-run, it will be helpful to the food safety risk community if add-ins can be developed, either as shareware or commercially, that incorporate sensitivity analysis methods that are more compatible with the characteristics of the models.

7.6 How Can One Discriminate the Importance Between Closely Ranked Inputs?

A key question in using sensitivity indices for ranking inputs based on a selected method is regarding how much the sensitivity indices of two inputs must differ in order to infer that the two inputs are substantially different with regard to sensitivity. Some sensitivity analysis methods provide analytical solutions for quantifying statistically significant differences in the sensitivity indices. For example, for the regression coefficients as sensitivity indices in regression analysis, the 95 percent confidence intervals identify the possible range of variation for each of the coefficients. The confidence intervals are estimated based on the assumption of normality in their corresponding distributions. Thus, inputs with overlapping confidence intervals associated with their regression coefficients have comparable importance.

For the cases that analytical solutions are not available, possible ranges of variation for sensitivity indices can be estimated using several simulations of the model with alternative random numbers. Each simulation of a model may include multiple variability iterations and/or uncertainty realizations. Alternative simulations of a model typically use different sets of random numbers for generating samples from probability distributions of the inputs. This typically leads to different estimates of the sensitivity index values for each input because of the random sampling error. In order to quantify the random sampling error in the sensitivity indices, if time

and resource are available, multiple model simulations should be run (e.g., 1000 simulations) using different sets of random numbers for generating values from probability distributions of inputs. If the selected sensitivity analysis method is applied to each simulation, a single sensitivity index would be estimated for each input, and hence, for each of the 1000 simulations. Thus, 1000 sensitivity indices would be estimated for each input. Each of the estimated sensitivity indices would typically differ because of random sampling error. A summary of the distribution of the sensitivity index values for each input, such as a 95 percent probability range, can help to quantify the possible range of variation in the sensitivity indices. When compared with uncertainty in the sensitivity index values for other inputs, it is possible to infer whether the rankings of the sensitivity of each input would be unambiguous or not.

Usually it is not practical to run the model for hundreds of simulations as each simulation may take a few hours. As an alternative, the empirical bootstrap simulation is introduced to generate sampling distributions of uncertainty for sensitivity index values. Using the empirical bootstrap approach, an alternative randomized version of the original Monte Carlo simulation is obtained by sampling with replacement from the original set of random values. This procedure is computationally faster than generating new random samples from the original specified probability distributions of the model inputs. Selected sensitivity analysis method should then be applied to the bootstrap samples to produce a distribution of the sensitivity index values.

7.7 What are the Implications of Sensitivity Analysis in a Two-Dimensional Probabilistic Framework Compared to the One-Dimensional Analysis of Variability and Uncertainty?

Because of the theoretical preference in correctly distinguishing between variability and uncertainty, two-dimensional analyses of variability and uncertainty are typically favored over one-dimensional approaches in which variability and uncertainty are treated individually or co-

mingled into a single dimension. The prediction of the exposure or risk may depend on the way that uncertainty and variability is separated, and that a high end exposure or risk may be overlooked if the distinction between uncertainty and variability is neglected. The estimation of sensitivity in a two-dimensional simulation provides key insights regarding the robustness of results and, hence, the confidence that can be placed in the rankings. This insight is needed to support decision making under uncertainty. The range of ranks and probability distribution of ranks of an input is of valuable importance to the decision maker. However, sensitivity analysis in a two dimensional simulation is potentially computationally intensive, and as a practical matter, it is difficult to apply sensitivity analysis methods in a two dimensional framework unless they can be automated. For the demonstrated case studies, the automation was done using SAS© macros.

A comparison of one-dimensional and two-dimensional results implied that the top ranked input was the same for each of the probabilistic simulation approaches. Furthermore, these two approaches identified approximately the same groups of inputs with comparable importance. Thus, it may be possible to develop a robust insight regarding key inputs using a simplified one-dimensional probabilistic simulation approach if resources do not permit a two dimensional approach. Alternatively, as part of model development, sensitivity analysis can initially be performed in a one-dimensional analysis to identify priorities for data collection. However, although the results obtained in the specific case studies in this paper may be indicative, they are not definitive. Thus, where resources permit, a two-dimensional sensitivity analysis is generally preferred unless a one-dimensional analysis is appropriately justified.

8.0 Recommendations for Future Work

Key recommendations of this study are:

(1) The computational time was identified as a major drawback for application of sensitivity analysis methods to the selected MFSPR model. Typically, for providing a dataset including sampled values from probability distributions of inputs and corresponding output values with reasonable sample size substantial amount of time was required. In order to reduce the computational burden of sensitivity analysis with respect to time, it is recommended that distributed runs of a selected model be practiced. In order to perform a distributed run, a simulation of size n (e.g., 100,000) can be broken into several simulations with smaller sample sizes (e.g., 10 simulations with size 10,000). Each new simulation should be run on a different computer simultaneously. However, in this case enough resources with respect to computer hardware should be available. This approach can reduce the computational time substantially.

(2) The methods evaluated in this study should be applied to other quantitative risk assessment models, especially the models with temporal and spatial dimensions in exposure and risk. Such models consider accumulative exposure for each individual, and hence, the task of identifying key sources of variability and uncertainty and corresponding critical control limits would be challenging. An example includes Stochastic Human Exposure and Dose (SHEDS) models developed by EPA.

(3) This dissertation did not evaluate variance-based sensitivity analysis methods such as FAST and Sobol via application to some case studies. Although the idea of using variance as an indicator of importance for inputs is not new, these methods have not been widely used with quantitative risk assessment models and specifically in

microbial food safety due to their complexity and lack of available framework for application. However, variance-based sensitivity analysis methods have been widely used in reactor safety and nuclear engineering as promising methods for sensitivity analysis. Variance-based sensitivity analysis methods should be evaluated via case studies with MFSPR models and the results compared with those from methods used here in order to better understand how alternative measures of sensitivity may lead to different outcomes.

(4) Due to special characteristics of the model considered as the testbad for the case studies, there was no dependency structure between inputs. However, for cases that inputs have dependency structures such as correlation, introduced sensitivity analysis methods will fail to provide valid results unless special attention is considered to quantify and incorporate such dependencies in the analysis. Introduced sensitivity analysis methods should be evaluated with respect to their functionality in the case of inputs with dependency. Approaches should be explored for quantification of different dependency structures between inputs and possible modification of sensitivity analysis methods for such cases.

(5) A framework for exposure and dose response phases of risk assessment models including modules for application of sensitivity analysis should be developed. Current risk assessment frameworks do not incorporate sensitivity analysis. Hence the application of sensitivity analysis on existing models is often tedious as the models do not follow a combined modeling framework. The framework should be provided through adapting, extending, and combining methods for risk analysis developed within the food safety domain, semi-formal description methods (in particular,

methods for object oriented modeling), and computerized tools. The framework should be tested and assessed on a real food safety model.

APPENDIX *A*

GLOSSARY OF TECHNICAL TERMS

Add-In Software Packages "Add-in" is a term used, especially by Microsoft, for a software utility or other program that can be added to a primary program. Examples of add-in software packages are Crystal Ball™ and @Risk.

Aggregation Refers to situations in which multiple numerical values are combined into one numerical value, such as sum or mean value. In some cases, aggregation reflects the underlying process.

Ambiguity of Ranking Degree of uncertainty in estimated rankings of importance of model inputs from a sensitivity analysis method because of uncertainty in the sensitivity index or lack of significant differences in the sensitivity index of the inputs being jointly evaluated.

ANOVA Analysis of Variance, a statistically-based method that can be used for sensitivity analysis.

Binning A summarization technique in which values of an internal input or output of interest are binned into predefined intervals. Binning leads to loss of one-to-one correspondence between an output of a module and its exogenous predecessor inputs. If binning exists in a model, it can affect the scope of sensitivity analysis.

Bootstrap simulation A numerical method for estimating sampling distributions and confidence intervals for statistics based on an assumption that the "true" population distribution is equal to the empirical sample distribution.

Branch Term used in CART to refer to the conditions on the input variables that determine which output data go into particular data strata in a regression tree.

CART Classification and Regression Tree, a statistically-based method that can be used for sensitivity analysis. Also known as Hierarchical-Based Regression Tree (HBTR).

Categorical Input Also referred to as a qualitative input. This is a specific type of inputs with values that are mutually exclusive. The categories are measured in nominal scale and cannot be ordered from "highest" to "lowest" in any meaningful way. Rather, the categories define different groups. Examples include gender, season, and consumption type.

CCP Critical Control Point (CCP). A CCP is defined as a point, step, or procedure at which control can be applied and food safety hazard can be prevented, eliminated, or reduced to an acceptable level.

CDF The cumulative distribution function for a variable indicates the probability that the variable is less than or equal to any particular value.

CCDF Complementary cumulative distribution function for a variable indicates the probability that the variable is greater than any particular value.

Classification Rule With respect to CART, a result obtained from CART that enables stratification of sample values of the output to a specific situation determined by specific cut-points for specific inputs. For example, a classification rule could provide insight into conditions that lead to the largest mean exposure.

Coefficient of Determination (R^2) A statistic that is widely used as a goodness-of-fit measure for regression analysis and ANOVA. R^2 represents the fraction of variation in the output that can be explained by the fitted regression or ANOVA model.

Conditional Analysis Conditional analysis includes the "what-if" scenario analysis of a model and also identification of factors contributing to high exposures and risks.

Confidence Interval The computed interval with a given probability (i.e., 95%) that the true value of the statistic, such as a mean, proportion, or rate, is contained within the interval.

Contrast A comparison between two levels (or two sets of levels) of an independent variable in Analysis of Variance (ANOVA).

Correlation Coefficient Sample correlation coefficients provide a measure of linear association between two variables. Rank correlation coefficients provide a measure of monotonic association between two variables.

Critical limit A term used in HACCP to refer to a criterion or range that must be met for each preventive measure associated with a CCP, such as minimum internal temperature, product dimension, and cooking time.

CSA Conditional Sensitivity Analysis is a graphical method, in which the sensitivity of the model to a small number of inputs is evaluated while other inputs are held at fixed values.

Cut-Point With respect to CART, a value of an input used as the basis for creating branches.

Discrete Input An input variable that can only assume certain integer values. An ordinal categorical input variable could also be called discrete variable.

DSA Differential Sensitivity Analysis, a mathematical local sensitivity analysis method based upon small perturbations of inputs at a specific point in the model space.

F Value The F value is the ratio of two sums of squares (i.e. estimates of a population variance, based on the information in two or more random samples). When employed in ANOVA, the obtained F value provides a test for the statistical significance of the observed differences among the means of two or more random samples.

Factor An independent variable or input in ANOVA.

Factorial Design An experimental design in which each level of a variable is paired with each level of every other variables.

Food Safety Process Risk Model A quantitative model evaluating the risk to humans exposed to specific food borne hazard on a population basis. Food safety process risk models aim to simulate the process of bringing food from the farm to the table.

Global Sensitivity Analysis Global sensitivity analysis apportions the uncertainty in the output to the uncertainty in the input factors. Global methods are applicable to situations in which model inputs are varied simultaneously over large ranges of values, typically based upon probability distributions assigned to each input.

Graphical Sensitivity Analysis Methods A category of sensitivity analysis methods that feature the use of graphical techniques to present how a model output responds to changes in a model input. Typical graphical techniques include scatter plots or conditional sensitivity analysis.

HACCP Hazard Analysis and Critical Control Points (HACCP). HACCP is a management system in which food safety is considered through analysis and control of biological, chemical and physical hazards including raw materials, handling, production, distribution and consumption of the final product**.**

Important Inputs. An input to which the model output is highly sensitive or that leads to an outcome of most interest to a decision maker.

Interaction A case in which the effect of an input to the model response depends on the value of another input. For example, interactions can be represented in regression and ANOVA in the form of cross-product terms, such as $X_i \times X_j$, where X_i and X_j are inputs to the model.

Intermediate Node With respect to CART, an intermediate node refers to a node at which the data can be successively split in a regression tree.

Internal input A variable that is an output from a predecessor module and an input to a successor module for modules that are in series.

Leaf Node With respect to CART, leaf node refers to the node at which the penultimate data was split in a regression tree. The leaf node is also referred to as a terminal node.

Level A concept used in ANOVA to refer to specific range of values for a factor.

Local Sensitivity Analysis Local sensitivity analysis focuses on the impact of changes in input values with respect to a specific point in the input domain. NRSA and DSA are examples of local sensitivity analysis methods.

Mathematical Sensitivity Analysis Method A category of sensitivity analysis methods that are typically applied to a deterministic mathematical model. Examples include NRSA and DSA.

Measure of Sensitivity An index by which sensitivity of model output to inputs can be prioritized or by which insight regarding the sensitivity of the output to inputs can be obtained. Examples include standardized regression coefficients in regression analysis and F values in ANOVA. Also referred to as sensitivity index.

Model Independent A sensitivity analysis method is said to be model independent if it does not require any specific assumption regarding the functional form of the model to which it is applied.

Model Validation Model validation is the comparison of model results to independent observations from the system which is being modeled.

Model Verification A process of ensuring that the mathematical structure of a model, its computer implementation, and its input assumptions are as intended.

Monotonic Relationship A relationship that occurs when there is an association between two variables that is either consistently increasing or consistently decreasing. For example, the degree of monotonic relationship is measured by the Spearman Correlation Coefficient.

Monte Carlo Simulation A numerical random sampling method for generating a representative distribution of values for each of the inputs in a generic exposure (or dose) equation to derive an output distribution of exposure (or dose) in the population.

Node With respect to CART, a node is an input variable at which data can be split in a regression tree.

Nominal Value A selected point value of a distribution, such as a minimum, mean, median, or maximum value.

Non-linearity A relationship between two variables in which the change in each variable is not simply proportional to the change in the other variable. Parabolic and exponential relationships are examples of non-linearity. Non-linearity may be diagnosed from bivariate scatter plots.

NRSA Nominal Range Sensitivity Analysis – a mathematical technique in which one input at a time is varied over a range from low to high, while other inputs are held constant at a nominal value.

P Value The probability of obtaining a more extreme value for a test statistic under the assumption that the null hypothesis is true. The P value is compared with the chosen significance level of the test and, if it is smaller, the result is said to be significant.

Path or Classification Rule With respect to CART, a set of conditions on the input variable from the root node leading to a leaf node in a regression tree.

Pearson Correlation Coefficient A type of correlation coefficient that evaluates the strength of linear association between paired random output samples of output and input values. It is also known as the sample correlation coefficient.

Probabilistic Analysis A numerical analysis in which frequency (or probability) distributions are assigned to represent variability (or uncertainty) in model inputs and random sampling techniques are used to generate model input samples that are then propagated through the model to obtain a sample from the distribution of an output variable(s) of interest.

"Push-Button" Techniques A term used to refer to sensitivity analysis techniques that are built-in features of commonly used software packages. For example, the ability to calculate sample correlation coefficients is a built-in feature of Crystal Ball™.

Qualitative Input See categorical input.

Quantitative Input An input that takes numerical values. Quantitative inputs can be continuous or discrete. Examples include income, age, temperature, and time.

Rank Correlation Coefficient See Spearman correlation coefficient.

Realization A realization refers to one model simulation based upon one randomly sampled value for each probabilistic input.

Refined Analysis An analysis that is intended to be more comprehensive and/or accurate than a *screening analysis*. A refined analysis typically requires greater resources of time, effort, and computation than a screening analysis, but is expected to provide more detailed and robust insights.

Response Surface Method A sensitivity analysis method that develops a simplified version of the original model so that it is possible to retain the key characteristics of the model and to shorten the amount of time required to predict the output for a given set of inputs.

Root Node With respect to CART, the first node at which data are split in a regression tree.

RTE Ready-to-Eat food. A food that can be consumed safely without further heat treatment, including reheating. Examples include deli salads, bagged salads, and hot dogs.

Sample Correlation Coefficient See Pearson correlation coefficient.

Sampling Distribution A probability distribution for a statistic, typically estimated based upon random sampling error. A sampling distribution is the basis for estimation of confidence intervals for a statistic.

Saturation point A saturation pint is a value for a model input above which there is no change in the output, but below which there can be substantial variation in the output. For example, the maximum population density.

Scatter Plot A figure in which individual data points are plotted in two-dimensional space.

Scenario A set of facts, assumptions, and inferences about a problem of interest that helps an analyst in evaluating, estimating, or quantifying the problem of interest.

Scenario Uncertainty Uncertainty reflecting that a defined scenario may fail to consider all the factors and conditions contributing to the variation of the output. In the case of the food safety process risk models, scenario uncertainty is typically associated with missing or incomplete information needed to fully define exposure and dose.

Screening Sensitivity Analysis Can be used to identify subset of inputs that controls most of the output variability with low computational effort especially in models that are

computationally extensive and have a large number of inputs. Screening methods are typically less accurate than refined methods and may be used in tiered approach.

Sensitivity Analysis The assessment of the impact of changes in input values or assumptions on generated output values.

Sensitivity Index A quantitative measure of the degree of association between changes in the values of a model output relative to changes in the values of a model input

Simulation A series of calculations that attempt to predict a value or outcome. Simulation is useful for exploring outcomes that are not observable under the conditions of interest, or an outcome whose occurrence is potentially dangerous.

Spatial Dimension of a Scenario The spatial considerations for a scenario including the relevant geographic area. Depending upon the level of detail of the model, it may be necessary to incorporate multiple scales of geographic information.

Spearman Correlation Coefficient A measure of the strength of the monotonic relationship between two random variables. Thus, it is applicable to situations where there is a monotonic nonlinear relationship. It is also known as the rank correlation coefficient.

Statistical Sensitivity Analysis Methods A category of sensitivity analysis methods that are typically applied to probabilistic models for which random samples are generated for model inputs and corresponding samples are estimated for model outputs. Examples of statistical sensitivity analysis methods include correlation analysis, regression analysis, ANOVA, and CART.

Statistical Significance The satisfaction of a statistical criterion determined by a priori specification of an acceptable probability of wrongly rejecting a true hypothesis (i.e., usually 0.05 percent).

Threshold A value in an input domain below which a model output does not respond to changes in the input. An example is a specific storage temperature below which zero growth is estimated for an organism.

Temporal Dimension of a Scenario The temporal dimension of a scenario typically include: (1) the time for each major step in a process; (2) the activity patterns of individuals or agencies; (3) "seasonal" effects, whether at a short time scale (e.g., daily, weekly) or a longer scale (e.g., monthly, quarterly, annual); and (4) the time period associated with occurrence of illness as a result of one or more exposures.

Treatment With respect to ANOVA, a specific combination of levels for different factors. For example, in a factorial design with two factors that have n and m levels, respectively, $n \times m$ treatments can be specified.

Two-Dimensional Probabilistic Simulation A probabilistic simulation in which variability and uncertainty in model input are simulated separately.

Uncertainty Lack of knowledge about the "true" value of a quantity, lack of knowledge regarding which of several alternative models best describes a mechanism of interest, or lack of knowledge about which of several alternative probability density functions should represent a quantity of interest. Uncertainty is a property of an analyst.

Variability Heterogeneity of values over time, space, or different members of a population. Variability is a property of nature.

What-If Scenario Analysis A type of conditional analysis in which specific goals with respect to risk mitigation can be established. As an example of a "what-if" scenario, risk managers may be interested in decreasing the mean exposure level by a specific amount.