UNCERTAINTY ANALYSIS OF TRANSPORT-TRANSFORMATION MODELS

BY SASTRY S. ISUKAPALLI

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Chemical and Biochemical Engineering

Written under the direction of

Dr. Panos G. Georgopoulos

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

January, 1999

ABSTRACT OF THE DISSERTATION

Uncertainty Analysis of Transport-Transformation Models

by Sastry S. Isukapalli

Dissertation Director: Dr. Panos G. Georgopoulos

Characterization of uncertainty associated with transport-transformation models is often of critical importance, as for example in cases where environmental and biological models are employed in risk assessment. However, uncertainty analysis using conventional methods such as standard Monte Carlo or Latin Hypercube Sampling may not be efficient, or even feasible, for complex, computationally demanding models.

This work introduces a computationally efficient alternative method for uncertainty propagation, the Stochastic Response Surface Method (SRSM). The SRSM approximates uncertainties in model outputs through a series expansion in normal random variables (polynomial chaos expansion). The unknown coefficients in series expansions are calculated using a limited number of model simulations. This method is analogous to approximation of a deterministic system by an algebraic response surface.

Further improvements in the computational efficiency of the SRSM are accomplished by coupling the SRSM with ADIFOR, which facilitates automatic calculation of partial derivatives in numerical models coded in Fortran. The coupled method, SRSM-ADIFOR, uses the model outputs and their derivatives to calculate the unknown coefficients.

The SRSM and the SRSM-ADIFOR are general methods, and are applicable to any model with random inputs. The SRSM has also been implemented as a black-box,

web-based tool for facilitating its easy use.

The SRSM and the SRSM-ADIFOR have been applied to a set of environmental and biological models. In all the case studies, the SRSM required an order of magnitude fewer simulations compared to conventional methods, and the SRSM-ADIFOR required even fewer simulations. In addition to their computational efficiency, these methods directly provide sensitivity information and individual contributions of input uncertainties to output uncertainties; conventional methods require substantially larger numbers of simulations to provide such information. Thus, the SRSM and the SRSM-ADIFOR provide computationally efficient means for uncertainty and sensitivity analysis.

Finally, this research addresses uncertainties associated with model structure and resolution with application to photochemical air quality modeling. A three dimensional version of the regulatory Reactive Plume Model (RPM), RPM-3D, has been developed and applied to understand model uncertainty.

Acknowledgements

I would like to express my sincere thanks and appreciation to my advisor, Panos G. Georgopoulos, for guidance, and for providing me with excellent facilities to pursue my work, and for ensuring financial support throughout my studies.

I am also grateful to the members of my committee for taking the time to guide me through my dissertation. I would like express my appreciation to Professor Alkis Constantinides for his insightful remarks, and to Professor Henrik Pedersen, and Professor David Kosson for their valuable comments and ideas. I am also grateful to Dr. Daewon Byun for his continued encouragement and support. I also thank Dr. Yee Chiew for introducing me to the concepts of Monte Carlo simulations, and for his insightful remarks on this thesis. I am also grateful to Dr. Amit Roy for his continued support, valuable guidance, and frank, critical comments throughout my research.

This work has benefited from various researchers. Special mention goes to Dr. Menner Tatang, whose work served as the starting point for this thesis. I would also like to acknowledge the following researchers and organizations who have provided models and input files for the evaluation of the methods developed here: Dr. Amit Roy for the PERC PBPK model and input data for the RPM model, the ozone modeling group (Dr. Shengwei Wang and Dr. Saravanan Arunachalam) for the input data for the UAM model, and Hydrogeologic Inc. for the input data for the EPACMTP model. I am also thankful to Dr. Ashwin Walia and Dr. Shengwei Wang for sharing their documentation on the UAM-IV and CB-IV (adapted here as Appendix C).

I am grateful to my colleagues at the CCL for their valuable discussions and suggestions. I am also thankful to Ms. Arlene Bicknell, Ms. Roberta Salinger, and Ms. Julie Greenman for their continued help. I thank the countless authors of all the free software that I have used during my research work – their selfless efforts have significantly aided

my work.

I would also like to acknowledge financial support provided by the U.S. EPA, NERL (National Exposure Research Laboratories), and U.S. DOE through CRESP (Consortium for Risk Evaluation with Stakeholder Participation).

Special thanks go to my friends, Ravi Bhupathiraju, Padma Isukapalli, and Ramana Isukapalli, for providing a constant source of encouragement and support, and for being there for me at all times.

This thesis would not have materialized but for the help of Amit Roy. As a friend, as a colleague, and as a mentor, he has helped me continuously, and has contributed tremendously to this thesis. I consider it as my fortune to be his friend, and to have worked with him.

Dedication

to my parents

and to Amit Roy

Table of Contents

List of Tables

List of Figures

Chapter 1
INTRODUCTION

1.1 Uncertainty Analysis

Mechanistic modeling of physical systems is often complicated by the presence of uncertainties. Environmental and biological modeling, for example, entails uncertainties in the estimates of toxicant emissions, transformation and transport parameters, etc., that impact the estimates of related health risks. The implications of these uncertainties are particularly important in the assessment of several potential regulatory options, for example, with respect to the selection of a strategy for the control of pollutant levels. Even though significant effort may be needed to incorporate uncertainties into the modeling process, this could potentially result in providing useful information that can aid in decision making.

A systematic uncertainty analysis provides insight into the level of confidence in model estimates, and can aid in assessing how various possible model estimates should be weighed. Further, it can lead to the identification of the key sources of uncertainty (such as data gaps) which merit further research, as well as the sources of uncertainty that are not important with respect to a given response.

The purpose of quantitative uncertainty analysis is to use currently available information in order for quantifying the degree of confidence in the existing data and models. The purpose is not to somehow "reduce" uncertainty – reduction in uncertainty can only come from gathering additional information and filling "data gaps". Even though the applicability of a model is limited by the model assumptions and the uncertainties in the evaluation data, understanding the judgments associated with the modeling process is more valuable than side-stepping the uncertainty analysis. In fact, it is precisely for problems where data are limited and where simplifying assumptions

have been used that a quantitative uncertainty analysis can provide an illuminating role, to help identify how robust the conclusions about model results are, and to help target data gathering efforts [72].

The following stages are involved in the uncertainty analysis of a model: (a) estimation of uncertainties in model inputs and parameter (*characterization of input uncertainties*), (b) estimation of the uncertainty in model outputs resulting from the uncertainty in model inputs and model parameters (*uncertainty propagation*), (c) characterization of uncertainties associated with different model structures and model formulations (*characterization of model uncertainty*), and (d) characterization of the uncertainties in model predictions resulting from uncertainties in the evaluation data.

1.2 Transport-Transformation Models

A wide range of environmental and biological models involve the fate and transport of a chemical species that originates from a source, travels through a medium ("transport"), undergoes changes due to chemical or radioactive processes ("transformation"), and eventually comes in contact with a receptor (e.g., a human, or specifically an organ, or a tissue).

In this thesis, the term "transport-transformation models" refers to a wide range of biological, environmental, and engineering problems. These models are similar in structure, often described by ordinary or partial differential equations. These models are based on the continuity equation, and on a mass-balance of chemical or radioactive species in a control volume. Even though the complexity of these models varies significantly, they are categorized together because the underlying physical and chemical processes, and the mathematical equations describing these systems, are similar.

The term "transport" refers to the movement of material through the surroundings as a result of associated flows and diffusive processes. The inhalation of chemicals, the movement of toxic substances along with the blood flow, or the absorption of pollutants by the layers of skin, are examples of transport. The flow of gaseous pollutants along with the wind, and the flow of contaminants through and along with surface and ground water are also examples of transport processes. Additionally, the diffusion of chemical

Table 1.1: Examples of transport-transformation components in biological and environmental models

	Component		
Model	Flow	Reaction	Source/Sink
Biological	Blood	Enzymatic	Absorption into tissues
Air Quality	Wind	Photochemical	Emissions/Deposition
Water Quality	Surfacewater/groundwater	Radioactive	Leachate

species across a concentration gradient is also a transport process.

The term "transformation" refers to the change of a species (physical, chemical, or biological). In biological systems,the enzymatic reactions account for transformations. In the atmosphere, transformation can result from a series of nonlinear photochemical reactions. In groundwater systems, transformation can sometimes result from radioactive processes, and sometimes due to chemical reactions.

Many transport-transformation models can be represented in a simplified form by the following equation:

$$\frac{dc}{dt} = F + R + S$$

where $\frac{dc}{dt}$ is the rate of change of a species concentration in a control volume under consideration, F is the net inflow of the species into the control volume, R is the net rate of production of the species due to chemical or radioactive processes, and S is the net rate of injection of the species into the control volume (also called as source/sink term). Table 1.1 presents examples of the components of some transport-transformation models of environmental and biological systems.

Uncertainty occurs in transport-transformation models due to a number of factors. The randomness inherent in natural systems, errors in the estimates of transport properties of a species in a medium, and inaccurate estimates of transformation properties, such as the reaction rates, contribute to the overall uncertainty in these models. The main sources of uncertainty in transport-transformation systems are presented in more detail in Chapter 2.

1.3 Limitations in Performing Uncertainty Analysis

The main limitation in performing comprehensive uncertainty analyses of transport-transformation models is the associated cost and effort. The computer resources required for uncertainty propagation using conventional methods (to be discussed in the following paragraph) can sometimes be prohibitively expensive. Further, the incorporation of uncertainty associated with structural and formulation aspects of the models requires significant effort. Finally, the data needs for characterizing input uncertainties are often substantial.

Conventional methods for uncertainty propagation typically require several model runs that sample various combinations of input values. The number of model runs can sometimes be very large, i.e., of the order of many thousands, resulting in substantial computational demands. On the other hand, in order to estimate the uncertainties associated with model formulation, several different models, each corresponding to a different formulation of the mathematical problem corresponding to the original physical system, have to be developed. The model results corresponding to all the possible combinations give an estimate of the range of the associated model uncertainty. Development and application of several alternative computational models can require substantial time and effort. Thus, the costs associated with uncertainty analysis may sometimes be prohibitively high, necessitating a large number of model simulations and/or the development of several alternative models.

1.4 Objective of the Thesis

The primary objective of this thesis is the development of computationally efficient methods for uncertainty propagation. This is addressed from the perspective of (a) computational requirements of the methods, (b) applicability of the methods to a wide range of models, and (c) ease of use of the methods.

Conventional methods for uncertainty propagation, such as the standard Monte Carlo and Latin Hypercube Sampling methods, may be computationally prohibitively expensive, as described in more detail in Chapter 2. The development of alternative methods that are applicable to a wide range of transport-transformation models could

substantially reduce the computational costs associated with uncertainty propagation (in terms of both the time and the resources required). In fact, such methods could facilitate the uncertainty analysis of complex, computationally demanding models where the application of traditional methods may not be feasible due to computational and time limitations. Additionally, there is a need for an extensive evaluation of the alternative methods with realistic case studies. Such an evaluation addresses the issues associated with the wider applicability of the methods. Further, the ease of use of any method is an important factor in its applicability. The extra effort required in understanding and applying a new method can sometimes offset other advantages that the method has to offer. Hence, attention must be paid to the development of auxiliary tools that facilitate the easy use of the new methods.

Thus, the primary objective of this thesis can be stated as

> "the development of computationally efficient alternative methods for uncertainty propagation that are applicable to a wide range of transport-transformation models, and the development of auxiliary tools that facilitate easy use of these methods."

Another objective of this thesis is to study the uncertainties associated with model structure and formulation, which fall beyond the scope of input uncertainty propagation. Such a study provides help in the selection of the appropriate level of model detail. Further, in conjunction with uncertainty propagation, it can be used to identify the relative magnitudes of uncertainties associated with model inputs and model formulation. This provides information as to where available resources should be focused: filling data gaps versus refining the model used for the problem.

These objectives are accomplished via the development of the Stochastic Response Surface Method (SRSM), and its evaluation for a range of transport-transformation models. Additionally, an easy to use, web-based interface is developed to facilitate black-box use of this method. Furthermore, the SRSM is coupled with an existing sensitivity analysis method, ADIFOR (Automatic DIfferentiation of FORtran), in order to further improve the computational efficiency of the SRSM. Finally, uncertainties

associated with model structure and formulation are addressed in the framework of photochemical air quality modeling.

1.5 Outline of the Thesis

Chapter 1 has presented a brief introduction to the concepts of uncertainty, and the importance of uncertainty analysis in the context of transport-transformation models. It has also discussed the steps involved in a systematic uncertainty analysis, and the associated limitations. Additionally, Chapter 1 summarizes the objectives and presents a brief overview of this thesis (Figure 1.1).

Chapter 2 presents the relevant background information in the following order:

- classification of uncertainty into *natural*, *model*, and *parametric/data* uncertainties, and into *reducible* and *irreducible* uncertainties,
- approaches for representing parametric/data uncertainty, such as, *Interval Mathematics*, *Fuzzy Theory*, and *Probabilistic Analysis* (discussed further in Appendix A), and
- conventional methods for sensitivity and sensitivity/uncertainty analysis, classified here as *sensitivity testing methods*, *analytical methods*, *sampling based methods*, and *computer algebra based methods*.

Additionally, Chapter 2 presents several applications of the conventional methods to the uncertainty analysis of transport-transformation models. The merits and limitations of each method are discussed, and an argument is made for alternative methods that can be used for combined sensitivity and uncertainty analysis. This argument is made both from the perspective of the computational cost reduction, and from the perspective of obtaining sensitivity information.

Chapter 3 presents the Stochastic Response Surface Method (SRSM), which is developed here as a computationally efficient alternative method for uncertainty analysis of numerical models. The formulation, implementation, and application of the SRSM are discussed in detail. Individual sections elaborate on the steps involved in the application

of the SRSM: (a) representation of random inputs in a "standardized manner", (b) expression of model outputs in a parametric form (e.g., series expansions), (c) estimation of the parameters of output approximation (e.g., coefficients in a series expansion), (d) calculation of the statistical properties of outputs, and (e) evaluation of the approximation. Chapter 3 also presents an example problem that illustrates the application of the SRSM, explaining all the steps in detail; case studies for the evaluation of the SRSM are presented in Chapter 5. Additionally, Chapter 3 presents a web-interface for the application of the SRSM through a browser, without requiring additional effort in studying the mathematical details of the method.

Chapter 4 presents the coupling of the SRSM with the Automated Differentiation method, ADIFOR (Automatic DIfferentiation of FORtran). ADIFOR is a computer algebra based method that reads the Fortran source code of a model and constructs the code that calculates first order partial derivatives of model outputs with respect to model inputs or parameters. In Chapter 4, first the rationale for coupling the SRSM with ADIFOR is presented, and is followed by a description of ADIFOR. Then, the coupling of the SRSM with ADIFOR is presented, and the application of the coupled method, SRSM-ADIFOR, is illustrated with an example problem.

Chapter 5 presents a comprehensive evaluation of both the SRSM and the SRSM-ADIFOR methods. This evaluation consists of four case studies, covering a wide range of transport-transformation models, as follows:

I: A zero-dimensional physiologically-based pharmacokinetic model.

II: A two-dimensional reactive atmospheric plume model.

III: A three-dimensional urban/regional scale photochemical air quality model.

IV: A one-dimensional ground water model with discontinuous probability density functions for inputs and parameters.

The SRSM is evaluated for all the four models, whereas the SRSM-ADIFOR is evaluated for the first two models. Each case study description contains the information about the

model and its application in exposure/risk assessment. Then, the uncertainties associated with the model are described, and the application of different uncertainty propagation methods is presented. The estimates from the uncertainty propagation methods are then presented graphically, in the form of probability density functions (*pdf*s), along with a brief discussion at the end.

Chapter 6 addresses the issues involved with *model uncertainty*, which arises due to uncertainties in model formulation and model structure. This chapter describes how such uncertainty can be addressed in the framework of photochemical air quality modeling. A study is presented involving a regulatory atmospheric photochemical plume model, the Reactive Plume Model (RPM). The development and implementation of the RPM-3D, a three-dimensional version of the RPM, is presented in this Chapter. Subsequently, case studies are presented to compare the estimates of RPM-3D and RPM, in order to characterize uncertainties associated with model resolution.

Chapter 7 presents the conclusions of this thesis, and recommendations for future work. This is followed by bibliography.

Appendix A presents background information on probabilistic analysis that is useful in understanding the SRSM and the SRSM-ADIFOR. Appendix B, Appendix C, and Appendix D present information on the models used in the case studies in Chapter 5. Appendix E lists the programs that are included with the accompanying CDROM.

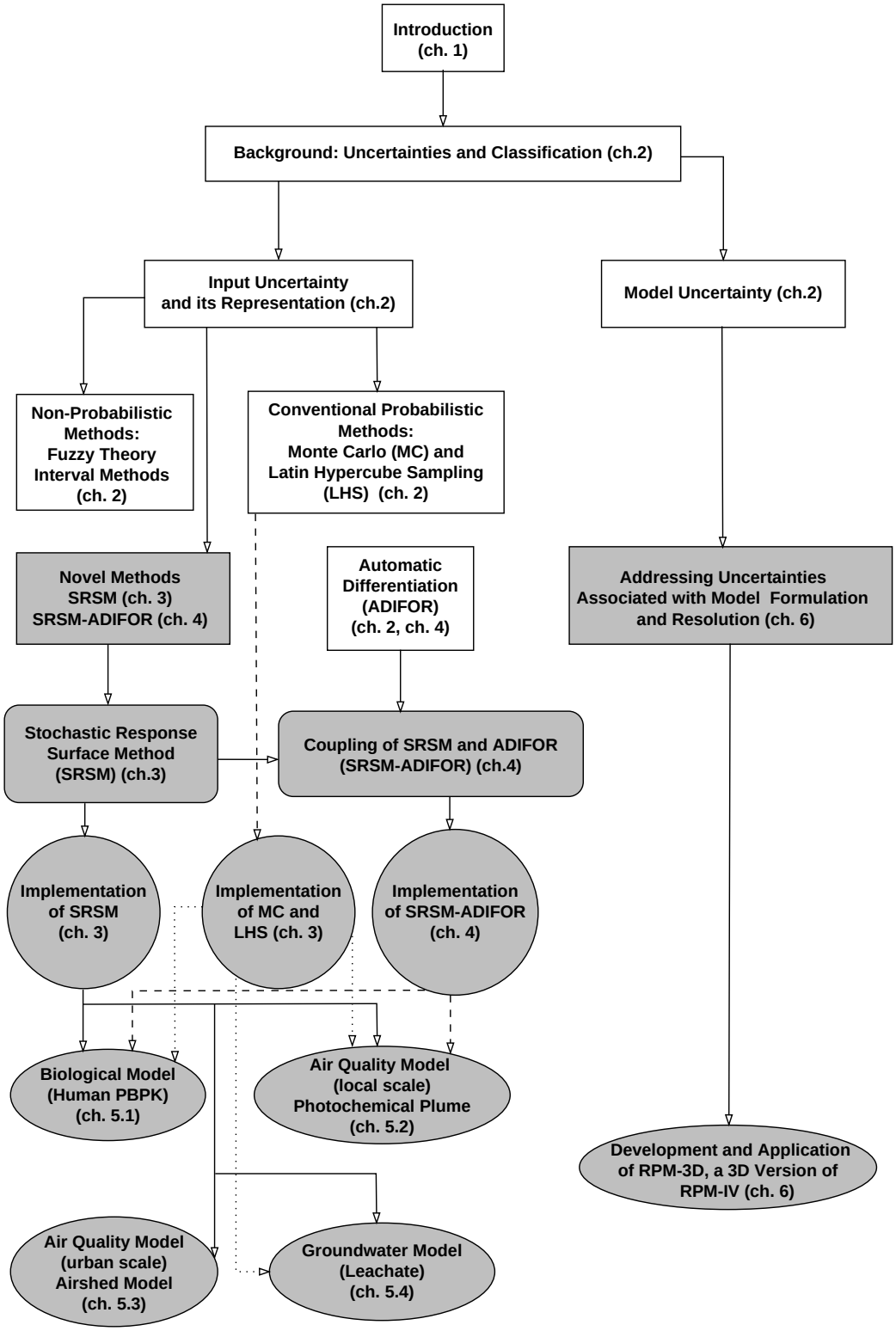


Figure 1.1: Schematic depiction of the outline of this thesis. The shaded areas indicate the contributions of this work

Chapter 2
BACKGROUND

2.1 Types and Origins of Uncertainty in Transport-Transformation Models

Uncertainties in transport-transformation models can be classified as *Natural*, *Model*, and *Data* uncertainties, depending on their origins and on how they can be addressed. They are briefly described in the following.

2.1.1 Natural Uncertainty and Variability

Environmental and biological systems are inherently stochastic due to unavoidable unpredictability (randomness). Some quantities are random even in principle, while some quantities that are precisely measurable are modeled as "random" quantities as a practical matter (due to cost and effort involved with continuous and precise measurement). For example, in air pollution systems, the turbulent atmosphere and unpredictable emission-related activities contribute to "natural uncertainty". In such cases, a precise estimation of system properties is not possible, and the uncertainty can be characterized through ensemble averages. On the other hand, the modeling of emissions from a source are sometimes modeled via a mean value and a "random error", since it is impractical to continuously monitor the emissions.

Further, some quantities vary over time, over space, or across individuals in a population; this is termed "variability". Variability is the heterogeneity between individual members of a population of some type, and is typically characterized through a frequency distribution. It is possible to interpret variability as uncertainty under certain conditions, since both can be addressed in terms of "frequency" distributions.

However, the implications of the differences in uncertainty and variability are relevant in decision making. For example, the knowledge of the frequency distribution for

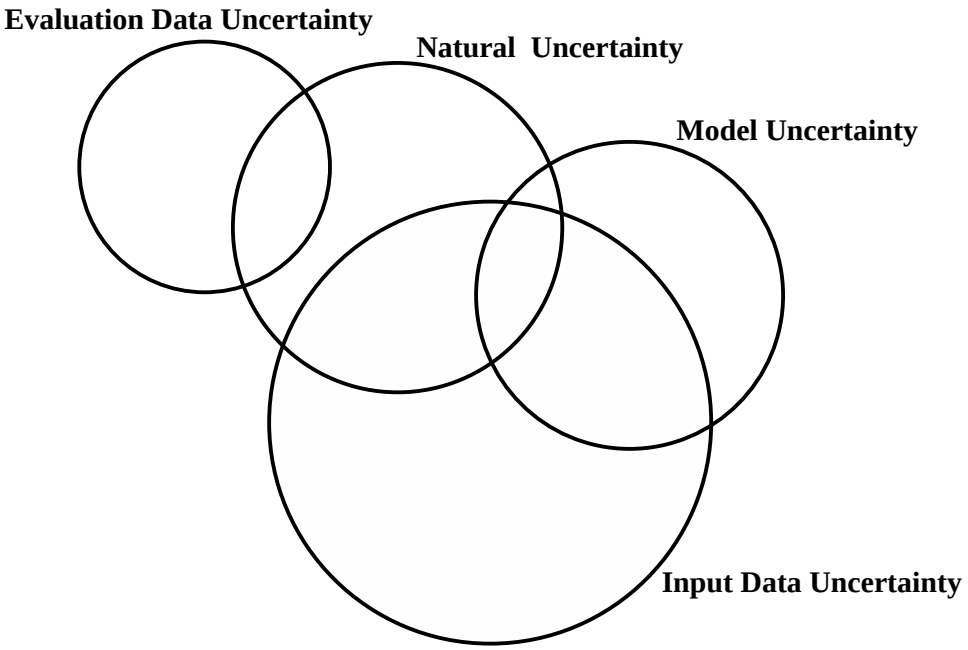


Figure 2.1: Types of uncertainty present in transport-transformation modeling applications and their interrelationships (adapted from [78])

variability can guide the identification of significant subpopulations which merit more focused study. In contrast, the knowledge of uncertainty can aid in determining areas where additional research or alternative measurement techniques are needed to reduce uncertainty.

2.1.2 Model Uncertainty

Mathematical models are necessarily simplified representations of the phenomena being studied and a key aspect of the modeling process is the judicious choice of model assumptions. The optimal mechanistic model will provide the greatest simplifications while providing an adequately accurate representation of the processes affecting the phenomena of interest. Hence, the structure of mathematical models employed to represent transport-transformation systems is often a key source of uncertainty. In addition to the significant approximations often inherent in modeling, sometimes competing models may be available. Furthermore, the limited spatial or temporal resolution (e.g., numerical grid cell size) of many models is also a type of approximation that introduces uncertainty into model results. In this context, different sources of model uncertainties

are summarized in the following:

Model Structure: Uncertainty arises when there are alternative sets of scientific or technical assumptions for developing a model. In such cases, if the results from competing models result in similar conclusions, then one can be confident that the decision is robust in the face of uncertainty. If, however, alternative model formulations lead to different conclusions, further model evaluation might be required.

Model Detail: Often, models are simplified for purposes of tractability. These include assumptions to convert a complex nonlinear model to a simpler linear model in a parameter space of interest. Uncertainty in the predictions of simplified models can sometimes be characterized by comparison of their predictions to those of more detailed, inclusive models.

Extrapolation: Models that are validated for one portion of input space may be completely inappropriate for making predictions in other regions of the parameter space. For example, a dose-response model based on high-dose, short duration animal tests may incur significant errors when applied to study low-dose, long duration human exposures. Similarly, in atmospheric modeling, models that are evaluated only for base case emission levels, may introduce uncertainties when they are employed in the study of future scenarios that involve significantly different emissions levels (also known as ΔE caveat).

Model Resolution: In the application of numerical models, selection of a spatial and/or temporal grid size often involves uncertainty. On one hand, there is a trade-off between the computation time (hence cost) and prediction accuracy. On the other hand, there is a trade-off between resolution and the validity of the governing equations of the model at such scales. Very often, a coarse grid resolution introduces approximations and uncertainties into model results. Sometimes, a finer grid resolution need not necessarily result in more accurate predictions. For example, in photochemical grid modeling, the atmospheric diffusion equation (ADE) (see Equation 6.1) is valid for grid cell sizes between 2 km and 20 km [137]. In this case, a coarse-grid model ignores significant "sub-grid" detail, and a fine-grid model would require more computer resources, and may even produce incorrect results, as the governing equation may not be appropriate.

Table 2.1: Examples of the sources of uncertainty in the formulation and application of transport-transformation models

Uncertainty in Model Formulation (Structural Uncertainty)	**Uncertainty in Model Application (Data/Parametric Uncertainty)**
Simplifications in Conceptual Formulation	Constitutive Parameter Selection
Simplifications in Mathematical Formulation	Design/Structural Parameter Selection
Ergodic-type Hypotheses	Input Data Development/Selection
Idealizations in Constitutive Relations	Source Information
Independence Hypotheses	Meteorology (air quality models)
Spatial Averaging	Hydrology (water quality models)
Temporal Averaging	Initial and Boundary Conditions
Process Decoupling	Operational Model Evaluation
Lumping of Parameters	Uncertainty in model estimates
Discretization	Uncertainty in observations
Numerical Algorithm/Operator Splitting	Nonexistence of observations
Approximations in Computer Coding	Response Interpretation
	ΔE caveat and other implicit hypotheses

This type of uncertainty is sometimes dealt with through an appropriate selection of model domain parameter values, or by comparing results based on different grid sizes.

Model Boundaries: Any model may have limited boundaries in terms of time, space, number of chemical species, types of pathways, and so on. The selection of a model boundary may be a type of simplification. Within the boundary, the model may be an accurate representation, but other overlooked phenomenon not included in the model may play a role in the scenario being modeled.

2.1.3 Parametric/Data Uncertainty

Uncertainties in model parameter estimates stem from a variety of sources. Even though many parameters could be measurable up to any desired precision, at least in principle,

there are often significant uncertainties associated with their estimates. Some uncertainties arise from measurement errors; these in turn can involve (a) random errors in analytic devices (e.g., the imprecision of continuous monitors that measure stack emissions), (b) systematic biases that occur due to imprecise calibration, or (c) inaccuracies in the assumptions used to infer the actual quantity of interest from the observed readings of a "surrogate" or "proxy" variable. Other potential sources of uncertainties in estimates of parameters include misclassification, estimation of parameters through a small sample, and estimation of parameters through non-representative samples. Further, uncertainty in model application arises from uncertainties associated with measurement data used for the model evaluation.

Table 2.1 lists some of the sources of model and parametric uncertainty associated with the formulation and the application of transport-transformation models.

2.2 Reducible and Irreducible Uncertainty

Uncertainty associated with model formulation and application can also be classified as "reducible" and "irreducible". Natural uncertainty is "inherent" or irreducible, whereas data and model uncertainty contain both reducible and irreducible components. The irreducible uncertainty in data and models is generally a result of the presence of natural uncertainty. Reducible uncertainty can be lowered, e.g., by better inventorying methods, improved instrumentation, improvements in model formulation, etc. Nevertheless, the distinction between reducible and irreducible model and data uncertainties is to a great extent a matter of convention since it may not be feasible to eliminate the presence of an error (reducible uncertainty) in measurement or modeling beyond a certain level. Furthermore, what is perceived as irreducible natural uncertainty may be quantified in a statistical sense, and via mechanistic modeling, better than "artificial" reducible uncertainty. Irreducible modeling uncertainty reflects the "current" model formulation and may actually change when improved theories describing the phenomena under consideration become available. Also, the averaging processes involved in model formulation unavoidably "lump" together natural and modeling uncertainty and only

a quantification of this lumped uncertainty may be possible or desirable. Figure 2.1 depicts schematically the types of uncertainties present in transport-transformation models, and their interrelationships.

2.3 Approaches for Representation of Uncertainty

Various approaches for representing uncertainty in the context of different domains of applicability are presented by Klir [125], and are briefly summarized in the following:

- *Classical set theory:* Uncertainty is expressed by sets of mutually exclusive alternatives in situations where one alternative is desired. This includes *diagnostic, predictive* and *retrodictive* uncertainties. Here, the uncertainty arises from the *nonspecificity* inherent in each set. Large sets result in less specific predictions, retrodictions, etc., than smaller sets. Full specificity is obtained only when one alternative is possible.

- *Probability theory:* Uncertainty is expressed in terms of a measure on subsets of a universal set of alternatives (events). The uncertainty measure is a function that, according to the situation, assigns a number between 0 and 1 to each subset of the universal set. This number, called *probability* of the subset, expresses the *likelihood* that the desired unique alternative is in this subset. Here, the uncertainty arises from the *conflict* among likelihood claims associated with the subsets of the universal set, each consisting of exactly one alternative. Since these alternatives are mutually exclusive, nonzero probabilities assigned to two or more events conflict with one another since only one of them can be the desired one. A detailed description of the probability theory is presented in Appendix A.

- *Fuzzy set theory:* Fuzzy sets, similar to classical sets, are capable of expressing nonspecificity. In addition, they are also capable of expressing *vagueness.* Vagueness is different from nonspecificity in the sense that vagueness emerges from imprecision of definitions, in particular definitions of linguistic terms. In fuzzy sets, the membership is not a matter of affirmation or denial, but rather a matter of degree.

- *Fuzzy measure theory:* This theory considers a number of special classes of measures, each of which is characterized by a special property. Some of the measures used in this theory are *plausibility* and *belief* measures, and the classical probability measures. Fuzzy measure theory and fuzzy set theory differ significantly: in the fuzzy set theory, the conditions for the membership of an element into a set are vague, whereas in the fuzzy measure theory, the conditions are precise, but the information about an element is insufficient to determine whether it satisfies those conditions.

- *Rough set theory:* A rough set is an imprecise representation of a *crisp* set in terms of two subsets, a *lower approximation* and *upper approximation.* Further, the approximations could themselves be imprecise or fuzzy.

Some of the widely used uncertainty representation approaches used in transport-transformation modeling include interval mathematics, fuzzy theory, and probabilistic analysis. These approaches are presented in the following sections.

2.3.1 Interval Mathematics

Interval mathematics is used to address data uncertainty that arises (a) due to imprecise measurements, and (b) due to the existence of several alternative methods, techniques, or theories to estimate model parameters. In many cases, it may not be possible to obtain the probabilities of different values of imprecision in data; in some cases only error bounds can be obtained. This is especially true in case of conflicting theories for the estimation of model parameters, in the sense that "probabilities" cannot be assigned to the validity of one theory over another. In such cases, interval mathematics can be used for uncertainty estimation, as this method does not require information about the type of uncertainty in the parameters [5, 24].

The objective of interval analysis is to estimate the bounds on various model outputs based on the bounds of the model inputs and parameters. In the interval mathematics approach, uncertain parameters are assumed to be "unknown but bounded", and each of them has upper and lower limits without a probability structure [184]; every uncertain parameter is described by an interval. If a parameter x_i of a model is known to be

between $\overline{x_i} - \epsilon_i$ and $\overline{x_i} + \epsilon_i$, the interval representation of x_i is given by $[\overline{x_i} - \epsilon_i, \overline{x_i} + \epsilon_i]$. Correspondingly, the model estimates would also belong to another interval. Special arithmetic procedures for calculation of functions of intervals used in this method are described in the literature [5, 151, 152, 155]. For example, if two variables, a and b are given by $[a_l, a_u]$ and $[b_l, b_u]$, where $a_l \leq a_u$ and $b_l \leq b_u$, simple arithmetic operations are given by the following:

$$\begin{aligned} a + b &= [a_l + b_l, a_u + b_u] \\ a - b &= [a_l - b_u, a_u - b_l] \\ a \cdot b &= [\min(a_l b_l, a_l b_u, a_u b_l, a_u b_u), \max(a_l b_l, a_l b_u, a_u b_l, a_u b_u)] \\ a/b &= [a_l, a_u] \cdot \left[\frac{1}{b_u}, \frac{1}{b_l}\right]; 0 \notin [b_l, b_u] \end{aligned} \tag{2.1}$$

Furthermore, a range of computational tools exist for performing interval arithmetic on arbitrary functions of variables that are specified by intervals [122, 132]. Symbolic computation packages such as Mathematica [10] and Maple [94] support interval arithmetic. In addition, extensions to FORTRAN language with libraries for interval arithmetic are also available [120, 121]. Therefore, in principle, the ranges of uncertainty in any analytical or numerical model (FORTRAN model) can be analyzed by existing tools.

Various applications of interval analysis in the literature include the treatment of uncertainty in the optimal design of chemical plants [69], in the cost benefit analysis of power distribution [184], and in decision evaluation [24]. Interval analysis has also been applied to estimate the uncertainty in displacements in structures due to the uncertainties in external loads [129]. Dong et al. presented a methodology for the propagation of uncertainties using intervals [52]. Hurme et al. presented a review on semi qualitative reasoning based on interval mathematics in relation to chemical and safety engineering [104]. Applications in transport-transformation modeling include uncertainty analysis of groundwater flow models [53, 143].

The primary advantage of interval mathematics is that it can address problems of uncertainty analysis that cannot be studied through probabilistic analysis. It may be useful for cases in which the probability distributions of the inputs are not known.

However, this method does not provide adequate information about the nature of output uncertainty, as all the uncertainties are forced into one arithmetic interval [134]. Especially when the probability structure of inputs is known, the application of interval analysis would in fact ignore the available information, and hence is not recommended.

2.3.2 Fuzzy Theory

Fuzzy theory is a method that facilitates uncertainty analysis of systems where uncertainty arises due to vagueness or "fuzziness" rather than due to randomness alone [65]. This is based on a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth – truth values between "completely true" and "completely false". It was introduced by Zadeh as a means to model the uncertainty of natural language [222]. Fuzzy theory uses the process of "fuzzification" as a methodology to generalize any specific theory from a crisp (discrete) to a continuous (fuzzy) form.

Classical set theory has a "crisp" definition as to whether an element is a member of a set or not. However, certain attributes of systems cannot be ascribed to one set or another. For example, an attribute of a system can be specified as either "low" or "high". In such a case, uncertainty arises out of vagueness involved in the definition of that attribute. Classical set theory allows for either one or the other value. On the other hand, fuzzy theory provides allows for a gradual degree of membership. This can be illustrated as follows:

In the classical set theory, the truth value of a statement can be given by the membership function $\mu_A(x)$, as

$$\mu_A(x) = \begin{cases} 1 & \text{iff} \quad x \in A \\ 0 & \text{iff} \quad x \notin A \end{cases} \tag{2.2}$$

On the other hand, fuzzy theory allows for a continuous value of μ_A between 0 and 1, as

$$\mu_A(x) = \begin{cases} 1 & \text{iff} \quad x \in A \\ 0 & \text{iff} \quad x \notin A \\ p; \quad 0 < p < 1 & \text{if} \quad x \text{ partially belongs to} A \end{cases} \tag{2.3}$$

In fuzzy theory, statements are described in terms of membership functions, that are continuous and have a range [0,1]. For example, given the measured value of a parameter, the membership function gives the "degree of truth" that the parameter is "high" or "low".

Further, fuzzy logic is defined by the following the set relationships:

$$\begin{aligned} \mu_{A'}(x) &= 1.0 - \mu_A(x) \\ \mu_{A\cap B}(x) &= \min(\mu_A(x), \mu_B(x)) \\ \mu_{A\cup B}(x) &= \max(\mu_A(x), \mu_B(x)) \end{aligned} \tag{2.4}$$

Using fuzzy arithmetic, based on the grade of membership of a parameter of interest in a set, the grade of membership of a model output in another set can be calculated. Fuzzy theory can be considered to be a generalization of the classical set theory. It must be noted that if the membership grades are restricted to only 0 and 1, the fuzzy theory simplifies to classical set theory.

A vast amount of literature is available on fuzzy theory, including extensive introductions to the concepts involved by Bezdek and by Smithson [15, 191]. Fuzzy arithmetic and its applications are described by Klir et al., by Kauffmann et al., and by Puri et al. [119, 126, 166]. Kraslawski et al. applied fuzzy set theory to study uncertainty associated with incomplete and subjective information in process engineering [130]. Ayyub et al. studied structural reliability assessment by using fuzzy theory to study the uncertainty associated with ambiguity [9]. Juang et al. demonstrated the applicability of fuzzy theory in the modeling and analysis of non-random uncertainties in geotechnical engineering [115]. Further literature on industrial applications of fuzzy theory is available for the interested reader [174, 221]. Specifically, it has been applied to characterize uncertainty in engineering design calculations [131,219], in quality control [167], in sludge application land selection [45], and in solute transport modeling [53,54,143]. However, drawbacks in its applicability to uncertainty analysis has been noted by some researchers [193,212]. Fuzzy theory appears to be more suitable for qualitative reasoning, and classification of elements into a fuzzy set, than for quantitative estimation of uncertainty.

2.3.3 Probabilistic Analysis

In the probabilistic approach, uncertainties are characterized by the *probabilities* associated with *events*. The *probability* of an event can be interpreted in terms of the frequency of occurrence of that event. When a large number of *samples* or *experiments* are considered, the probability of an event is defined as the ratio of the number of times the event occurs to the total number of samples or experiments. For example, the statement that the probability that a pollutant concentration c lies between c_1 and c_2 equals p means that from a large number of independent measurements of the concentration c, under identical conditions, the number of times the value of c lies between c_1 and c_2 is roughly equal to the fraction p of the total number of samples.

Probabilistic analysis is the most widely used method for characterizing uncertainty in physical systems, especially when estimates of the probability distributions of uncertain parameters are available. This approach can describe uncertainty arising from stochastic disturbances, variability conditions, and risk considerations. In this approach, the uncertainties associated with model inputs are described by probability distributions, and the the objective is to estimate the output probability distributions. This process comprises of two stages:

- *Probability encoding of inputs:* This process involves the determination of the probabilistic distribution of the input parameters, and incorporation of random variations due to both natural variability (from, e.g., emissions, meteorology) and "errors." This is accomplished by using either statistical estimation techniques or expert judgments. Statistical estimation techniques involve estimating probability distributions from available data or by collection of a large number or representative samples. Techniques for estimating probability distributions from data are presented in the literature [37,103]. In cases where limited data are available, an expert judgment provides the information about the input probability distribution. For example, a uniform distribution is selected if only a range of possible values for an input is available, but no information about which values are more likely to occur is available. Similarly, normal distribution is typically

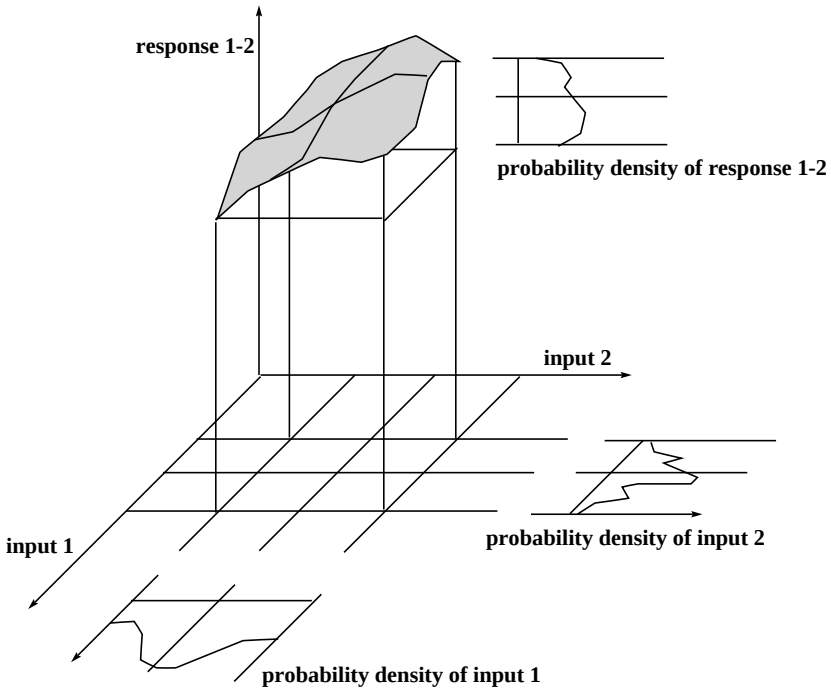


Figure 2.2: A schematic depiction of the propagation of uncertainties in transport-transformation models

used to describe unbiased measurement errors. Table 2.2 presents a list of some commonly used probability distributions in the uncertainty analysis of transport-transformation models.

- *Propagation of uncertainty through models:* Figure 2.2 depicts schematically the concept of uncertainty propagation: each point of the response surface (i.e., each calculated output value) of the model to changes in inputs "1" and "2" will be characterized by a probability density function (*pdf*) that will depend on the *pdf*s of the inputs. The techniques for uncertainty propagation are discussed in the following.

Appendix A presents some background information on probabilistic analysis and random variables. A number of excellent texts on probabilistic analysis are available in the literature for the interested reader [75, 160, 201].

Table 2.2: Selected probability density functions useful in representing uncertainties in environmental and biological model inputs

Distribution	Parameters & Conditions	Probability density function	Moments
Uniform	a, b	$\frac{1}{b-a}$	Mean $= \frac{a+b}{2}$ Var $= \frac{(b-a)^2}{12}$
Normal	$\mu, \sigma\ , \sigma > 0$	$\frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	Mean $= \mu$ Var $= \sigma^2$ Mode $= \mu$
Lognormal	$\mu, \sigma\ , \sigma > 0$	$\frac{1}{x\sigma\sqrt{2\pi}}\ e^{-\frac{(\log(x)-\mu)^2}{2\sigma^2}}$	Mean $= e^{(\mu + \sigma^2/2)}$ Var $= (e^{\sigma^2} - 1)e^{(2\mu + \sigma^2)}$ Mode $= e^{(\mu - \sigma^2)}$
Gamma	$a, b,$ $a > 0, b > 0$	$\frac{1}{\Gamma(a)b^a}\ x^{a-1}e^{-\frac{x}{b}}\ \ x > 0$	Mean $= ab$ Var $= ab^2$ Mode $= (a-1)b$
Exponential	$\lambda, \lambda > 0$	$\lambda e^{-\lambda x}\ ,\ x > 0$	Mean $= \frac{1}{\lambda}$ Var $= \frac{1}{\lambda^2}$ Mode $= 0$
Weibull	a	$ax^{a-1}e^{-x^a}\ \ (x \geq 0)$	Mean $= \Gamma(1 + \frac{1}{a})$ Var $= \Gamma(1 + \frac{2}{a}) - \Gamma(1 + \frac{1}{a})$ Mode $= (1 - \frac{1}{a})^{\frac{1}{a}}, a \geq 1$
Extreme Value		$e^{-x} - e^{-x}$	Mean $= 0$ Var $= 1$ Mode $= 0$

2.4 Sensitivity and Sensitivity/Uncertainty Analysis

The aim of sensitivity analysis is to estimate the rate of change in the output of a model with respect to changes in model inputs. Such knowledge is important for (a) evaluating the applicability of the model, (b) determining parameters for which it is important to have more accurate values, and (c) understanding the behavior of the system being modeled. The choice of a sensitivity analysis method depends to a great extent on (a) the sensitivity measure employed, (b) the desired accuracy in the estimates of the sensitivity measure, and (c) the computational cost involved.

In general, the meaning of the term "sensitivity analysis" depends greatly on the sensitivity measure that is used. Table 2.3 presents some of the sensitivity measures that are often employed in the sensitivity analysis of a mathematical model of the form

$$\mathcal{F}(\boldsymbol{u}, \boldsymbol{k}) = 0 \tag{2.5}$$

where $\boldsymbol{k}$ is a set of m parameters, and $\boldsymbol{u}$ is a vector of n output variables.

Based on the choice of a sensitivity metric and the variation in the model parameters, sensitivity analysis methods can be broadly classified into the following categories:

- *Variation of parameters or model formulation:* In this approach, the model is run at a set of sample points (different combinations of parameters of concern) or with straightforward changes in model structure (e.g., in model resolution). Sensitivity measures that are appropriate for this type of analysis include the response from arbitrary parameter variation, normalized response and extrema. Of these measures, the extreme values are often of critical importance in environmental applications.

- *Domain-wide sensitivity analysis:* Here, the sensitivity involves the study of the system behavior over the entire range of parameter variation, often taking the uncertainty in the parameter estimates into account.

- *Local sensitivity analysis:* Here, the focus is on estimates of model sensitivity to input and parameter variation in the vicinity of a sample point. This sensitivity is often characterized through gradients or partial derivatives at the sample point.

Table 2.3: A summary of sensitivity measures employed in sensitivity analysis, adapted from McRae et al., 1982 [148]

Sensitivity Measure	Definition
Response from arbitrary parameter variation	$\boldsymbol{u} = \boldsymbol{u}(\overline{\boldsymbol{k}} + \delta\boldsymbol{k}) - \boldsymbol{u}(\boldsymbol{k})$
Normalized Response	$D_i = \dfrac{\delta u_i}{u_i(\overline{\boldsymbol{k}})}$
Average Response	$\overline{u_i(\overline{\boldsymbol{k}})} = \dfrac{\int_{\boldsymbol{k}}\cdots\int u_i(\overline{\boldsymbol{k}})d\boldsymbol{k}}{\int_{\boldsymbol{k}}\cdots\int d\boldsymbol{k}}$
Expected Value	$\langle u_i(\boldsymbol{k})\rangle = \int_{\boldsymbol{k}}\cdots\int u_i(\boldsymbol{k}P(\boldsymbol{k})d\boldsymbol{k}$
Variance	$\delta_i^2(\boldsymbol{k}) = \langle u_i(\boldsymbol{k})^2\rangle - \langle u_i(\boldsymbol{k})\rangle^2$
Extrema	$\max\,[u_i(\boldsymbol{k})]$, $\min\,[u_i(\boldsymbol{k})]$
Local Gradient Approximation	$\delta\boldsymbol{u} \approx [\boldsymbol{S}]\delta\boldsymbol{k}$; $S_{ij} = \dfrac{\partial u_i}{\partial k_j}$
Normalized Gradient	$S_{ij}^n = \dfrac{\overline{k}_j}{u_i(\overline{\boldsymbol{k}})}\dfrac{\partial u_i}{\partial k_j}$

2.5 Conventional Sensitivity/Uncertainty Analysis Methods

Conventional methods for sensitivity analysis and uncertainty propagation can be broadly classified into four categories: (a) "sensitivity testing", (b) analytical methods, (c) sampling based methods, and (d) computer algebra based methods.

Sensitivity testing involves studying model response for a set of changes in model formulation, and for a selected model parameter combinations. Analytical methods involve either the differentiation of model equations and subsequent solution of a set

of auxiliary sensitivity equations, or the reformulation of original model using stochastic algebraic/differential equations. On the other hand, the sampling based methods involve running the original model for a set of input/parameter combinations (sample points) and estimating the sensitivity/uncertainty using the model outputs at those points. Yet another sensitivity analysis method is based on direct manipulation of the computer code of the model, and is termed automatic differentiation. These methods are elaborated in the following.

2.5.1 Sensitivity Testing Methods

In this approach, the model is run for a set of sample points for the parameters of concern or with straightforward changes in model structure (e.g., in model resolution). This approach is often used to evaluate the robustness of the model, by testing whether the model response changes significantly in relation to changes in model parameters and structural formulation of the model. The application of this approach is straightforward, and it has been widely employed in conjunction with transport-transformation modeling.

Sistla et al. [189] and Al-Wali et al. [4] used this approach for the sensitivity analysis of urban air quality models. Roselle [173] used this approach to study the effect of biogenic emission uncertainties by performing model simulations at three emission biogenic estimate levels. Vieux et al. [207] and Vanderperk [206] have used a similar approach in water quality modeling.

The primary advantage of this approach is that it accommodates both qualitative and quantitative information regarding variation in the model. However, the main disadvantage of this approach is that detailed information about the uncertainties is difficult to obtain using this approach. Further, the sensitivity information obtained depends to a great extent on the choice of the sample points, especially when only a small number of simulations can be performed.

2.5.2 Analytical Methods

Some of the widely used analytical methods for sensitivity/uncertainty are: (a) differential analysis methods, (b) Green's function method, (c) spectral based stochastic finite element method, and (d) coupled and decoupled direct methods.

Differential Analysis Methods

Differential analysis methods include the Neumann expansion [3, 197], and the perturbation method [3, 198]. The Neumann expansion method involves finding the inverse of the model operator through the expansion of the model equations, and hence has limitations on the type of model equations it can address. The perturbation method involves expansion of model outputs as a series in terms of small random perturbations in model parameters, and the subsequent solution of the series coefficients. The Neumann expansion and perturbation based methods have been applied in uncertainty analysis of ground water models [1, 2, 224], and in the design of structures [26, 118, 140]. The main limitation of these methods is the requirement that the perturbation terms be small. Further, these methods are in general difficult to apply in conjunction with the modeling of complex, nonlinear systems, as the model equations are often mathematically intractable.

Green's Function Method

In the Green's function method, the sensitivity equations of a model are obtained by differentiating the model equations. The sensitivity equations are then solved by constructing an auxiliary set of Green's functions. This method minimizes the number of differential equations that are solved for sensitivity, and replaces them with integrals that can be easily evaluated [55, 56]. This approach has been applied to the sensitivity analysis of atmospheric chemistry models [106, 209].

Spectral Based Stochastic Finite Element Method

This method relies on the use of representing stochastic processes in terms of a series expansion, specifically the Karhunen-Loeve expansion [83, 160]. For finite element

method problems, this approach results in a set of linear matrix equations with deterministic matrices multiplied by random vectors. The matrix equations are solved either using operator expansions or by using the Galerkin's method [208]. One of the main features of this method is the representation of random parameters in terms of orthogonal functions of a set of standardized random variables; the expansion is also known as "polynomial chaos expansion", and forms the basis for the development of the "Stochastic Response Surface Method" (SRSM). The polynomial chaos expansion and the SRSM are discussed in detail in Chapter 3.

Coupled/Decoupled Direct Method

The direct method involves the differentiation of model equations and the subsequent solution of the sensitivity equations. The sensitivity equations are then solved along with the original model equations (Coupled Direct Method) [200], or separately (Decoupled Direct Method) [60, 165]. The decoupled method is advantageous both in terms of computational efficiency and stability of the solution. This method has evolved into a standard module in conjunction with commercial sensitivity analysis codes for chemical kinetics, such as CHEMKIN [41], and GCKP86 [168]. The decoupled method is also reported to be more efficient than the Green's function method [60].

The analytical methods require access to the governing model equations, and may involve writing additional computer code for the solution of the auxiliary equations, which may be impractical and sometimes impossible. For example, reformulating an existing computational model developed by others could require prohibitive amounts of resources.

2.5.3 Sampling Based Methods

Sampling based methods do not require access to model equations or even the model code. These methods involve running a set of model at a set of sample points, and establishing a relationship between inputs and outputs using the model results at the sample points. Some of the widely used sampling based sensitivity/uncertainty analysis methods are: (a) Monte Carlo and Latin Hypercube Sampling methods, (b) Fourier

Amplitude Sensitivity Test (FAST) (c) reliability based methods, and (d) response surface methods.

Monte Carlo and Latin Hypercube Sampling Methods

Monte Carlo (MC) methods are the most widely used means for uncertainty analysis, with applications ranging from aerospace engineering [12] to zoology [28]. These methods involve random sampling from the distribution of inputs and successive model runs until a statistically significant distribution of outputs is obtained. They can be used to solve problems with physical probabilistic structures, such as uncertainty propagation in models or solution of stochastic equations, or can be used to solve non-probabilistic problems, such as finding the area under a curve [49,75]. Monte Carlo methods are also used in the solution of problems that can be modeled by the sequence of a set of random steps that eventually converge to a desired solution. Problems such as optimization and the simulation of movement of fluid molecules are often addressed through Monte Carlo simulations [75]. For the interested reader, a wide range of literature describing the methodology, tools, and the applicability of the Monte Carlo methods is available in the literature [51, 71, 116, 177, 192].

Since these methods require a large number of samples (or model runs), their applicability is sometimes limited to simple models. In case of computationally intensive models, the time and resources required by these methods could be prohibitively expensive.*.

A degree of computational efficiency is accomplished by the use of Modified Monte Carlo (MMC) methods that sample from the input distribution in an efficient manner, so that the number of necessary solutions compared to the simple Monte Carlo method is significantly reduced.

The Latin Hypercube Sampling [107, 146, 194] is one such widely used variant of the standard Monte Carlo method. In this method, the range of probable values for

*For example, in order to perform the uncertainty analysis of a grid-based photochemical model, such as the Urban Airshed Model (UAM-IV), which requires about six hours of CPU time for the simulation of air quality near the vicinity of New Jersey (see the example in Section 5.3), a thousand samples would require more than eight months of computer time

each uncertain input parameter is divided into ordered segments of equal probability. Thus, the the whole parameter space, consisting of all the uncertain parameters, is partitioned into cells having equal probability, and they are sampled in an "efficient" manner such that each parameter is sampled once from each of its possible segments. The advantage of this approach is that the random samples are generated from all the ranges of possible values, thus giving insight into the extremes of the probability distributions of the outputs. A detailed description of the implementation of Latin Hypercube Sampling is presented in Section 3.8.1.

Monte Carlo methods and Latin Hypercube Sampling have been applied in probabilistic assessment of risk in physiologically based pharmacokinetic models by considering uncertainty in the estimates of the model parameters [44, 67, 187]. These methods have also been applied in groundwater contamination models [14, 89, 143, 157, 172], in air pollution modeling [34, 48, 74, 84, 179, 190], in process engineering [169, 215], and in the reliability assessment of structures [8].

Guidelines for the application of Monte Carlo and Latin Hypercube Sampling methods in risk assessment are provided by the U.S.EPA [203, 204]. Several computer packages containing routines for Monte Carlo and Latin Hypercube Sampling methods are reported in the literature [108, 112]. The EPA approved FORTRAN package for Latin Hypercube Sampling by Iman et al. [108] is available through the EPA's Exposure Models Library [202]. Commercial plugins based on spreadsheet software include Crystal ball [46], @RISK [218], RISKMAN [162], and MonteCarlo [164]. This list is only representative of the available software tools. In the present work, a Latin Hypercube Sampling utility in FORTRAN has been developed as a part of the implementation of the Stochastic Response Surface Method, and is freely available to the research community.

Fourier Amplitude Sensitivity Test (FAST)

Fourier Amplitude Sensitivity Test (FAST) is a method based on Fourier transformation of uncertain model parameters into a frequency domain, thus reducing the a multi-dimensional model into a single dimensional one. For a model with m model parameters,

$k_1, k_2, \ldots, k_m$, and n outputs, $u_1, u_2, \ldots, u_m$, such that

$$u_i = f_i(t; k_1, k_2, \ldots, k_m); \quad i = 1, 2, \ldots, n,$$

the FAST method involves the transformation of the parameters into a frequency domain spanned by a scalar s, as follows:

$$k_l = G_l(\sin w_l s), \quad l = 1, 2, \ldots, m$$

The outputs are then approximated as:

$$\begin{aligned} \overline{u}_i(t) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} u_i(t; k_1(s), k_2(s), \ldots, k_m(s)) ds \\ \sigma_i^2(t) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} u_i^2(t; k_1(s), k_2(s), \ldots, k_m(s)) ds - \overline{u}_i^2 \end{aligned} \tag{2.6}$$

These integrals are evaluated by repeatedly sampling the parameter space of s, which corresponds to the sampling in the multidimensional model parameter space. The details of the transformation of model parameters into the frequency domain, and the subsequent sampling are explained by Koda et al. [127], and a computational implementation of FAST is presented by McRae et al. [149].

FAST has been applied in the sensitivity/uncertainty analysis of atmospheric photochemical mechanisms [66], in conjunction with land surface processes in global climate models models [40, 145], and in the disposal of radioactive waste [124]. A review article by Helton [96] reports that Monte Carlo methods are more widely applicable than FAST.

Reliability Based Methods (FORM and SORM)

First- and second-order reliability methods (FORM and SORM, respectively) are approximation methods that estimate the probability of an event under consideration (typically termed "failure"). For example, these methods can provide the probability that a contaminant concentration exceeds a target level at a location (or, the probability of failure). In addition, these methods provide the contribution to the probability of failure from each input random variable, at no additional computational effort [92,117,147].

These methods are useful in uncertainty analysis of models with a single failure criterion.

For a model with random parameters

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_n),$$

and a failure condition

$$g(X_1, X_2, \ldots, X_n) < 0$$

the objective of the reliability based approach is to estimate the probability of failure. In case of pollutant contamination exceedance, the failure condition can be defined as

$$g(\boldsymbol{X}) \;=\; C_R - C(\boldsymbol{X}) \;<\; 0,$$

where C_R is a pre-specified maximum permissible concentration at a location of interest.

If the joint probability density function for the set $\boldsymbol{X}$ is given by $f_{\mathbf{X}}$, then the probability of failure is given by the n-fold integral:

$$P_F = P\{g(\boldsymbol{X}) < 0\} = P\{C_R < C(\boldsymbol{X})\} = \int_{g(\mathbf{X}) < 0} f_{\mathbf{X}} d\boldsymbol{X}$$

where the integration is carried out over the failure domain. The evaluation of this integral becomes computationally demanding as the number of random variables (the dimension of the integration) increases; in fact if m is the number of function calls of the integrand per dimension, and n is the dimension, the computation time grows as m^n [97]. In addition, since the value of the integrand is small, the numerical inaccuracies can be considerably magnified when integrated over a multi-dimensional space [23].

FORM and SORM use analytical schemes to approximate the probability integral, through a series of the following simple steps, as illustrated by Bjerager [20]:

- mapping the basic random variables $\boldsymbol{X}$ and the failure function $g(\boldsymbol{X})$, into a vector of standardized and uncorrelated normal variates $\boldsymbol{U}$, as $X(\boldsymbol{U})$ and $G(\boldsymbol{U})$ respectively,

- approximating the function $G(\boldsymbol{U})$ by a tangent (FORM) or a paraboloid (SORM) at a failure point $\boldsymbol{u}^*$ closest to the origin, and

- calculating the probability of failure as a simple function of $\boldsymbol{u}^*$.

These methods are reported to be computationally very efficient compared to Monte Carlo methods, especially for scenarios corresponding to low probabilities of failure [92]. Further, SORM is more accurate than FORM, but computationally more intensive, since it involves a higher order approximation. These methods have been applied to problems of ground-water flow and contaminant transport [90–92, 111] and to problems in safety assessment [39, 147].

The main drawbacks of FORM and SORM are that the mapping of the failure function on to a standardized set, and the subsequent minimization of the function, involve significant computational effort for nonlinear black box numerical models [82]. In addition, simultaneous evaluation of probabilities corresponding to multiple failure criteria would involve significant additional effort. Furthermore, these methods impose some conditions on the joint distributions of the random parameters [92], thus limiting their applicability.

Response Surface Methods

The response surface methods consist of (i) screening to determine a subset of important model input parameters, (ii) making multiple runs of the computer model using specific values and pairings of these input parameters, and (iii) fitting a general polynomial model to the model data (using the method of least squares). This fitted response-surface is then used as a replacement or proxy for the computer model, and all inferences related to sensitivity/uncertainty analysis for the original model are derived from this fitted model. This approach is sometimes termed as a "secondary model technique" [68]. Box et al. [21, 22] describe the adaptive nature of these methods, and the methodology for identifying the sampling points and for subsequent analysis is presented in detail in the literature [59, 70, 123].

Helton [96] studied the applicability of response surface methods in performance assessment of radioactive waste disposal. It has also been used in conjunction with soil erosion modeling [33], with vegetative plant growth modeling [141], and with structural reliability problems [27].

2.5.4 Computer Algebra Based Methods

Table 2.4: A representative list of packages available for automatic differentiation

Package	Authors	Year	Language	Applications/Features
ADIC	Bischof et al. [19]	1997	C	3-D CFD grid generator [19]
ADIFOR	Bischof et al. [17]	1996	Fortran	A wide range of applications [29, 101, 105, 110]
ADGEN	Worley et al. [220]	1989	Fortran	Radionuclide modeling [98]
ADOL-C	Griewank [87]	1996	C/C++	Groundwater flow [42]
ADOL-F	Shiriaev et al. [186]	1996	Fortran 90	
ATOMFT	Change et al. [32]	1994	Fortran	Solution of ODEs [32]
GRESS	Horwedel [100]		Fortran	Radionuclide transport [99, 156, 161]
Odyssee	Rostaing et al. [175]	1993	Fortran	Process engineering [150]
PADRE2	Kubota [133]	1991	Fortran	Estimates rounding errors
PCOMP	Dobmann et al. [50]	1995	Fortran	
SMS	Korelc [128]	1997	Mathematica	Automatic derivation of finite element formulae [128]
TEXPANDER	Ozaki et al. [158]	1993	Fortran	Optimization of structures [158]
Unnamed	Rich et al. [171]	1992	Matlab	

Computer algebra based methods involve the direct manipulation of the computer code, typically available in the form of a high level language code (such as C or FORTRAN), and estimation of the sensitivity and uncertainty of model outputs with respect to model inputs. These methods do not require information about the model structure or the model equations, and use mechanical, pattern-matching algorithms, to generate a "derivative code" based on the model code. One of the main computer algebra based methods is the automatic differentiation, which is sometimes also termed automated differentiation.

Automatic differentiation involves direct manipulation of a model code to generate

a corresponding "derivative calculating code". Given the source code, and the information about what the dependent and independent variables of interest are, the automatic differentiation tools can generate derivative information without further user intervention. This approach is essentially a "black-box" approach, as it does not require access to the model formulation or model equations. The main advantages of this approach are that partial derivatives can be calculated with the accuracy of the machine precision [17], and this approach results in substantial computer time savings [105, 113].

An overview of the methodology and applications of automatic differentiation is presented by Griewank [85, 86]. In addition, various reviews of automatic differentiation in the fields of engineering design, process engineering and numerical analysis can be found in the literature [13, 35, 109]. A large number of packages for automatic differentiation of model codes written in various programming languages are currently available. Table 2.4 lists a majority of the automatic differentiation tools currently available, along with some of their applications reported in the literature.

Two of the notable automatic differentiation tools that have been applied in conjunction with transport-transformation models are GRESS and ADIFOR. GRESS (GRadient Enhanced Software System) [100] is a Fortran preprocessor system that searches the model equations based on the appearance of the "=" symbol, and analyzes the functional dependence of independent and dependent variables. The differentiation is carried out for each statement and the results are stored numerically at each stage of computation. This method has been applied to models describing transport of radionuclides through groundwater [156, 161]. It has also been applied a model describing the atmospheric dispersion of radionuclides, AIRDOS [98, 152]. The main limitation of GRESS is that the particulars of every operation performed in the code are stored in the memory at each step. This interpretation overhead associated with the storage, and the potentially great memory demands, can sometimes limit its applicability [16].

ADIFOR (Automatic DIfferentiation of FORtran) [17] applies the rules of automatic differentiation, and rewrites the original code, inserting statements for computing the first-order derivatives. The generated code can then be compiled just like a normal program, and can be executed to calculate derivatives in the model, along with the

model outputs. This approach avoids the limitation of GRESS, with respect to the interpretation overhead.

In the area of photochemical air quality modeling, Carmichael et al. [29] applied the ADIFOR method to calculate the sensitivity of ozone levels to the initial concentrations (84 species) and all reaction rate constants (178 chemical reactions) for six different chemical regimes. Further, Hwang et al. [105] studied the applicability of automatic differentiation techniques in numerical advection schemes in air quality models. They calculated the sensitivity of concentration to a global perturbation of wind velocity in advection models and comparing that with other methods. Their results indicate that ADIFOR-generated code can produce exact sensitivity information up to the machine epsilon, and that the CPU requirements were lower for the ADIFOR method.

ADIFOR has also been applied in several other research areas, such as groundwater modeling [214], in weather modeling [18], in biostatistics [110], and in aeronautics [101].

The details of the ADIFOR approach and its application are presented in Chapter 4 where the coupling of the Stochastic Response Surface Method and ADIFOR is described.

2.6 Need for Alternative Sensitivity/Uncertainty Analysis Methods

Traditional sampling methods for sensitivity and uncertainty analysis, such as the Monte Carlo and Latin Hypercube Sampling, require a substantial number of model runs to obtain a good approximation of the output *pdf*s, especially for cases involving a several inputs. On the other hand, analytical methods require the information about the mathematical equations of a model, and often are restricted in their applicability to cases where the uncertainties are small. Therefore there is a need for a computationally efficient method for uncertainty propagation that is robust and also applicable to a wide range of complex models. The following chapter describes the development of the Stochastic Response Surface Method (SRSM), which is a computationally efficient uncertainty analysis method developed as part of this work.

Chapter 3
THE STOCHASTIC RESPONSE SURFACE METHOD: DEVELOPMENT AND IMPLEMENTATION

3.1 Overview of the Method

The Stochastic Response Surface Method (SRSM), as the name suggests, can be viewed as an extension to the classical deterministic Response Surface Method (RSM), on which extensive literature is available [21,22,70,123]. The main difference between the SRSM and the RSM is that in the former the inputs are random variables, where as in the latter, the inputs are deterministic variables.

The SRSM was developed as a part of this work with the objective of reducing the number of model simulations required for adequate estimation of uncertainty, as compared to conventional methods. This is accomplished by approximating both inputs and outputs of the uncertain system through series expansions of standard random variables; the series expansions of the outputs contain unknown coefficients which can be calculated from the results of a limited number of model simulations.

Conceptually, the propagation of input uncertainty through a model using the SRSM consists of the following steps: (1) input uncertainties are expressed in terms of a set of standard random variables, (2) a functional form is assumed for selected outputs or output metrics, and (3) the parameters of the functional approximation are determined.

The SRSM is founded on the principle that random variables with well-behaved (square-integrable) *pdf*s can be represented as functions of independent random variables [25,49,83,198]. The above integrability requirement is usually satisfied by uncertain quantities of interest, so this method is applicable to a wide variety of of transport-transformation models. Random variables with normal distributions, N(0,1), are often

selected to represent input uncertainties due to the mathematical tractability of functions of these random variables [49, 160]. In the present work these random variables are referred to as "Standard Random Variables" (*srv*s).

Once the inputs are expressed as functions of the selected *srv*s, the output metrics can also be represented as functions of the same set of *srv*s. The minimum number of *srv*s needed to represent the inputs is defined as the "number of degrees of freedom" in input uncertainty. Since model outputs are deterministic functions of model inputs, they have at most the same number of degrees of freedom in uncertainty.

Consider a model $\mathbf{y} = \underline{\underline{\mathcal{F}}}(\mathbf{x})$, where the set of random inputs is represented by the vector $\mathbf{x}$, and a set of selected random outputs or output metrics is represented by the vector $\mathbf{y}$. First, the vector of input random variables is expressed as a function of the form $\mathbf{x} = h(\boldsymbol{\xi})$, where $\boldsymbol{\xi}$ denotes the vector of the selected *srv*s. Then, a functional representation for outputs, of the form $\mathbf{y} = f(\boldsymbol{\xi}, \boldsymbol{a})$, where $\boldsymbol{a}$ denotes a parameter vector, is assumed. The specific method employed in the estimation of parameters of approximation depends on the complexity of the model, which can be broadly categorized into the following:

- Models consisting of simple linear algebraic operators: the outputs can be directly obtained as analytical functions of *srv*s, as $\mathbf{y} = \underline{\underline{\mathcal{F}}}^{-1}(\mathbf{x}) = g(\boldsymbol{\xi})$. For many cases, even a specific functional form for $\mathbf{y}$ need not be assumed a priori.

- Models consisting of nonlinear operators with mathematically manipulable equations: a functional form for outputs, such as $\mathbf{y} = f(\boldsymbol{\xi}, \boldsymbol{a})$, where $\boldsymbol{a}$ denotes the parameters of approximation, is assumed. The parameters are then calculated by substituting the transformed inputs and outputs into the model equations, and minimizing an appropriate norm of the resultant error in approximation. Galerkin's method used in the Deterministic Equivalent Modeling Method [198, 208] is an example of one such case.

- Black-box type models (possibly reflecting the numerical implementation of a

complex mechanistic model that one may decide to treat as a "black-box" input-output system): a functional representation for the output is assumed, as described above, and model outputs for several realizations of the *srv*s are used as "data" for the estimation of parameters of the functional approximation. This approach can be used for cases where it is not possible to directly substitute the input representations into the model. Sometimes this approach may be preferable in the case of complex nonlinear models, where the direct solution requires a significant amount of algebraic manipulation.

Figure 3.1 further offers a schematic illustration of the application of the SRSM for various scenarios.

The following sections describe the SRSM in more detail, and elaborate on the following steps that are involved in the application of the SRSM:

I representation of uncertain inputs,

II representation of model outputs,

III estimation of approximation parameters,

IV calculation of statistical properties of the outputs, and

V evaluation of the approximation of model outputs.

3.2 Step I: Representation of Stochastic Inputs

The first step in the application of the SRSM is the representation of all the model inputs in terms of a set of "standardized random variables". This is analogous to normalization process used in transforming deterministic variables. In this work, normal random variables are selected as *srv*s as they have been extensively studied and their functions are typically well behaved [25, 49, 83, 160]. Here, the *srv*s are selected from a set of independent, identically distributed (*iid*) normal random variables, $\{\xi_i\}_{i=1}^{n}$, where n is the number of independent inputs, and each ξ_i has zero mean and unit variance. When the input random variables are independent, the uncertainty in the ith model input X_i, is expressed directly as a function of the ith *srv*, ξ_i; i.e., a transformation of X_i to ξ_i

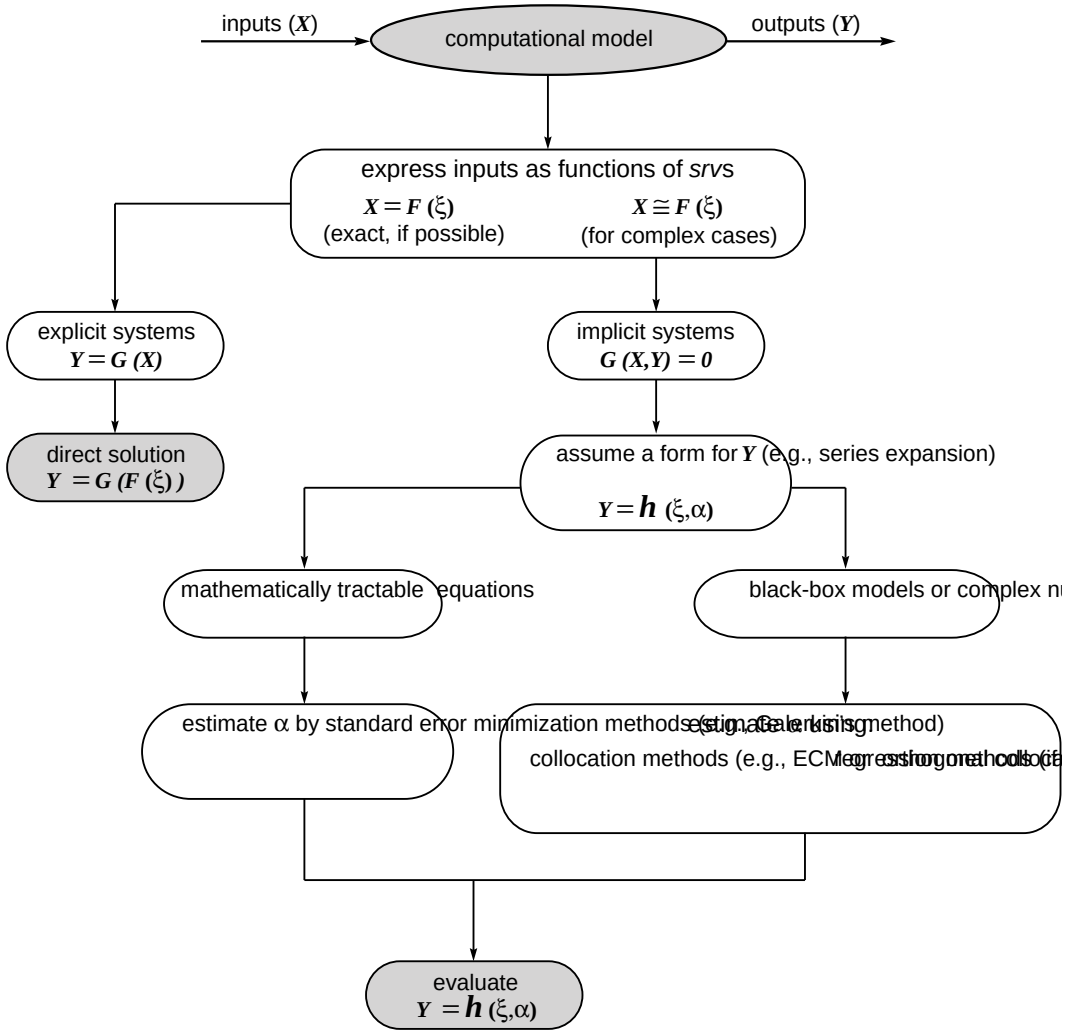


Figure 3.1: Schematic depiction of the Stochastic Response Surface Method

is employed. Such transformations are useful in the standardized representation of the random inputs, each of which could have very different distribution properties.

One of two approaches may be taken to represent the uncertain inputs in terms of the selected *srv*s: (a) direct transformation of inputs in terms of the *srv*s, and (b) series approximations in *srv*s. Devroye [49] presents transformation techniques and approximations for a wide variety of random variables. Input random variables whose transformations are not found in literature can be approximated either by algebraic manipulation or by series expansion techniques. These techniques are described in detail in the following sections.

Table 3.1: Representation of common univariate distributions as functionals of normal random variables

Distribution Type	Transformation[a]
Uniform (a,b)	$a+(b-a)\left(\frac{1}{2}+\frac{1}{2}\text{erf}(\xi/\sqrt{2})\right)$
Normal (μ,σ)	$\mu+\sigma\xi$
Lognormal (μ,σ)	$\exp(\mu+\sigma\xi)$
Gamma (a,b)	$ab\left(\xi\sqrt{\frac{1}{9a}}+1-\frac{1}{9a}\right)^3$
Exponential (λ)	$-\frac{1}{\lambda}\log\left(\frac{1}{2}+\frac{1}{2}\text{erf}(\xi/\sqrt{2})\right)$
Weibull (a)	$y^{\frac{1}{a}}$
Extreme Value	$-\log(y)$

[a] ξ is normal (0,1) and y is exponential (1) distributed

3.2.1 Direct Transformations

Direct transformations for expressing a given random variable as a function of another random variable are typically derived by employing the following relation [160]:

If $Y=h(X)$, where X is a random variable with the *pdf* $f_X(x)$, then the *pdf* of Y, $f_Y(y)$, is given by

$$f_Y(y)=\frac{f_X(x_1)}{|h'(x_1)|}+\cdots+\frac{f_X(x_n)}{|h'(x_n)|}+\cdots \tag{3.1}$$

where $x_1,\ldots,x_n,\ldots$ are the real roots of the equation $y=h(x)$.

Using the above equation, given $f_Y(y)$, and $f_X(x)$, the transformation $Y=h(X)$ is identified by means of trial and error. Transformations are available in the literature for some common distributions. For example, if X is uniform [0,1], then $-\frac{1}{\lambda}\log X$ is exponential (λ) distributed, and if X is normal, the X^2 is gamma (1,1) distributed.

Although a large number of transformations for distributions exist in the literature, they are not directly applicable for the case of normal random variables because of the following reasons: (a) these transformations are optimized for computational speed as

they are used as random number generators, and (b) a majority of these transformations are in terms of the uniform random numbers. However, some transformations in terms of the standard normal random variables are presented by Devroye [49]. In addition, some transformations were derived during the course of the development of the SRSM. Table 3.1 presents a list of transformations for some probability distributions commonly employed in transport-transformation modeling.

3.2.2 Transformation via Series Approximation

When a model input follows a non-standard distribution, direct transformation in terms of standard random variables is often difficult to obtain. This may be true for some distributions for which direct transformations are not available in the literature, and are difficult to derive. In such cases, the approximation of the random variable can be obtained as follows:

- the moments of the random variable are computed based on the given information (e.g., measurements or the given non-standard or empirical *pdf*),
- a series expansion in terms of an *srv* is assumed to represent the random variable,
- the moments of the corresponding series expansion are derived as functions of unknown coefficients,
- the error between the derived moments and the calculated moments is minimized (e.g., through least squares), thus resulting in a set of equations in the unknown coefficients, and
- the process is repeated using a higher order approximation, until the desired accuracy is obtained.

3.2.3 Transformation of Empirical Distributions

In the case of empirical distributions, the probability distribution function of an input is specified in terms of a set of measurements (i.e., a sample of random data points, or a frequency histogram). In some cases the uncertainty in a model input $\boldsymbol{x}$ is specified by an empirical cumulative distribution function (*cdf*), in the form $F_{\boldsymbol{x}}(x) = g(x)$, where

$g(x)$ can be one of the following: (a) an algebraic function, or (b) a look-up table, or (c) a numerical subroutine. In such cases, $\boldsymbol{x}$ can be approximated in terms of an *srv* ξ, as follows:

- Generate a sample of *srv* ξ
- Generate a uniform random variate $\boldsymbol{z}$ from ξ as

$$\boldsymbol{z} = \left(\frac{1}{2} + \frac{1}{2}\text{erf}(\xi/\sqrt{2})\right)$$

  ($\boldsymbol{z}$ corresponds to Uniform(0,1))

- For the calculated value of $\boldsymbol{z}$, obtain the value of the input $\boldsymbol{x}$, for which $F_{\boldsymbol{x}}(x)$ equals the sample value of $\boldsymbol{z}$
- Repeat the above steps for additional samples.

In this process, $\boldsymbol{x}$ is sampled uniformly from the all its percentiles, and hence the samples represent the true distribution of $\boldsymbol{x}$. These steps can be represented by the following transformation equation:

$$\boldsymbol{x} = g^{-1}(\boldsymbol{z}), \text{ where } \boldsymbol{z} = \left(\frac{1}{2} + \frac{1}{2}\text{erf}(\xi/\sqrt{2})\right), \text{ and } g(x) \text{ is the } cdf \text{ of } x \tag{3.2}$$

3.2.4 Transformation of Correlated Inputs

For cases in which the random inputs are correlated, the interdependence of a set of n correlated random variables is often described by a covariance matrix. The covariance matrix, $\boldsymbol{\Sigma}$, as described in Appendix A, is a symmetric matrix with the following elements:

$$\Sigma_{i,j} = \text{E}\{(\boldsymbol{x_i} - \eta_{x_i})(\boldsymbol{x_j} - \eta_{x_j})\},$$

where η_{x_i} and η_{x_j} are the means of $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$, respectively.

The transformation process for n correlated random variables with zero mean, and similar distribution functions, is presented by Devroye [49], and is described briefly here:

To generate a random vector $\boldsymbol{A}$, with n components, with mean 0 and covariance $\boldsymbol{C}$, we start with a vector $\boldsymbol{B}$ comprising of n *iid* unit random variables. Then, assuming a linear transformation $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{B}$, the following conditions are satisfied:

$$\mathrm{E}\{\boldsymbol{B}\boldsymbol{B}^T\} = 1$$

$$\mathrm{E}\{\boldsymbol{A}\} = \boldsymbol{H}\mathrm{E}\{\boldsymbol{B}\} = 0$$

$$\mathrm{E}\{\boldsymbol{A}\boldsymbol{A}^T\} = \boldsymbol{H}\mathrm{E}\{\boldsymbol{B}\boldsymbol{B}^T\}\boldsymbol{H}^T = \boldsymbol{H}\boldsymbol{H}^T = \boldsymbol{C}$$

The problem reduces to finding a nonsingular matrix $\boldsymbol{H}$ such that $\boldsymbol{H}\boldsymbol{H}^T = \boldsymbol{C}$. Since covariance matrices are positive semidefinite, one can always find a nonsingular $\boldsymbol{H}$.

The above method is generalized here for the case of transformation of an arbitrary random vector $\boldsymbol{X}$ with a covariance matrix $\boldsymbol{\Sigma}$. If the ith random variable, $\boldsymbol{x}_i$, has a probability density function $f_i(x)$, mean μ_i, and standard deviation σ_i, the steps involved in the transformation into functions of *srv*s are as follows:

- construct a new vector $\boldsymbol{Y}$ such that $\boldsymbol{y}_i = (\boldsymbol{x}_i - \mu_i)/\sigma_i$
- construct a new covariance matrix $\boldsymbol{\Sigma}^*$ such that $\Sigma^*_{i,j} = \Sigma_{i,j}/(\sigma_i\sigma_j)$
- solve for a nonsingular matrix $\boldsymbol{H}$ such that $\boldsymbol{H}\boldsymbol{H}^T = \boldsymbol{\Sigma}^*$
- obtain transformations $g_i(\xi_i)$ for the ith random variable $\boldsymbol{y}_i$, using transformation methods for univariate random variables, as described in Table 3.1
- obtain the required vector $\boldsymbol{X}$, using $\boldsymbol{x}_i = \mu_i + \sigma_i\boldsymbol{z}_i$, where $\boldsymbol{Z} = \boldsymbol{H}\boldsymbol{Y}$

Thus, given a random vector $\boldsymbol{X}$ with n elements and a covariance matrix $\boldsymbol{\Sigma}$, the sampling from $\boldsymbol{X}$ involves the following steps:

- sampling of n *iid* normal random variables, ξ_i's,
- calculating the vector $\boldsymbol{Y}$ through $y_i = g_i(\xi_i)$,
- calculating the vector $\boldsymbol{Z}$ through $\boldsymbol{Z} = \boldsymbol{H}\boldsymbol{Y}$, and
- calculating the vector $\boldsymbol{X}$ through $\boldsymbol{x}_i = \mu_i + \sigma_i\boldsymbol{z}_i$.

Specific Example: The Dirichlet Distribution

The Dirichlet distribution provides a means of expressing quantities that vary randomly, independent of each other, yet obeying the condition that their sum remains fixed. In environmental modeling, it can be used to represent uncertainty in chemical composition; for example, mole fractions vary independently, but their sum always equals unity. This distribution can also be used to represent uncertainty in compartmental mass fractions in physiological modeling.

Each component of a Dirichlet distribution is specified by a χ^2 distribution, given by

$$f(x;a) = \begin{cases} \dfrac{1}{\Gamma(a/2)2^{a/2}} x^{\frac{a}{2}-1} e^{-x/2}, & x > 0 \\ 0, \text{ elsewhere} \end{cases}$$

The mean of the χ^2 distributions is a, and is related to the mean of the ith component, $x_{i,m}$, and the Dirichlet distribution parameter Θ, as follows:

$$\mathrm{E}[X_i] \quad = \quad x_{i,m}\Theta$$

The ith component can now be directly obtained as:

$x_i = \dfrac{X_i}{X_1 + X_2 + \ldots + X_n}$ where n is the number of components in the Dirichlet distribution.

3.3 Step II: Functional Approximation of Outputs

An output of a model may be influenced by any number of model inputs. Hence, any general functional representation of uncertainty in model outputs, should take into account uncertainties in all inputs. For a deterministic model with random inputs, if the inputs are represented in terms of the set $\{\xi_i\}_{i=1}^{n}$, the output metrics can also be represented in terms of the same set, as the uncertainty in the outputs is solely due to the uncertainty of the inputs [198]. This work addresses one specific form of representation, the series expansion of normal random variables, in terms of Hermite polynomials; the expansion is called "polynomial chaos expansion" [83].

When normal random variables are used as *srv*s, an output can be approximated

by a polynomial chaos expansion on the set $\{\xi_i\}_{i=1}^n$, given by:

$$\begin{aligned} y &= a_0 + \sum_{i_1=1}^{n} a_{i_1}\Gamma_1(\xi_{i_1}) + \sum_{i_1=1}^{n}\sum_{i_2=1}^{i_1} a_{i_1 i_2}\Gamma_2(\xi_{i_1},\xi_{i_2}) \\ &+ \sum_{i_1=1}^{n}\sum_{i_2=1}^{i_1}\sum_{i_3=1}^{i_2} a_{i_1 i_2 i_3}\Gamma_3(\xi_{i_1},\xi_{i_2},\xi_{i_3}) + \dots , \end{aligned} \tag{3.3}$$

where, y is any output metric (or random output) of the model, the $a_{i_1\dots}$'s are deterministic constants to be estimated, and the $\Gamma_p(\xi_{i_1},\dots,\xi_{i_p})$ are multi-dimensional Hermite polynomials of degree p, given by

$$\Gamma_p(\xi_{i_1},\dots,\xi_{i_p}) = (-1)^p\, e^{\frac{1}{2}\boldsymbol{\xi}^T\boldsymbol{\xi}}\, \frac{\partial^p}{\partial\xi_{i_1}\dots\partial\xi_{i_p}} e^{-\frac{1}{2}\boldsymbol{\xi}^T\boldsymbol{\xi}} , \tag{3.4}$$

where $\boldsymbol{\xi}$ is the vector of p *iid* normal random variables $\{\xi_{i_k}\}_{k=1}^p$, that are used to represent input uncertainty. Hermite polynomials on $\{\xi_i\}_{i=1}^n$are random variables, since they are functions of the random variables $\{\xi_i\}_{i=1}^n$. Furthermore, the Hermite polynomials defined on $\{\xi_i\}_{i=1}^n$are orthogonal with respect to an inner product defined as the expectation of the product of two random variables [83]. Thus,

$$\mathrm{E}[\Gamma_p\Gamma_q] = 0 \quad \text{iff} \quad \Gamma_p \neq \Gamma_q ,$$

It is known that the set of multi-dimensional Hermite polynomials forms an orthogonal basis for the space of square-integrable *pdf*s, and that the polynomial chaos expansion is convergent in the mean-square sense [83]. In general, the accuracy of the approximation increases as the order of the polynomial chaos expansion increases. The order of the expansion can be selected to reflect accuracy needs and computational constraints.

For example, an uncertain model output U, can be expressed as second and third order Hermite polynomial approximations, U_2 and U_3 as follows:

$$U_2 = a_{0,2} + \sum_{i=1}^{n} a_{i,2}\xi_i + \sum_{i=1}^{n} a_{ii,2}(\xi_i^2 - 1) + \sum_{i=1}^{n-1}\sum_{j>i}^{n} a_{ij,2}\xi_i\xi_j \tag{3.5}$$

$$\begin{aligned} U_3 &= a_{0,3} + \sum_{i=1}^{n} a_{i,3}\xi_i + \sum_{i=1}^{n} a_{ii,3}(\xi_i^2 - 1) + \sum_{i=1}^{n} a_{iii,3}(\xi_i^3 - 3\xi_i) \\ &+ \sum_{i=1}^{n-1}\sum_{j>i}^{n} a_{ij,3}\xi_i\xi_j + \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ijj,3}(\xi_i\xi_j^2 - \xi_i) + \sum_{i=1}^{n-2}\sum_{j>i}^{n-1}\sum_{k>j}^{n} a_{ijk,3}\xi_i\xi_j\xi_k \end{aligned} \tag{3.6}$$

where n is the number of *srv*s used to represent the uncertainty in the model inputs, and the coefficients $a_{i,m2}, a_{i,m3}, a_{ij,m2}, a_{ij,m3}$, and $a_{ijk,m3}$ are the coefficients to be estimated.

From the above equation, it can be seen that the number of unknowns to be determined for second and third order polynomial chaos expansions of dimension n, denoted by N_2 and N_3, respectively, are:

$$N_2 = 1 + 2n + \frac{n(n-1)}{2} \tag{3.7}$$

$$N_3 = 1 + 3n + \frac{3n(n-1)}{2} + \frac{n(n-1)(n-2)}{6}. \tag{3.8}$$

3.4 Step III: Estimation of Parameters in the Functional Approximation

The unknown coefficients in the polynomial chaos expansion ($a_{ik\ldots}$'s) can be estimated by one of the following methods, depending on the complexity of the model:

- For cases in which the model is invertible, output metrics can be directly calculated as explicit functions of the *srv*s (ξ_is); the expressions for input variables in terms of ξ_i s are substituted into the inverted model to obtain explicit functions for the output metrics in terms of ξ_i s. In such simple cases, there is no need to assume a functional form for the output metrics.

- For nonlinear operators with mathematically manipulable equations, the unknown coefficients in the polynomial chaos expansion can be determined by minimizing an appropriate norm of the residual, after substituting the transformed inputs into the model equations. Gelarkin's method [208] is commonly used with weight function corresponding to the expectation of the random variables [198].

- For cases in which the model equations are not easy to manipulate, or when the model is of "black-box" type, the unknown coefficients can be obtained by a collocation method [198, 208]. This method imposes the requirement that the estimates of model outputs are exact at a set of selected collocation points, thus making the residual at those points equal to zero. The unknown coefficients are estimated by equating model outputs and the corresponding polynomial chaos

expansion, at a set of collocation points in the parameter space; the number of collocation points should be equal to the number of unknown coefficients to be found. Thus, for each output metric, a set of linear equations results with the coefficients as the unknowns; these equations can be readily solved using standard linear solvers.

3.4.1 Deterministic Equivalent Modeling Method/Probabilistic Collocation Method (DEMM/PCM)

The Deterministic Equivalent Modeling Method (DEMM) [198] incorporates the polynomial chaos expansion technique, along with symbolic computation methods and compiler technology, to provide a prototype language for automated uncertainty analysis. This method uses the Galerkin's method to relate the input and output approximations for models with tractable equations. Depending on the complexity of the model, DEMM uses either the Galerkin's method or the Probabilistic Collocation Method (PCM), described here.

As a part of DEMM, the Probabilistic Collocation Method (PCM) [159, 199], is used for applying DEMM in the case of black-box models. The coefficients of the polynomial chaos expansion are obtained using the model outputs at selected collocation points. Then, the next order polynomial chaos expansion is employed, and the solution process is repeated. If the estimates of the *pdf*s of the output metrics of concern, are approximately equal, the expansion is assumed to have converged; the higher order approximation is used to estimate the *pdf*s of output metrics.

In DEMM/PCM the collocation points are selected following the orthogonal collocation method suggested by Villadsen and Michelsen [208]. The collocation points correspond to the roots of the polynomial of one degree higher than the order of the polynomial chaos expansion. For one-dimensional problems, this method gives the same results as Galerkin's method [208], and hence is regarded as an "optimal method". This approach is adapted for multi-dimensional cases in DEMM [198]. For example, in order to solve for a two dimensional second order polynomial chaos expansion, the roots of the third order Hermite polynomial, $\sqrt{3}$,$-\sqrt{3}$ and 0 are used, hence the possible collocation

points are (0,0), $(-\sqrt{3},0)$, $(0,\sqrt{3})$, $(0,-\sqrt{3})$, $(-\sqrt{3},-\sqrt{3})$, $(\sqrt{3},\sqrt{3})$, $(\sqrt{3},0)$, $(-\sqrt{3},\sqrt{3})$ and $(\sqrt{3},-\sqrt{3})$. There are nine possible collocation points, but from Equations 3.6 and 3.6 with n=2 and taking terms only up to Γ_2, there are only six unknowns. Similarly, for higher dimension systems and higher order approximations, the number of available collocation points is always greater than the number of collocation points needed, which introduces a problem of selecting the appropriate collocation points. In the absence of selection criteria at present, the collocation points are typically selected at random from the set of available points.

Limitations of DEMM/PCM

As noted above, the selection of the required collocation points in an efficient manner from the large number of possible points is not addressed in the present implementation of DEMM. In fact, as the number of degrees of freedom increases, the number of available collocation points increases exponentially. For example, for the case of $n = 6$, using Equation 3.3, the number of collocation points required for second and third order expansions are 28 and 84 respectively. Since the collocation points are selected from the combinations of the roots of one order higher Hermite polynomial, the number of collocation points available for second and third order expansions are $3^6 = 729$ and $4^6 = 4096$, respectively - there are three roots for a third order Hermite polynomial (used for obtaining collocation points for a second order expansion), and there are six variables, hence the number of possible collocation points is 3^6, and similarly 4^6 for a third order expansion. As the number of inputs and the order of expansion increase, the number of available collocation points increases exponentially.

In principle, any choice of collocation points from the available ones, should give adequate estimates of the polynomial chaos expansion coefficients. However, different combinations of collocation points may result in substantially different estimates of the *pdf*s of output metrics; this poses the problem of the optimal selection of collocation points. Furthermore, in a computational setting, some of the collocation points could be outside the range of the algorithmic or numerical applicability of the model, and model results at these points cannot be obtained. These issues are addressed in the following

section, where an algorithm for collocation point selection method is described.

Another shortcoming of the standard collocation technique suggested in DEMM is that sometimes the roots of the Hermite polynomials do not correspond to high probability regions and these regions are therefore not adequately represented in the polynomial chaos expansion. For example, for a third order approximation, the roots of the fourth order Hermite polynomial, namely $\sqrt{3+\sqrt{6}}, -\sqrt{3+\sqrt{6}}, \sqrt{3-\sqrt{6}}$, and $-\sqrt{3-\sqrt{6}}$ are used as the collocation points. However, the origin corresponds to the region of highest probability for a normal random variable. Hence, the exclusion of the origin as a collocation point could potentially lead to a poor approximation.

The limitations of DEMM/PCM are shown in the context of a case study involving a human Physiologically Based Pharmacokinetic (PBPK) Model, and the results are presented in Section 5.1.4.

3.4.2 Efficient Collocation Method (ECM)

This method addresses some limitations of DEMM/PCM. Here, the collocation points for the estimation of the coefficients of the polynomial chaos expansion are selected based on a modification of the standard orthogonal collocation method of [198, 208]. The points are selected so that each standard normal random variable ξ_i takes the values of either zero or of one of the roots of the higher order Hermite-polynomial.

A simple heuristic technique is used to select the required number of points from the large number of potential candidates: for each term of the series expansion, a "corresponding" collocation point is selected. For example, the collocation point corresponding to the constant is the origin; i.e., all the standard normal variables (ξ's) are set to value zero. For terms involving only one variable, the collocation points are selected by setting all other ξ's to zero value, and by letting the corresponding variable take values as the roots of the higher order Hermite polynomial. For terms involving two or more random variables, the values of the corresponding variables are set to the values of the roots of the higher order polynomial, and so on. If more points "corresponding" to a set of terms are available than needed, the points which are closer to the origin are preferred, as they fall in regions of higher probability. Further, when there is still an

unresolved choice, the collocation points are selected such that the overall distribution of the collocation points is more symmetric with respect to the origin. If still, more points are available, the collocation point is selected randomly.

The advantage of this method is that the behavior of the model is captured reasonably well at points corresponding to regions of high probability. Furthermore, singularities can be avoided in the resultant linear equations for the unknown coefficients, as the collocation points are selected to correspond to the terms of the polynomial chaos expansion. Thus, the *pdf*s of the output metrics are likely to be approximated better than by a random choice of collocation points, while following a technique similar to the "orthogonal collocation" approach.

Applicability and Limitations of ECM

The ECM method was applied to a case study involving a human Physiologically Based Pharmacokinetic (PBPK) model, which is presented in Section 5.1.4. This method resulted in accurate estimates of the output *pdf*s while requiring much fewer model simulations as compared to standard Monte Carlo methods, as shown in Figures 5.4–5.7 in Section 5.1.4.

The ECM method was applied to a more complex model, the two-dimensional atmospheric photochemical plume model, the Reactive Plume Model, version 4, (RPM-IV), as described in Section 5.2.3. For that case study, the ECM method failed to converge for a third order polynomial chaos approximation. This is attributed to the inherent instability of the collocation based approaches, as described by Atkinson [6].

3.4.3 Regression Based SRSM

Collocation methods are inherently unstable, especially with polynomial approximations of high orders, since the requirement that the polynomial (curve or a surface) has to pass through all the collocation points. Thus any one collocation point in the model space could significantly alter the behavior of the polynomial [6]. In this context, regression-based methods provide slightly more computationally expensive, but robust

means of estimation of coefficients of the functional approximation, because model results at more points are used in this approach, and the influence of each sample point is moderated by all other sample points. This extension to the collocation method uses regression in conjunction with an improved input sampling scheme to estimate the unknown coefficients in the polynomial expansion. In the regression based method, a set of sample points is selected in the same manner as the ECM method, as described earlier. The number of sample points selected must be higher than the number of unknown coefficients to be estimated; selecting a number of points equaling twice the number of coefficients is recommended for obtaining robust estimates, and is the approach used in the case studies presented in this work. The model outputs at the selected sample points are equated with the estimates from the series approximation, resulting in a set of linear equations with more equations than unknowns. This system of equations is then solved using the singular value decomposition method [163].

The regression method has the advantage that it is more robust than the collocation method, but the method requires a higher number of model simulations to obtain estimates of output uncertainty. The advantages of this method are demonstrated by comparing the results of this method with those of ECM for a the case study involving RPM-IV, an atmospheric photochemical plume model, described in Section 5.2.3.

Based on preliminary case studies (presented in Chapter 5), the regression based method is found to be robust and the method resulted in good approximations of model outputs. Hence this method is recommended for use in general. The higher computational burden, compared to the ECM, is offset by the robustness this method provides for many cases. Throughout the rest of this document, the ECM method is denoted as "ECM" and the regression based method is denoted simply as "SRSM". Unless otherwise specified, the term SRSM in the case studies refers to a regression based SRSM.

3.5 Step IV: Estimation of the Statistics of the Output Metrics

Once the coefficients used in the series expansion of model outputs are estimated, the statistical properties of the outputs, such as the density functions, moments; joint

densities; joint moments; correlation between two outputs, or between an output and an input; etc., can be readily calculated. One way of accomplishing this is through the generation of a large number of realizations of the *srv*s, and the calculation of the values of inputs and outputs from the transformation equations. This results in a large number of samples of inputs and outputs. These samples can then be statistically analyzed using standard methods.

It must be noted that the calculation of model inputs and outputs involves evaluation of simple algebraic expressions, and does not involve model runs, and hence substantial savings in the computer time are accomplished.

As an example, if the inputs $\boldsymbol{x}_i$'s are represented as $\boldsymbol{x}_i = F_i(\xi_i)$, and if the outputs $\boldsymbol{y}_j$'s are estimated as $\boldsymbol{x}_i = G_i(\xi_1, \xi_2 \ldots \xi_n)$, then the following steps are involved in the estimation of the statistics of the inputs and outputs.

- generation of a large number of samples of $(\xi_1\ \xi_2\ \ldots\ \xi_{n,i})$,
- calculation of the values of input and output random variables from the samples,
- calculation of the moments using Equation 3.9, and
- calculation of density functions and joint density functions using the sample values.

From a set of N samples, the moments of the distribution of an output $\boldsymbol{y}_i$ can be calculated as follows:

$$\begin{aligned}
\eta_{y_i} &= \text{Mean}(\boldsymbol{y}_i) &&= \text{E}\{\boldsymbol{y}_i\} &&= \frac{1}{N}\sum_{j=1}^{N} y_{i,j} \\
\sigma^2_{y_i} &= \text{Var}(\boldsymbol{y}_i) &&= \text{E}\{(\boldsymbol{y}_i - \eta_{y_i})^2\} &&= \frac{1}{N-1}\sum_{j=1}^{N}(y_{i,j} - \eta_{y_i})^2 \\
\gamma_{1,y_i} &= \text{Skew}(\boldsymbol{y}_i) &&= \frac{1}{\sigma^3_{y_i}}\text{E}\{(\boldsymbol{y_i} - \eta_{y_i})^3\} &&= \frac{1}{N\sigma^3_{y_i}}\sum_{j=1}^{N}(y_{i,j} - \eta_{y_i})^3 \\
\gamma_{2,y_i} &= \text{Kurt}(\boldsymbol{y}_i) &&= \frac{1}{\sigma^4_{y_i}}\text{E}\{(\boldsymbol{y_i} - \eta_{y_i})^4\} &&= \frac{1}{N\sigma^4_{y_i}}\sum_{j=1}^{N}(y_{i,j} - \eta_{y_i})^4
\end{aligned} \tag{3.9}$$

Further, the correlation coefficient of an input $\boldsymbol{x}_k$ and an output $\boldsymbol{y}_i$ can be calculated

using the following:

$$r_{x_k,y_i} = \frac{\mathrm{E}\{(\boldsymbol{x_k} - \eta_{x_k})(\boldsymbol{y}_i - \eta_{y_i})\}}{\sigma_{x_k}\sigma_{y_i}} = \frac{1}{N\sigma_{x_k}\sigma_{y_i}}\sum_{j=1}^{N}(x_{k,j} - \eta_{x_k})(y_{i,j} - \eta_{y_i}) \tag{3.10}$$

Similarly, higher moments of outputs, or the correlation between two outputs, or between an input and an output can be directly calculated.

3.6 Step V: Evaluation of Convergence, Accuracy and Efficiency of Approximation

Once the coefficients of the polynomial chaos expansion are obtained using one of the methods described above, the convergence of the approximation is determined through comparison with the results from a higher order approximation. The next order polynomial chaos expansion is used, and the process for the estimation of unknown coefficients is repeated. If the estimates of *pdf*s of output metrics agree closely, the expansion is assumed to have converged, and the higher order approximation is used to calculate the *pdf*s of output metrics. If the estimates differ significantly, yet another series approximation, of the next order, is used, and the entire process is repeated until convergence is reached.

In the present work, the accuracy of the approximation is evaluated by comparing the results from SRSM with the results obtained from Monte Carlo analyses. This verification is performed to ensure that the approximation converges to the true distribution. The efficiency of the SRSM is further evaluated by comparing the number of simulations required for the SRSM method, with the number required for a Latin Hypercube sampling method.

3.7 An Illustration of the Application of the SRSM

As an illustration, consider a computational model with three independent random inputs X_1, X_2, and X_3, and two outputs Y_1 and Y_2, where the input random variables

X_1, X_2 and X_3 are given by

$$\begin{aligned} X_1 &= \text{Uniform}(p_1, q_1) \\ X_2 &= \text{Lognormal}(p_2, q_2) \\ X_3 &= \text{Gamma}(p_3, q_3) \end{aligned} \tag{3.11}$$

The input random variables can be represented by three *srv*s, ξ_1, ξ_2, ξ_3, using Table 3.1 as follows:

$$\begin{aligned} X_1 &= p_1 + (q_1 - p_1)\left(\frac{1}{2} + \frac{1}{2}\text{erf}(\xi_1/\sqrt{2})\right) \\ X_2 &= \exp(p_2 + q_2\xi_2) \\ X_3 &= p_3 q_3 \left(\xi_3\sqrt{\frac{2}{9p_3}} + 1 - \frac{2}{9p_3}\right)^3 \end{aligned} \tag{3.12}$$

Where ξ_1, ξ_2, ξ_3 are *iid* N(0,1) random variables. A second order polynomial chaos approximation for Y_1 and Y_2 in terms of ξ_1, ξ_2, ξ_3 is given by

$$\begin{aligned} Y_1 &= a_0 + a_1\xi_1 + a_2\xi_2 + a_3\xi_3 + a_4(\xi_1^2 - 1) + a_5(\xi_2^2 - 1) + a_6(\xi_3^2 - 1) \\ &\quad + a_7\xi_1\xi_2 + a_8\xi_2\xi_3 + a_9\xi_1\xi_3 \\ Y_2 &= b_0 + b_1\xi_1 + b_2\xi_2 + b_3\xi_3 + b_4(\xi_1^2 - 1) + b_5(\xi_2^2 - 1) + b_6(\xi_3^2 - 1) \\ &\quad + b_7\xi_1\xi_2 + b_8\xi_2\xi_3 + b_9\xi_1\xi_3 \end{aligned} \tag{3.13}$$

In order to estimate the 10 unknown coefficients (for each output) from the above equation, the selection of a set of N sample points (equaling about twice the number of sample points, i.e., 20 in this case) is recommended for regression based SRSM, in the form

$$(\xi_{1,1}\ \xi_{2,1}\ \xi_{3,1}), (\xi_{1,2}\ \xi_{2,2}\ \xi_{3,2}), \ldots, (\xi_{1,N}\ \xi_{2,N}\ \xi_{3,N})$$

These sample points can be readily generated using a simple algorithm developed as a part of the SRSM implementation. The selection of the sample points is discussed in detail in Section 3.4.2, in the context of the ECM method.

These sample points correspond to the original model input samples, $(x_{1,1}\ x_{2,1}\ x_{3,1})\ldots$

$(x_{1,N}\ x_{2,N}\ x_{3,N})$, as follows:

$$\begin{pmatrix} \xi_{1,i} \\ \xi_{2,i} \\ \xi_{3,i} \end{pmatrix} \rightarrow \begin{pmatrix} x_{1,i} \\ x_{2,i} \\ x_{3,i} \end{pmatrix} = \begin{pmatrix} p_1 + (q_1 - p_1)\left(\frac{1}{2} + \frac{1}{2}\mathrm{erf}(\xi_{1,i}/\sqrt{2})\right) \\ \exp(p_2 + q_2\xi_{2,i}) \\ p_3 q_3 \left(\xi_{3,i}\sqrt{\frac{2}{9p_3}} + 1 - \frac{2}{9p_3}\right)^3 \end{pmatrix} \tag{3.14}$$

for $i = 1 \ldots N$. After obtaining the original model input sample points, the model simulation is performed at the points given by $(x_{1,1}\ x_{2,1}\ x_{3,1}) \ldots (x_{1,N}\ x_{2,N}\ x_{3,N})$. Then the outputs at these sample points, $y_{1,1} \ldots y_{1,N}$ and $y_{2,1} \ldots y_{2,N}$, are used to calculate the coefficients $a_1 \ldots a_{10}$ and $b_1 \ldots b_{10}$ by solving the following linear equations through singular value decomposition.

$$\boldsymbol{Z}^T \times \begin{bmatrix} a_0 & b_0 \\ a_1 & b_1 \\ a_2 & b_2 \\ \vdots & \vdots \\ a_8 & b_8 \\ a_9 & b_9 \end{bmatrix} = \begin{bmatrix} y_{1,1} & y_{2,1} \\ y_{1,2} & y_{2,2} \\ \vdots & \vdots \\ y_{1,N} & y_{2,N} \end{bmatrix}, \text{ where} \tag{3.15}$$

$$\boldsymbol{Z} = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ \xi_{1,1} & \xi_{1,2} & \xi_{1,3} & \ldots & \xi_{1,N} \\ \xi_{2,1} & \xi_{2,2} & \xi_{2,3} & \ldots & \xi_{2,N} \\ \xi_{3,1} & \xi_{3,2} & \xi_{3,3} & \ldots & \xi_{3,N} \\ \xi_{1,1}^2 - 1 & \xi_{1,2}^2 - 1 & \xi_{1,3}^2 - 1 & \ldots & \xi_{1,N}^2 - 1 \\ \xi_{2,1}^2 - 1 & \xi_{2,2}^2 - 1 & \xi_{2,3}^2 - 1 & \ldots & \xi_{2,N}^2 - 1 \\ \xi_{3,1}^2 - 1 & \xi_{3,2}^2 - 1 & \xi_{3,3}^2 - 1 & \ldots & \xi_{3,N}^2 - 1 \\ \xi_{1,1}\xi_{2,1} & \xi_{1,2}\xi_{2,2} & \xi_{1,3}\xi_{2,3} & \ldots & \xi_{1,N}\xi_{2,N} \\ \xi_{2,1}\xi_{3,1} & \xi_{2,2}\xi_{3,2} & \xi_{2,3}\xi_{3,3} & \ldots & \xi_{2,N}\xi_{3,N} \\ \xi_{1,1}\xi_{3,1} & \xi_{1,2}\xi_{3,2} & \xi_{1,3}\xi_{3,3} & \ldots & \xi_{1,N}\xi_{3,N} \end{bmatrix} \tag{3.16}$$

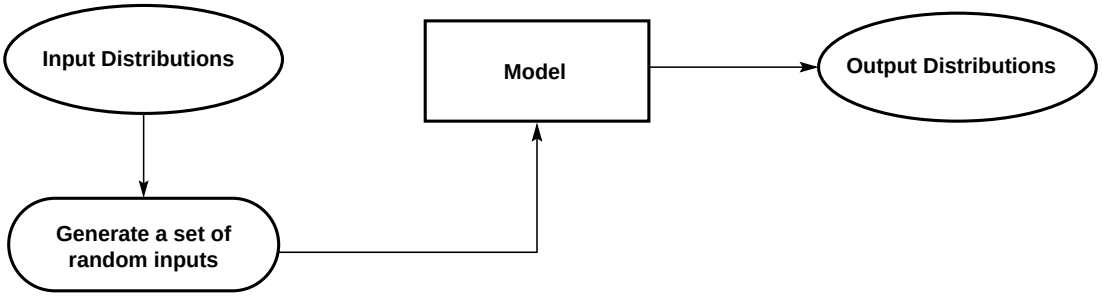


Figure 3.2: Schematic depiction of the steps involved in the application of conventional Monte Carlo method

In the above equations, $\boldsymbol{Z}$ can be calculated from the values of ξ's at each sample point, whereas, $y_{1,i}$ and $y_{2,i}$ are the corresponding model outputs. Thus, the only unknowns in Equation 3.15, the coefficients a_i's and b_i's, can be readily estimated. Once the coefficients are estimated, the distributions of Y_1 and Y_2 are fully described by the polynomial chaos expansions as shown in Equation 3.14.

The next step involves determination of the statistical properties of the outputs, such as the individual moments, joint moments, correlations between the outputs and correlations between inputs and outputs. This process is done by numerically generating a large number of random samples of $(\xi_{1,i}\ \xi_{2,i}\ \xi_{3,i})$, calculating the values of model inputs and outputs for those samples, and computing the statistical properties from the values of inputs and outputs as described in Section 3.5.

3.8 Computational Implementation

3.8.1 Implementation of the Conventional Uncertainty Propagation Methods

Modules for the application of conventional Monte Carlo and the Latin Hypercube Sampling methods were developed as part of this work. Though these modules are mainly used in this work for the evaluation of alternative methods, they can be used in isolation.

As shown in Figure 3.2, the conventional Monte Carlo method involves the following steps: (a) obtaining random samples from the probability distributions of the

inputs, (b) performing model simulations for the combination of the sampled inputs, and (c) statistically analyzing the model outputs.

The random numbers for sampling the input distributions were generated depending on the case study considered. For all the cases, the normal random numbers were generated through the function `GASDEV` provided with the mathematical routines of the Numerical Recipes [163]. Gamma random variables were generated by first generating a normal random number, and applying the transformation shown in Table 3.1. In the case of Dirichlet distributions, the independent random numbers were generated first, and the corresponding ratios were calculated using Equation 3.2.4. In one specific case, in the study involving a physiologically based pharmacokinetic model, described in Section 5.1, the model was implemented on the SimuSolv modeling platform [58], and the Monte Carlo method involved sampling from random numbers generated by the pseudo-gaussian random number generator function in SimuSolv, `R-GAUSS`.

In all the case studies described in this work, the original numerical models were modified accordingly to accept the uncertain model inputs and parameters as inputs from a file, and model simulations were performed at the sampled values of the inputs. The outputs of the simulations were then analyzed by "binning" them into a set of intervals; a bin number associated with an interval contains the number of outputs realizations that belong to that interval, from all the model simulations performed. The ratio of the bin number to the total number of samples, gives the probability that the output belongs to that interval. The ratio of this probability to the interval size gives the probability density at the mid-point of the interval. The probability density is then plotted against the range of values that the output can belong to, as presented in the case studies. Further, FORTRAN routines to calculate the mean, variance, and higher moments were also implemented as part of the SRSM; these were based on equations presented in Section 3.5.

The Latin Hypercube Sampling (LHS) method was developed and implemented from basic principles, mainly because of the lack of availability of a standardized routine readily usable on Unix operating system. The EPA approved routine for the LHS, developed by Iman et al. [108], is developed for the VAX/VMS operating system, and

other commercial software such as Crystal ball [46] and @RISK [218], for the PCs have LHS routines embedded in the modeling platform (typically spreadsheet based), and thus cannot be used in conjunction with stand-alone numerical codes. The main aspects of the LHS implementation developed here are presented in the following.

In order to generate samples from input distributions, the range of probable values for each uncertain input parameter is divided into M segments of equal probability. Thus, the whole parameter space, consisting of N parameters, is partitioned into M^N cells, with each cell corresponding to equal probability regions from the input sampling space. For example, for the case of 3 input parameters and 4 segments, the parameter space is divided into $4 \times 4 \times 4$ cells. The next step involves the generation of M samples from M^N cells. This is accomplished as follows: first, a random sample is generated, and its cell number is calculated. The cell number indicates the segment number the sample belongs to, with respect to each of the parameters. For example, a cell number (2,1,2) indicates that the sample lies in the segment 2 with respect to first parameter, segment 1 with respect to second parameter, and segment 2 with respect to third parameter. At each successive step, a random sample is generated, and is accepted only if it does not agree with any previous sample on any of the segment numbers. Thus, no two samples have any input corresponding to the same segment. The advantage of this approach is that the random samples are generated from all the ranges of possible values, thus providing samples from the tails of the probability distributions (i.e., regions that have very low probabilities of occurrence, but correspond to extreme values). Further, for every M samples, each parameter is sampled from each of its M subranges.

The LHS FORTRAN subroutine developed here requires the following inputs: (a) the total number of samples to be generated, N_{tot}, (b) the number of random inputs, N, and (c) the input probability distributions, consisting of the distribution type (a number), and the distribution parameters (dependent on the distribution type). Further, the number of ranges the parameter range of each input should be divided (M) is also required as an input. The routine generates a set of M sample points in each iteration, till N_{tot} samples are generated. A typical value of M used in the case studies is about 50. A much higher value results in too small a range for each sampled input,

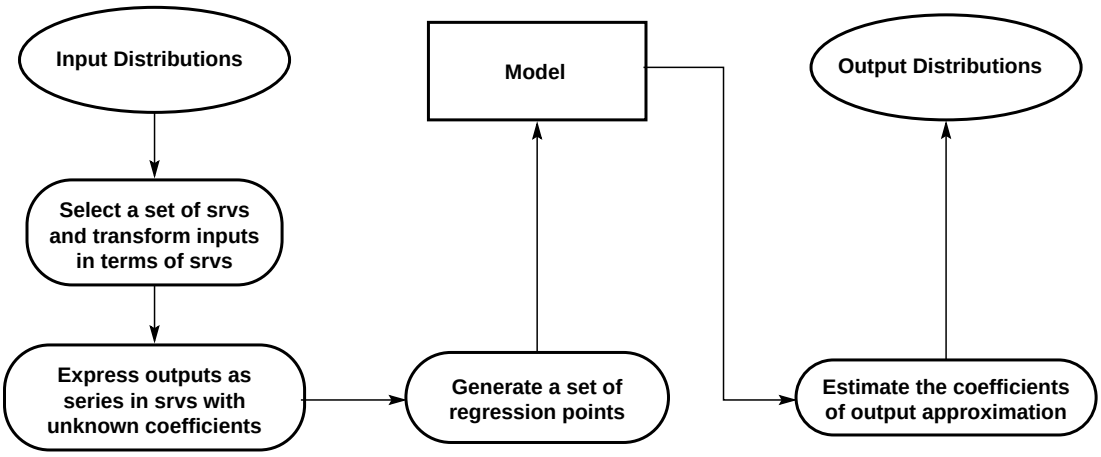


Figure 3.3: Schematic depiction of the steps involved in the application of the Stochastic Response Surface Method

requiring a large number of trial samples before an acceptable sample is generated, whereas much smaller values do not adequately represent all ranges of possible values, especially the tails of the distributions. *The routine, along with all other routines is included in the CDROM accompanying this document.*

3.8.2 Implementation of the Stochastic Response Surface Method

The steps involved in the application of the SRSM are presented in Figure 3.3. The implementation of the SRSM is based on the following routines:

- A symbolic calculation routine in Maple [94], to generate polynomial chaos expansions of arbitrary order, consisting of an arbitrary number of parameters.
- A FORTRAN subroutine to generate polynomial chaos expansions of up to fifth order for an arbitrary number of parameters. The extension to higher orders is straightforward, but not pursued because in all the case studies considered, the third order expansion was the highest required. This routine is developed using the results from the Maple routine.
- A FORTRAN subroutine to generate the sample points based on the input distributions and the order of the polynomial chaos expansion selected.

- A set of FORTRAN utilities to sample distributions in the form of the *srv*s, to calculate the corresponding transformed random variables, and to calculate the numerical values of the polynomial chaos expansions.

- A FORTRAN subroutine to perform collocation or regression (based on the scenario) on a set of sample points and the corresponding model outputs. These are used to estimate the coefficients of a polynomial chaos expansion.

- A FORTRAN subroutine to generate the probability density functions from the coefficients of a polynomial chaos expansion in *srv*s, by obtaining a large number of samples of *srv*s, and calculating the statistics of the corresponding outputs.

- FORTRAN subroutines to calculate the moments, percentiles, and other statistical properties of both the inputs and outputs.

3.9 Implementation of a Web Interface to the SRSM

All numerical routines associated with the development of the SRSM have been incorporated into a a stand-alone application, that could be used for uncertainty propagation with black-box models. This has also been implemented as a "Web Based" tool, that can be run from remote web browsers. The interface was implemented by using the Common Gateway Interface (CGI) programming using the Perl language. An operational version can be accessed through the location

`http://www.ccl.rutgers.edu/srsm.html`

The web server is located in the Computational Chemodynamics Laboratory, at the Environmental and Occupational Sciences Institute, a joint project of University of Medicine and Dentistry, New Jersey, and Rutgers, The State University of New Jersey.

Here, the application of the SRSM method for a simple black-box model is presented. In order to use the SRSM, first the number of inputs, their probability distributions, the number of outputs, and the order of polynomial chaos expansion desired, have to be specified, as shown in Figure 3.4. The program responds with the recommended sample points, and the model outputs at these sample points have to be input, as shown in

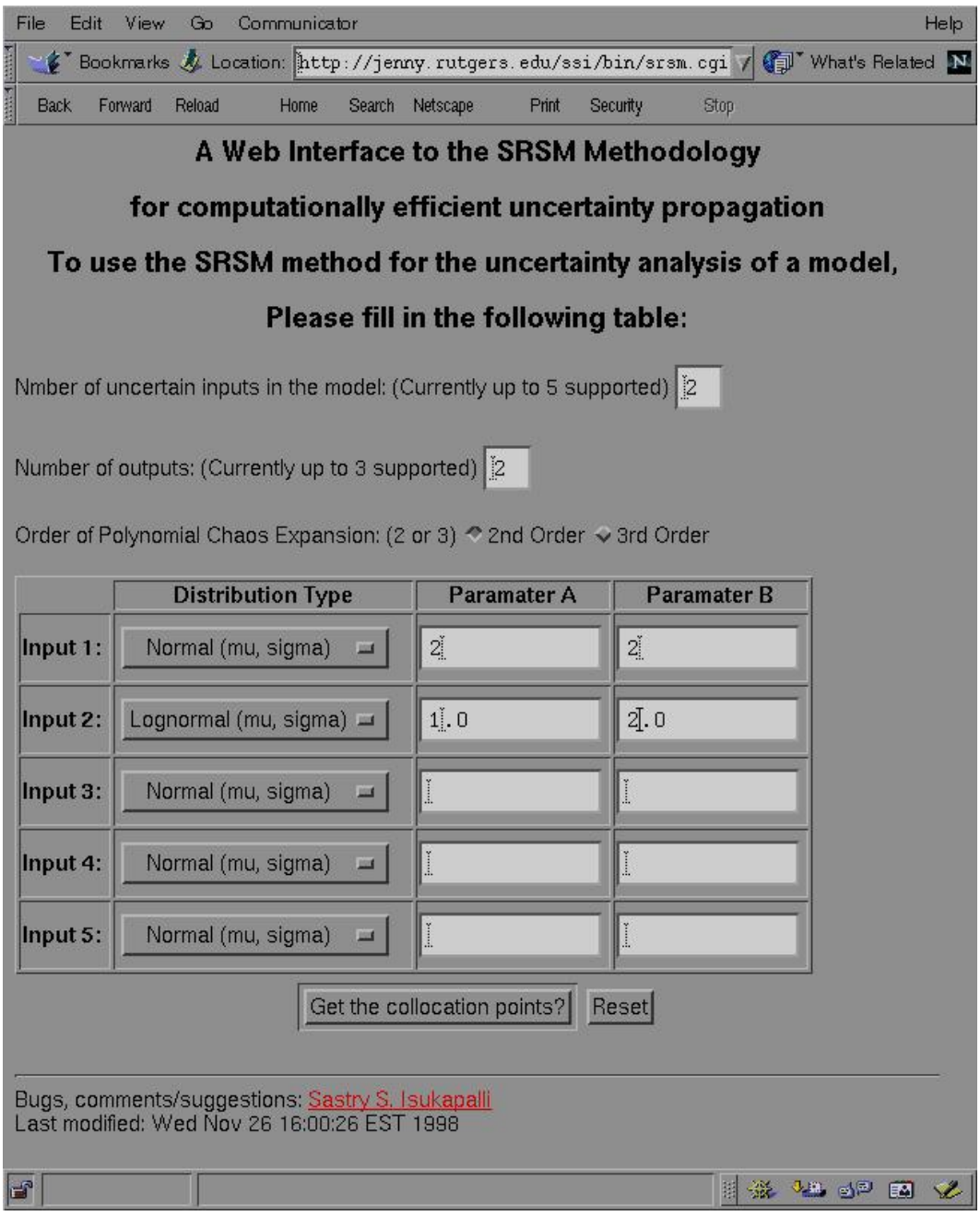


Figure 3.4: Application of the web based SRSM tool: specification of input uncertainties

Figure 3.5. Additionally, a range for plotting the output probability density functions is also required as input. The program uses the model outputs at these sample points and calculates, and displays the probability density functions for each of the outputs, as shown in Figure 3.6.

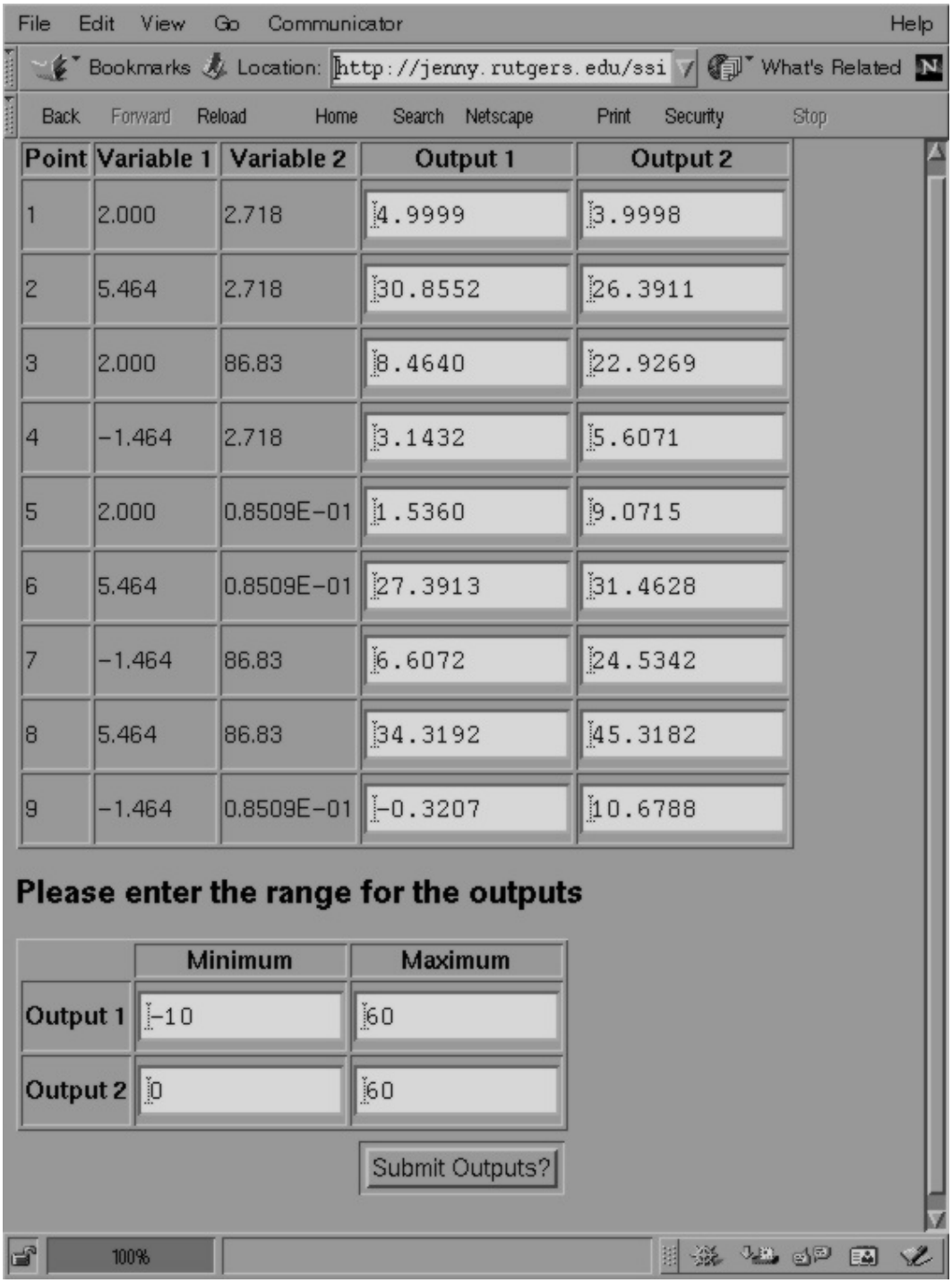


Figure 3.5: Application of the web based SRSM tool: recommended sample points and the corresponding model outputs

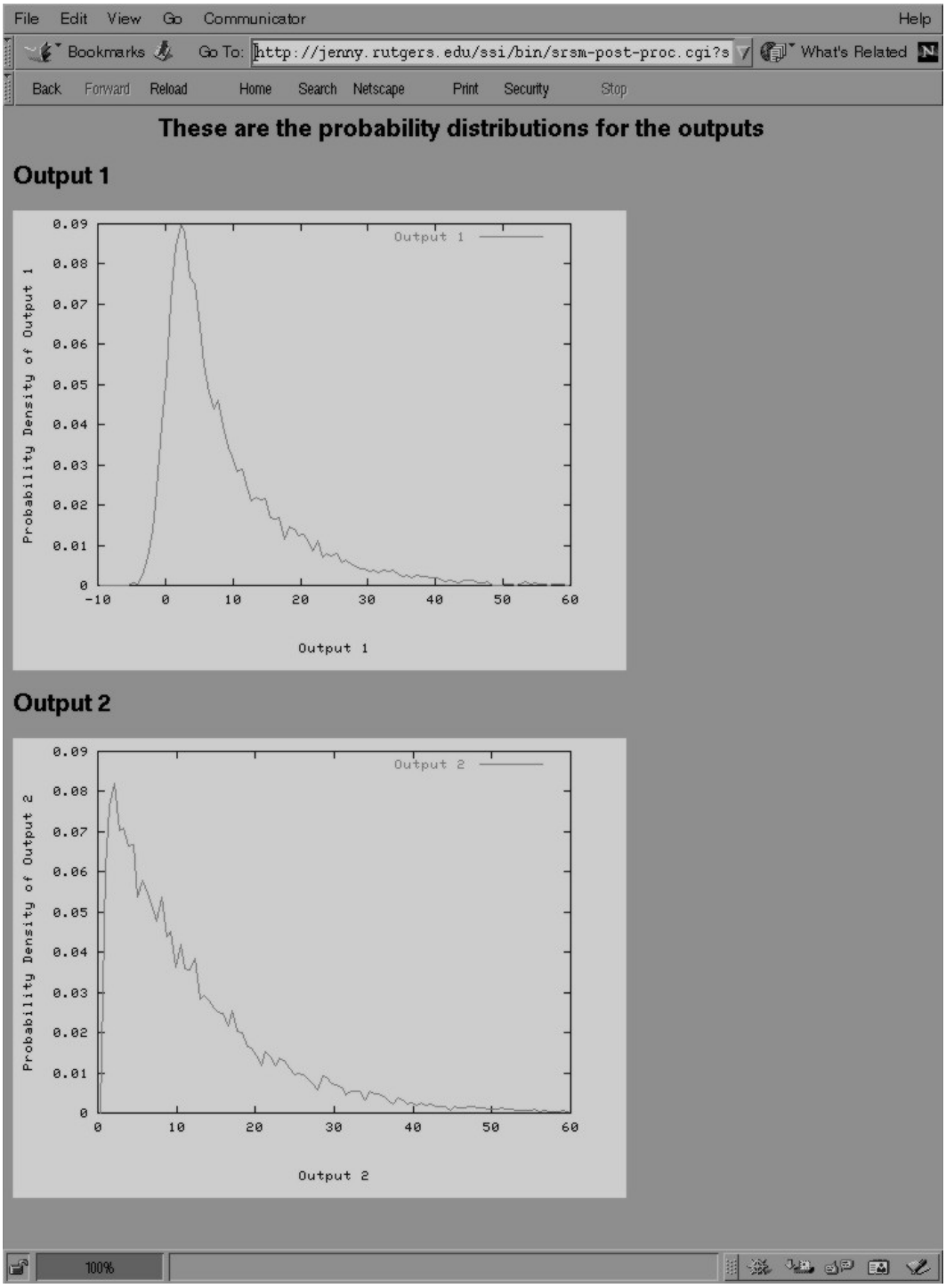


Figure 3.6: Application of the web based SRSM tool: estimates of the pdfs of the outputs

Chapter 4

COUPLING OF THE SRSM WITH SENSITIVITY ANALYSIS METHODS: DEVELOPMENT AND IMPLEMENTATION OF SRSM-ADIFOR

The objective of this effort is to couple the Stochastic Response Surface Method (SRSM) with sensitivity analysis methods, to obtain savings in the computer resources required for uncertainty propagation. In the construction of an approximate model, sensitivity analysis methods can provide information that can help reduce the number of model simulations required to estimate the approximation parameters. Consequently, the use of sensitivity information in conjunction with uncertainty propagation methods can reduce the model simulations required to estimate uncertainties in model outputs.

The rationale for the coupling of the SRSM with a sensitivity analysis method is illustrated in Figure 4.1 for the simple case of approximation of a curve in a 2-D Euclidean space using a set of sample points. Figure 4.1 (A) represents the model response, which can be approximated accurately using the five sample points, as shown in Figure 4.1 (B). In contrast, using three sample points, only a poor approximation as shown in Figure 4.1 (C) can be obtained. However, using the derivative information at the three sample points (tangents at these points), a good approximation, as shown in Figure 4.1 (D) can be obtained. As an extension of this reasoning to multi-dimensional systems, the partial derivatives of model outputs with respect to model inputs (i.e., tangents in the multi-dimensional space) can be used to reduce the number of sample points required to construct an adequately approximate model.

An overview of various sensitivity metrics is presented in Section 2.4, and the metrics are listed in Table 2.3. Of these sensitivity metrics, the derivative information can be

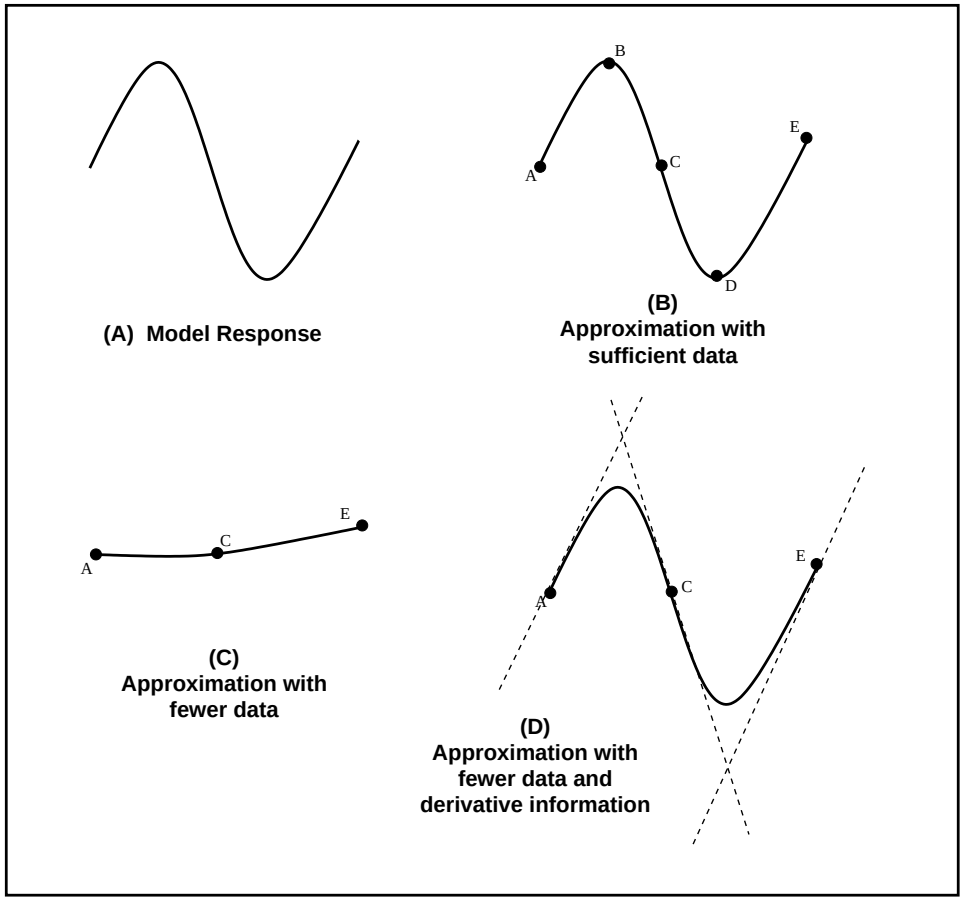


Figure 4.1: Rationale for coupling of the SRSM with a sensitivity analysis method

directly utilized in constructing an approximate model. For example, consider a model

$$z = f(x, y)$$

for which the output z is to be approximated as a function of inputs x and y through the expression:

$$z = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 y^2 + a_5 xy$$

In order to estimate the coefficients a_i's, the model outputs for z are needed at six sample points, requiring six model runs. This results in a set of six equations with the left hand side consisting of the model outputs and the right hand side consisting of linear equations in the coefficients.

For the same case, if all the first order partial derivatives are available at each sample point, the number of model runs can be reduced to two, and the derivative information

at two points, (x_1, y_1) and (x_2, y_2) can be used to construct the following closed system of equations:

$$\begin{aligned}
z|_{(x_1,y_1)} &= a_0 + a_1 x_1 + a_2 y_1 + a_3 {x_1}^2 + a_4 {y_1}^2 + a_5 x_1 y_1 \\
\left.\frac{\partial z}{\partial x}\right|_{(x_1,y_1)} &= a_1 + 2a_3 x_1 + a_5 y_1 \\
\left.\frac{\partial z}{\partial y}\right|_{(x_1,y_1)} &= a_2 + 2a_4 y_1 + a_5 x_1 \\
z|_{(x_2,y_2)} &= a_0 + a_1 x_2 + a_2 y_2 + a_3 {x_2}^2 + a_4 {y_2}^2 + a_5 x_2 y_2 \\
\left.\frac{\partial z}{\partial x}\right|_{(x_2,y_2)} &= a_1 + 2a_3 x_2 + a_5 y_2 \\
\left.\frac{\partial z}{\partial y}\right|_{(x_2,y_2)} &= a_2 + 2a_4 y_2 + a_5 x_2
\end{aligned}$$

Similarly, for a model with M parameters, the number of model runs required to estimate the unknown coefficients decreases by a factor of $M + 1$, since the first order partial derivatives with respect to M parameters result in M additional equations at each sample point, whereas calculation of the model output at a given point results in only one equation.

4.1 Methods for Estimation of Partial Derivatives

The following approaches are commonly used to estimate derivatives for a model [16, 29]:

(a) *Variational Equations:* When the model equations are available, one can, in principle, compute the derivatives from these equations by directly differentiating the equations and performing the necessary algebraic manipulations. This approach results in an additional set of equations that can be solved either in a coupled manner [200], or in a decoupled manner [60]. For lower dimensional, linear models, these equations can be solved directly along with the model equations to estimate the derivatives. However, for nonlinear functions the derivatives are generally more complicated than the function itself. Further, if the original model has n outputs and m model parameters, in order to estimate the partial derivatives of all the outputs with respect to all the parameters,

the number of equations required would be $m \times n$. Hence, this approach requires a considerable amount of effort for generating all the necessary model equations. Further, this approach requires access to the original model formulation equations.

(b) *"Brute-force" approach:* The two commonly used techniques in this approach are:

- *Central differences*: For the model represented by Equation 2.5, the partial derivative of u_i with respect to a parameter k_j at a point $(k_{1,0}, \ldots, k_{j,0}, \ldots, k_{m,0})$ is estimated by solving the model at two points $(k_{1,0}, \ldots, k_{j,0} - h, \ldots, k_{m,0})$ and $(k_{1,0}, \ldots, k_{j,0} + h, \ldots, k_{m,0})$, and the respective outputs are obtained as $u_{i,l}$ and $u_{i,r}$. The partial derivative is then given by

$$\frac{\partial u_i}{\partial k_j} \approx \frac{u_{i,r} - u_{i,l}}{2h} \; ; \; h \to 0$$

- *One-sided differences*: Here, the model is solved at a point $(k_{1,0}, \ldots, k_{j,0}, \ldots, k_{m,0})$, resulting in output $u_{i,0}$ and either at the point $(k_{1,0}, \ldots, k_{j,0} - h, \ldots, k_{m,0})$ or at the point $(k_{1,0}, \ldots, k_{j,0} + h, \ldots, k_{m,0})$. The left-side derivative approximation is given by

$$\frac{\partial u_i}{\partial k_j} \approx \frac{u_{i,0} - u_{i,l}}{h} \; ; \; h \to 0$$

and the right-side derivative approximation is given by

$$\frac{\partial u_i}{\partial k_j} \approx \frac{u_{i,r} - u_{i,0}}{h} \; ; \; h \to 0$$

In order to estimate the partial derivatives of outputs with respect to m parameters, the central difference method requires solution of the model equations $2m$ times (i.e., $2m$ "model runs"), that is two times for each parameter considered. On the other hand, the one-sided difference method requires $m + 1$ model runs, one for each parameter considered, and one run at the point under consideration.

The primary disadvantage of these methods is the computational cost associated with a large number of model runs. Another disadvantage with these methods involves the selection of a small enough value of h; a small h results in *"roundoff errors"*, resulting from subtracting two almost equal numbers, and a large value results in truncation

errors, as the omitted higher order terms become significant.

(c) *Symbolic Differentiation:* When the model equations are mathematically tractable, symbolic manipulators such as Maple [94] and MACSYMA [95] can be used for manipulating the model equations and obtaining expressions for the partial derivatives. The practical limitations are numerous: (a) many computational models do not have mathematically tractable equations, (b) the symbolic manipulators are in general, unable to deal with constructs such as branches, loops or subroutines that are inherent in computer codes, and (c) the derivative code expressions could be very complex, because, for every operation, the derivative expression in essence doubles, leading to a combinatorial explosion [16].

(d) *Automated Differentiation:* Automated differentiation, also known as automatic differentiation, relies on the fact that every function is executed on a computer as a sequence of elementary operations such as additions, multiplications, and elementary functions. By applying the chain rule

$$\frac{\partial}{\partial t} f(g(t))\bigg|_{t=t_0} = \left(\frac{\partial}{\partial s} f(s)\bigg|_{s=g(t_0)} \right) \left(\frac{\partial}{\partial t} g(t)\bigg|_{t=t_0} \right)$$

over and over again to the composition of those elementary operations, one can compute, in a completely mechanical fashion, derivatives of f that are correct up to the machine precision [16]. This approach essentially consists of using a computational tool that can "understand" the model code and subsequently produce the code for estimating partial derivatives. The main advantage of this method is that the partial derivatives can be estimated using a single model run; another advantage is that this method can be used in conjunction with complex numerical models, that can be considered as "black-box" structures.

The automatic differentiation method has been widely applied in sensitivity analysis of transport/transformation modeling, as mentioned in Section 2.5.4. ADIFOR (Automatic DIfferentiation of FORtran) is used in this work for coupling with the SRSM. The following sections describe the ADIFOR methodology and the coupling with the SRSM.

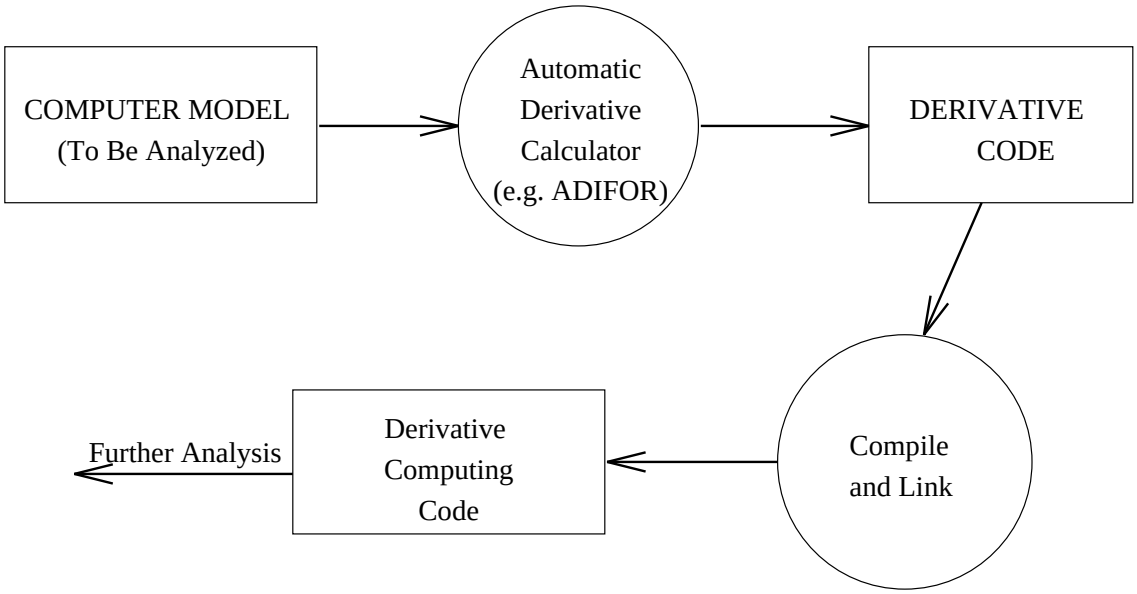


Figure 4.2: Schematic illustration of the Automatic Differentiation method (adapted from Bischof et al., 1994 [16])

4.2 Automatic Differentiation using ADIFOR

ADIFOR (Automatic DIfferentiation of FORtran) [17] transforms the model source code based on the specification of the dependent and independent variables, and produces source code that calculates the derivatives of the dependent variables with respect to the independent variables. Figure 4.2 presents a schematic depiction of the automatic differentiation procedure using ADIFOR.

Figure 4.3 presents an example of the transformation of source code using ADIFOR. Figure 4.3 (a) shows a segment of a sample Fortran program for which derivatives are desired. The dependent variable is `y` and the independent variables are `x(1)` through `x(10)`. Figure 4.3 (b) shows the derivative code produced through the application of ADIFOR. Using the resultant code, the partial derivatives of `y` with respect to independent variable `x(i)` can be obtained through the corresponding element `dy(i)` from the array `dy`.

This example illustrates some key advantages of using ADIFOR over other methods: ADIFOR steps through the conditional execution loops and iterative loops (e.g., the `if` and the `do` statements in Figure 4.3 (a)), and ADIFOR can calculate derivatives for

assignment and iteration statements (e.g., the statement `y =x(i) * y *y`). Further, the iterative loops and conditional loops are common in computer codes, while they do not appear in algebraic or differential equations. It must be noted that the value of the variable `y` depends to a great extent the "sign" of the various `x(i)`'s, specifically of `x(n)`; such situations are difficult to express in algebraic forms, whereas they can be readily expressed in computer code. Since ADIFOR methodically follows the model code, all the complex logic of the model code is followed exactly by the derivative code.

In principle, this method can be easily extended to compute derivatives for computer code of arbitrary length and complexity. Furthermore, in the application of ADIFOR, the user has to only (a) specify the information about the dependent and independent variables, and (b) add statements to the derivative code to output the derivatives.

4.3 Coupling of SRSM and ADIFOR

The coupling of SRSM and ADIFOR follows the same steps as the SRSM with respect to input and output transformations. The coupled method, SRSM-ADIFOR, SRSM approximates uncertain model outputs in terms of a set of "standard random variables" (*srv*s), denoted by the set $\{\xi_i\}_{i=1}^{n}$. The following steps are involved in the application of the SRSM-ADIFOR:

- the model inputs are expressed as functions of selected *srv*s, as described in Section 3.2,
- an approximation for the model outputs is assumed in the same form as presented in Section 3.3:

$$\begin{aligned} y_r &= a_{r,0} + \sum_{i_1=1}^{n} a_{r,i_1}\Gamma_1(\xi_{i_1}) + \sum_{i_1=1}^{n}\sum_{i_2=1}^{i_1} a_{r,i_1 i_2}\Gamma_2(\xi_{i_1},\xi_{i_2}) \\ &+ \sum_{i_1=1}^{n}\sum_{i_2=1}^{i_1}\sum_{i_3=1}^{i_2} a_{r,i_1 i_2 i_3}\Gamma_3(\xi_{i_1},\xi_{i_2},\xi_{i_3}) + \ldots \end{aligned} \tag{4.1}$$

  where, y_r is the rth output metric (or random output) of the model, the $a_{r,i_1\ldots}$'s are deterministic constants to be estimated, and the $\Gamma_p(\xi_{i_1},\ldots,\xi_{i_p})$ are multi-dimensional Hermite polynomials of degree p, given by

$$\Gamma_p(\xi_{i_1},\ldots,\xi_{i_p}) = (-1)^p\, e^{\frac{1}{2}\boldsymbol{\xi}^T\boldsymbol{\xi}}\, \frac{\partial^p}{\partial\xi_{i_1}\ldots\partial\xi_{i_p}} e^{-\frac{1}{2}\boldsymbol{\xi}^T\boldsymbol{\xi}}, \tag{4.2}$$

```
dimension x(10)
y = 1.0
do i = 1,n
 if (x(i) .gt. 0.0) then
  y = x(i) * y * y
 endif
enddo
```

(a)

```
dimension x(10)
dimension dy(10), dx(10,10)
do j=1,10
 dx(i,i) = 1.0
enddo
 do j = 1, 10
 dy(j) = 0.0
enddo
y = 1.0

do i = 1, n
 if (x(i) .gt. 0.0) then
  a = x(i)*y
  b = y*y
  c = a + y*x(i)
  do j = 1, 10
   dy(j) = b*dx(j,i) + c*dy(j)
  enddo
  y = a*y
 endif
enddo
```

(b)

Figure 4.3: FORTRAN derivative code generation using ADIFOR: (a) Sample code fragment and (b) ADIFOR generated derivative code

- correspondingly, the first order partial derivatives of the rth model output with respect to jth $srv\xi_j$, given by $\frac{\partial y_r}{\partial \xi_j}$ are expressed as

$$\begin{aligned}\frac{\partial y_r}{\partial \xi_j} &= \sum_{i_1=1}^{n} a_{r,i_1} \frac{\partial \Gamma_1(\xi_{i_1})}{\partial \xi_j} + \sum_{i_1=1}^{n} \sum_{i_2=1}^{i_1} a_{r,i_1 i_2} \frac{\partial \Gamma_2(\xi_{i_1}, \xi_{i_2})}{\partial \xi_j} \\ &+ \sum_{i_1=1}^{n} \sum_{i_2=1}^{i_1} \sum_{i_3=1}^{i_2} a_{r,i_1 i_2 i_3} \frac{\partial \Gamma_3(\xi_{i_1}, \xi_{i_2}, \xi_{i_3})}{\partial \xi_j} + \ldots \end{aligned} \tag{4.3}$$

- the model derivative code is generated using the original model code and ADIFOR and modified so that it can calculate the first order partial derivatives of model outputs with respect to the *srv*s,

- a set of sample points is selected using the sampling algorithm presented in Section 3.4.2; the number of sample points selected is smaller than the number of coefficients to be estimated, as shown in Equation 4.4,
- the model outputs and the partial derivatives with respect to the *srv*s are calculated at these sample points,
- outputs at these points are equated with the polynomial chaos approximation (Equation 4.1, resulting in a set of linear equations in a_i's, and partial derivatives at these points are equated with the corresponding approximation given by Equation 4.3,
- The resultant equations are solved using singular value decomposition, to obtain estimates of a_i's, as illustrated in Section 4.4.

4.3.1 Selection of Sample Points

In the application of SRSM-ADIFOR to the uncertainty analysis of a model with M inputs, P outputs, for a given order of expansion, Equation 4.1 is used to calculate the number of coefficients to be estimated (say K). Thus, K coefficients need to be estimated for each output. The execution of the model derivative code at one sample point gives the model calculations for the outputs and M first order partial derivatives for each output. Thus, equations 4.1 and 4.3 in conjunction with these calculations result in $M+1$ linear equations at each sample point.

Here, the number of recommended sample points is based on the rationale behind the regression based SRSM: the number of resulting equations should be higher than the number of coefficients estimated in order to obtain robust estimates of the coefficients. Here, the recommended number of equations is about twice the number of coefficients to be estimated.

Since for each sample point, the number of resultant equations is $M+1$, the number of sample points required for SRSM-ADIFOR, N, is approximately given by*:

$$N \approx \frac{2K}{M+1} \tag{4.4}$$

*the number is approximate because $2K$ may not always be exactly divisible by $M+1$

4.4 An Illustration of the Application of the SRSM-ADIFOR

This section presents an illustration of the SRSM-ADIFOR method by revisiting the example presented in Section 3.7 involving the application of the SRSM. The probability distributions of the three model inputs, X_1, X_2, and X_3, are given by Equation 3.12, and the two model outputs, Y_1 and Y_2, are approximated by polynomial chaos expansions

$$\begin{aligned} Y_1 &= a_0 + a_1\xi_1 + a_2\xi_2 + a_3\xi_3 + a_4(\xi_1^2 - 1) + a_5(\xi_2^2 - 1) + a_6(\xi_3^2 - 1) \\ &\quad + a_7\xi_1\xi_2 + a_8\xi_2\xi_3 + a_9\xi_1\xi_3 \\ Y_2 &= b_0 + b_1\xi_1 + b_2\xi_2 + b_3\xi_3 + b_4(\xi_1^2 - 1) + b_5(\xi_2^2 - 1) + b_6(\xi_3^2 - 1) \\ &\quad + b_7\xi_1\xi_2 + b_8\xi_2\xi_3 + b_9\xi_1\xi_3 \end{aligned} \tag{4.5}$$

Clearly, the partial derivatives of Y_1 and Y_2 with respect to the *srv*s are given by:

$$\begin{aligned} \frac{\partial Y_1}{\partial \xi_1} &= a_1 + 2a_4\xi_1 + a_7\xi_2 + a_9\xi_3 \\ \frac{\partial Y_1}{\partial \xi_2} &= a_2 + 2a_5\xi_2 + a_7\xi_1 + a_8\xi_3 \\ \frac{\partial Y_1}{\partial \xi_3} &= a_3 + 2a_6\xi_3 + a_8\xi_2 + a_9\xi_1 \\ \frac{\partial Y_2}{\partial \xi_1} &= b_1 + 2b_4\xi_1 + b_7\xi_2 + b_9\xi_3 \\ \frac{\partial Y_2}{\partial \xi_2} &= b_2 + 2b_5\xi_2 + b_7\xi_1 + b_8\xi_3 \\ \frac{\partial Y_2}{\partial \xi_3} &= b_3 + 2b_6\xi_3 + b_8\xi_2 + b_9\xi_1 \end{aligned} \tag{4.6}$$

Based on Equation 4.4, the number of sample points, N, required to estimate the 10 unknown coefficients equals 5. Thus, using SRSM-ADIFOR, the uncertainty in the model outputs can be estimated by obtaining the model outputs at $N = 5$ sample points, compared to 20 sample points required in the application of the SRSM.

The model equations corresponding to the estimation of unknown coefficients, are given by the following equations:

$$\boldsymbol{Z}^T \times \begin{bmatrix} a_0 & b_0 \\ a_1 & b_1 \\ a_2 & b_2 \\ \vdots & \vdots \\ a_8 & b_8 \\ a_9 & b_9 \end{bmatrix} = \begin{bmatrix} y_{1,1} & y_{2,1} \\ \dfrac{\partial y_1}{\partial \xi_{1,1}} & \dfrac{\partial y_2}{\partial \xi_{1,1}} \\ \dfrac{\partial y_1}{\partial \xi_{2,1}} & \dfrac{\partial y_2}{\partial \xi_{2,1}} \\ \dfrac{\partial y_1}{\partial \xi_{3,1}} & \dfrac{\partial y_2}{\partial \xi_{3,1}} \\ \vdots & \vdots \\ \vdots & \vdots \\ y_{1,N} & y_{2,N} \\ \dfrac{\partial y_1}{\partial \xi_{1,N}} & \dfrac{\partial y_2}{\partial \xi_{1,N}} \\ \dfrac{\partial y_1}{\partial \xi_{2,N}} & \dfrac{\partial y_2}{\partial \xi_{2,N}} \\ \dfrac{\partial y_1}{\partial \xi_{3,N}} & \dfrac{\partial y_2}{\partial \xi_{3,N}} \end{bmatrix}, \quad \text{where} \tag{4.7}$$

$$\boldsymbol{Z} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & \dots & 1 & 0 & 0 & 0 \\ \xi_{1,1} & 1 & 0 & 0 & \dots & \dots & \xi_{1,N} & 1 & 0 & 0 \\ \xi_{2,1} & 0 & 1 & 0 & \dots & \dots & \xi_{2,N} & 0 & 1 & 0 \\ \xi_{3,1} & 0 & 0 & 1 & \dots & \dots & \xi_{3,N} & 0 & 0 & 1 \\ \xi_{1,1}^2 - 1 & 2\xi_{1,1} & 0 & 0 & \dots & \dots & \xi_{1,N}^2 - 1 & 2\xi_{1,N} & 0 & 0 \\ \xi_{2,1}^2 - 1 & 0 & 2\xi_{2,1} & 1 & \dots & \dots & \xi_{2,N}^2 - 1 & 0 & 2\xi_{2,N} & 1 \\ \xi_{3,1}^2 - 1 & 0 & 0 & 2\xi_{3,1} & \dots & \dots & \xi_{3,N}^2 - 1 & 0 & 0 & 2\xi_{3,N} \\ \xi_{1,1}\xi_{2,1} & \xi_{2,1} & 0 & \xi_{3,1} & \dots & \dots & \xi_{1,N}\xi_{2,N} & \xi_{2,N} & 0 & \xi_{3,N} \\ \xi_{2,1}\xi_{3,1} & \xi_{1,1} & \xi_{3,1} & 0 & \dots & \dots & \xi_{2,N}\xi_{3,N} & \xi_{1,N} & \xi_{3,N} & 0 \\ \xi_{1,1}\xi_{3,1} & 0 & \xi_{2,1} & \xi_{1,1} & \dots & \dots & \xi_{1,N}\xi_{3,N} & 0 & \xi_{2,N} & \xi_{1,N} \end{bmatrix} \tag{4.8}$$

4.5 Implementation of the SRSM-ADIFOR

The implementation of the SRSM-ADIFOR involves the modification of the model source code, by first processing the code using ADIFOR, and then making changes in

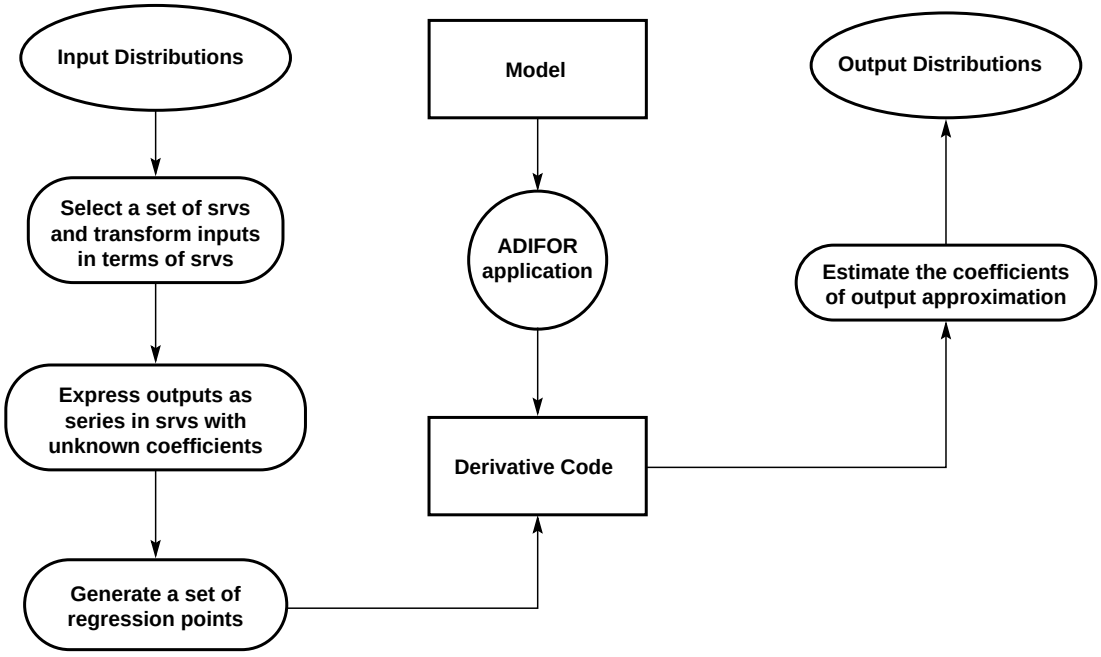


Figure 4.4: Schematic depiction of the steps involved in the application of the SRSM-ADIFOR

the resultant code so that the derivatives with respect to the *srv*s can be calculated. Subsequently the derivative code has to be coupled with a main routine that produces the derivatives of concern as program outputs. The calculated derivatives are then coupled with the SRSM to estimate the coefficients of the polynomial chaos approximations of the outputs. In order to accomplish this, a set of routines is developed to generate expressions for the derivatives of a polynomial chaos expansion, and to generate a matrix similar to the illustration shown in Equation 4.7. The subsequent procedure is similar to the application of the SRSM. Figure 4.4 presents the steps involved in the application of the SRSM-ADIFOR.

Chapter 5

CASE STUDIES FOR THE EVALUATION OF THE SRSM AND THE SRSM-ADIFOR

This chapter presents four case studies to evaluate the SRSM and the SRSM-ADIFOR. These models cover a range of applications, both from the perspective of model application (such as biological, air quality, and groundwater) and from the perspective of model complexity (ranging from a zero dimensional model to a three-dimensional urban/regional scale photochemical model). The execution times of the models also vary from a fraction of a second to a few hours per model run. The main objective of these case studies is to evaluate the applicability of the SRSM and the SRSM-ADIFOR to a wide range of transport-transformation models, and potentially to numerical models used in other areas.

The case studies are presented here in the following order:

I: A zero-dimensional pharmacokinetic model.

II: A two-dimensional atmospheric plume model.

III: A three-dimensional urban/regional scale air quality model.

IV: A one-dimensional ground water model with discontinuous probability density functions.

The order corresponds to both the chronological evaluation of the SRSM method, as well as to the associated complexity. The computational time for each model simulation in the first three cases is of the order of a fraction of a second (case I), through a few seconds (case II), to approximately 6 hours (case III). The groundwater model required about a few seconds per model simulation, but is presented at the end, since the complexity with respect to the specification of uncertainties in inputs is high.

Each case study description contains the information about the model and its application in exposure/risk assessment. Then, the uncertainties associated with the model are described, and the application of different uncertainty propagation methods is presented. The estimates from the uncertainty propagation methods are then presented graphically, in the form of probability density functions (*pdf*s), along with a brief discussion at the end.

5.1 Case Study I: A Zero Dimensional Physiological System

Uncertainty Analysis of a Physiologically Based PharmacoKinetic Model for Perchloroethylene

5.1.1 Description of the Case Study

There is often significant uncertainty associated with human Physiologically Based PharmacoKinetic (PBPK) model parameters since human *in vivo* experimental data are not usually available for toxic chemicals. Thus many of the parameters in human PBPK models are generally estimated by *in vitro* experimentation and by inter-species scale-up of animal PBPK model parameters. The uncertainty in human PBPK parameters includes a significant amount of natural variability, reflecting the interindividual variability inherent in human populations. It is desirable to estimate the effects that uncertainties (including variability) associated with model inputs and parameters have on output metrics such as internal and biologically effective doses.

Here, a PBPK model for perchloroethylene (PERC) [61] is considered for uncertainty analysis. Human exposure to PERC, and the subsequent metabolism of PERC in the body, is of concern since PERC is a potential carcinogen [61]. Further, PERC is used as a cleaning solvent in the dry cleaning industry, and hence human exposures to it are very common. Three dose surrogates for PERC inhalation are considered in this study: (a) area under the arterial blood concentration (AUCA), (b) area under venous blood from liver (AUCL), and (c) the cumulative amount metabolized in the liver (CML).

CML is considered to be an alternative indicator of risk than simply the total uptake of PERC, since the health risks are thought to be associated with the metabolism of

PERC in the body. AUCA and AUCL are indicators of the amount of time PERC resides in the arterial blood and liver, respectively, and thus are also considered as alternative indicators of risk.

Appendix B presents the formulation of a PBPK model, and additional details can be found in the literature [78, 176]. Figure 5.1 presents the schematic of the PERC PBPK model used in this case study; the human body is considered to consist of a set of compartments corresponding to different organs and tissues. These are modeled as a series of interconnected continuous stirred tank reactors (CSTRs).

The effect of uncertainty in the parameters of the PERC PBPK on the three dose surrogates is analyzed using the following:

- Standard Monte Carlo approach,
- Latin Hypercube Sampling (LHS),
- Deterministic Equivalent Modeling Method/Probabilistic Collocation Method (DEMM/PCM),
- Stochastic Response Surface Method with Efficient Collocation (ECM),
- Regression based SRSM, also referred to simply as SRSM, and
- Combined Application of the Stochastic Response Surface Method and Automatic Differentiation Method (SRSM-ADIFOR).

The uncertainty analysis consisted of two stages: In the first stage, the effect of uncertainty in the metabolic constants (K_m and $V_{\max}$) and the partition coefficients are considered (see Table 5.1). The uncertainty in these six parameters are described by independent log-normally distributed random variables. In the second stage, the uncertainty in compartmental volume proportions are included in the analysis. These five additional parameters are described by a set of mutually dependent Dirichlet distributed random variables. In the first stage (6 parameters), the DEMM approach and the ECM are evaluated through comparison with the results from a standard Monte Carlo simulation. For the second stage (11 parameters), the ECM, Regression Based

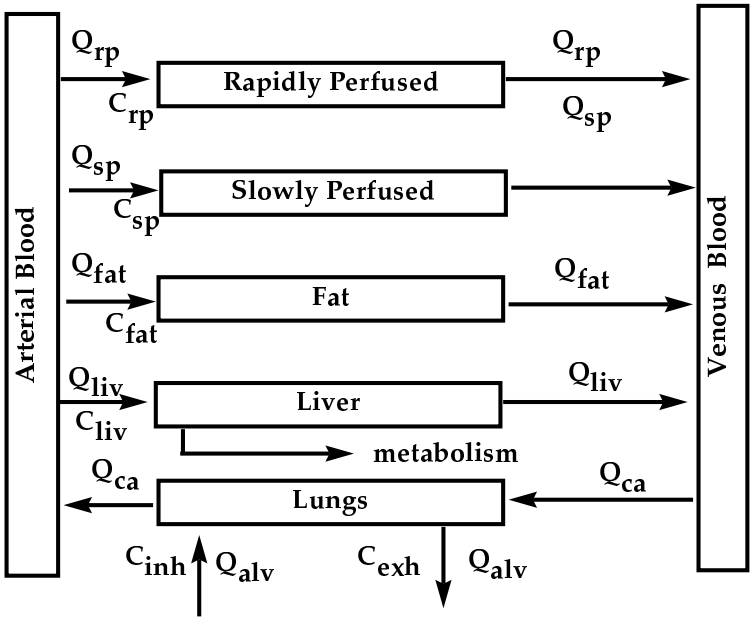


Figure 5.1: Schematic representation of a PBPK model for PERC

SRSM and SRSM-ADIFOR are evaluated by comparing the results from these methods with the results from the standard Monte Carlo and LHS methods.

In an earlier study, Farrar et al. [67] examined the effect of PBPK model parameter uncertainty on the three dose surrogates, CML, AUCA, and AUCL. The PBPK model parameter uncertainties used in the present work are adopted from their work, and are specified by the preferred values (PVs) and uncertainty factors (UFs) given in Table 5.1. The preferred values for all parameters other than metabolic constants, correspond to the nominal values in the PERC PBPK model [61], updated as suggested by Reitz and Nolan [170]. The preferred values of metabolic constants are given by the geometric mean of values suggested by Hattis et al [93], and by Reitz and Nolan [170]. The uncertainty factor for a parameter is defined such that the interval from X_P/UF to X_P UF is the 95% confidence interval, where X_P is the preferred value of parameter X.

5.1.2 Specification of Parameter Uncertainty

Partition coefficients and metabolic constants are assumed to be lognormally distributed, according to Farrar et al [67]. The *pdf* for a given lognormally distributed parameter is specified by assuming that the preferred value of the parameter corresponds

Table 5.1: Deterministic and uncertain parameters used in the uncertainty analysis of the PERC PBPK model

Symbol	Description	Preferred Value	UF
BW	body weight [Kg]	70	—
	Partition Coefficients		
$P_{\text{blood/air}}$	blood:air partition coefficient	12	1.7
$P_{\text{fat/blood}}$	fat:blood partition coefficient	102	2.15
$P_{\text{sp/blood}}$	slowly perfused tissue:blood partition coefficient	2.66	11.0
$P_{\text{rp/blood}}$	rapidly perfused tissue:blood partition coefficient	5.05	5.69
$P_{\text{liv/blood}}$	liver:blood partition coefficient	5.05	9.37
	Blood Flows		
Q_c	cardiac output [liters/hr]	348	1.12
Q_p	alveolar ventilation [liters/hr]	288	1.50
Q_{fat}	blood flow to fat [liters/hr]	17.4	1.09
Q_{sp}	blood flow to slowly perfused tissue [liters/hr]	87.0	1.04
Q_{rp}	blood flow to rapidly perfused tissue [liters/hr]	153	1.25
Q_{liv}	blood flow to liver [liters/hr]	90.6	1.35
	Compartment Mass Fractions		
V_{fat}	mass of fat compartment [Kg]	23.0%	1.09
V_{sp}	mass of slowly perfused tissue compartment [Kg]	62.0%	1.04
V_{rp}	mass of rapidly perfused tissue compartment [Kg]	5.0%	1.25
V_{liv}	mass of liver compartment [Kg]	2.6%	1.35
	Metabolic Constants		
K_m	Michaelis-Menten constant for metabolism [mg/liter]	1.47	12.3
$V_{\text{max},c}$	maximum rate of metabolism [mg/hr]	0.33	2.84

to the median of the distribution. This is convenient, since the mean and standard deviation that specify the corresponding normal distribution are then given by $\ln X_P$ and $\ln(\mathrm{UF})/1.96$ respectively.

Compartment volume fractions are assumed to be specified by a Dirichlet distribution [67]. This implies that each compartment volume is specified by a χ^2 distribution [114, 201], the *pdf* of which is given by:

$$f(x;a) = \begin{cases} \dfrac{1}{\Gamma(a/2)2^{a/2}} x^{\frac{a}{2}-1} e^{-x/2}, & x > 0 \\ 0, \text{ elsewhere} \end{cases} \tag{5.1}$$

The mean of the χ^2 distributions is a, and is related to the preferred value and the Dirichlet distribution parameter Θ, as follows:

$$\mathrm{E}[X_i] = x_{ip}\Theta$$

where X_i is the ith compartment volume, $\mathrm{E}[X_i]$ is its mean, and x_{ip} is the preferred value of the fraction X_i, as given in Table 5.1. From the values of Θ and the preferred values for each compartment volume fraction, the parameters of the *pdf*s of the compartment volumes can be obtained.

Compartment blood flows are assumed to be specified as deterministic functions of compartment volumes and preferred values of compartmental blood flows. In sleeping humans, the blood flow to compartment j is given by:

$$Q_j = Q_{j,P}\frac{V_j}{V_{j,P}} ,$$

where the notation X_P for the preferred value of the parameter X is employed. The total cardiac output is given by

$$Q_c = Q_{\mathrm{liv}} + Q_{\mathrm{fat}} + Q_{\mathrm{sp}} + Q_{\mathrm{rp}} ,$$

and the total cardiac output in waking humans is given by:

$$Q_{cw} = Q_c + (Q_{cw,P} - Q_{c,P})q_{c\epsilon} ,$$

where $q_{c\epsilon}$ is a lognormally distributed random variable, the preferred value and uncertainty factor for which are given in Table 5.1. The form of the expression for Q_{cw}

ensures that $Q_{cw} > Q_c$ always. The *pdf* for $q_{c\epsilon}$ is constructed in a manner analogous to that of the other lognormally distributed parameters as described above.

The increase in cardiac output in waking humans, is accounted for by an increase in blood flow to the slowly perfused compartment, as follows:

$$Q_{spw} = Q_{sp,P} + (Q_{cw,P} - Q_{c,P})q_{c\epsilon}.$$

Alveolar ventilation rate is a deterministic function of the cardiac output, and the preferred values for ventilation are given in Table 5.1. Alveolar ventilation rates in sleeping and waking humans are given by:

$$Q_p = Q_{p,P}\,\frac{Q_c}{Q_{c,P}}\ , \text{ and } \quad Q_{pw} = Q_{pw,P}\,\frac{Q_c w}{Q_{cw,P}}$$

5.1.3 Implementation of the PERC PBPK Model

The PERC PBPK model has been implemented in SimuSolv [58], a software environment for numerical modeling and simulation. The PBPK model consisted of a set of coupled ordinary differential equations, which describe the dynamic variation in the amount of PERC in each of the compartments shown in Figure 5.1, resulting from uptake, distribution, metabolism and excretion processes. The random parameters in the model have been sampled for the simulations using the pseudo-gaussian random number generator function in SimuSolv, `RGAUSS`. The representation of *pdf*s of uncertain parameters in terms of standard gaussian random variables, each with zero mean and unit variance, is described here. Daily values of dose surrogates, CML, AUCA, and AUCL, are computed using the PBPK model to simulate a 24 hour inhalation exposure to 10 ppb of PERC, which is a typical high concentration in urban/industrial areas [102].

A lognormally distributed random variable X, with median μ_X, and an uncertainty factor UF_X, can be represented in terms of a standard normal random variable $\xi = \mathrm{N}(0,1)$, having zero mean and unit variance, as follows [201]:

$$X = \exp\left(\ln(\mu_X) + \frac{\ln(\mathrm{UF}_X)}{1.96}\xi\right) \tag{5.2}$$

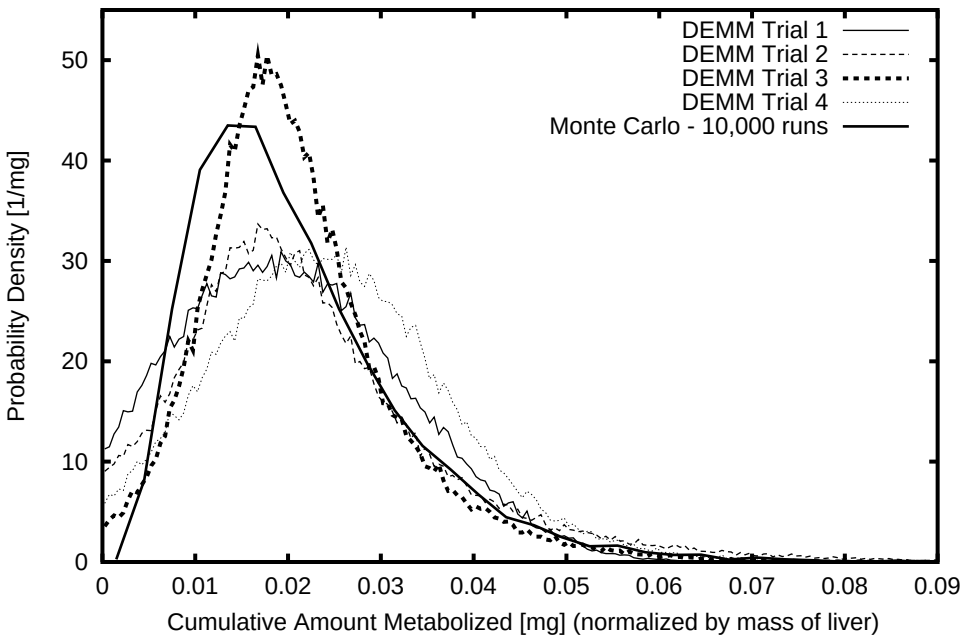


Figure 5.2: Evaluation of DEMM/PCM: uncertainty in the cumulative amount of PERC metabolized, over a period of one day, resulting from 6 uncertain parameters

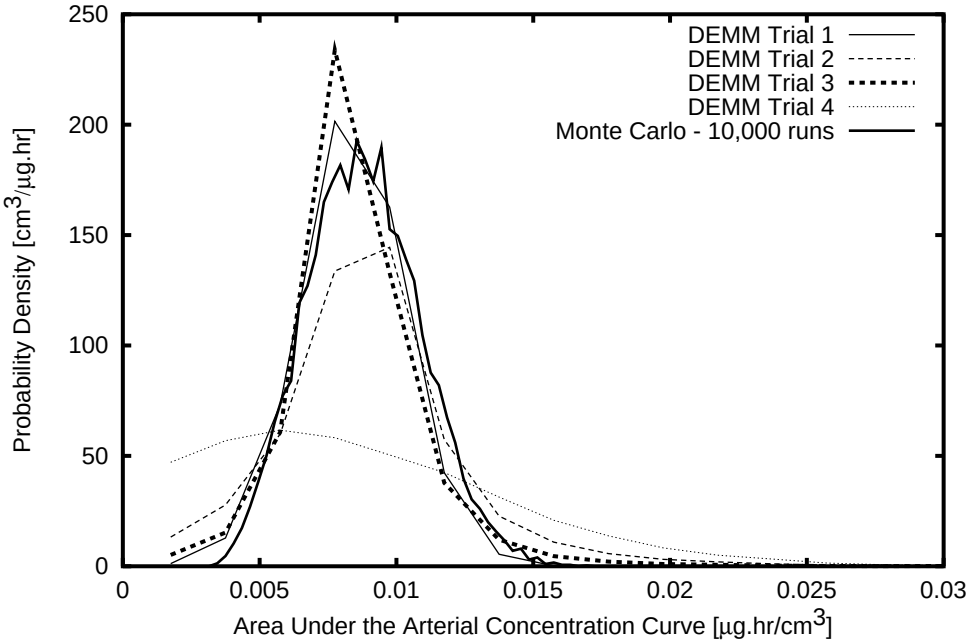
A χ^2 distributed random variable with parameter a, can be approximated in terms of a normal random variable ξ as follows [217]:

$$Y = a\left(\xi\sqrt{\frac{2}{9a}} + 1 - \frac{2}{9a}\right)^3 \tag{5.3}$$

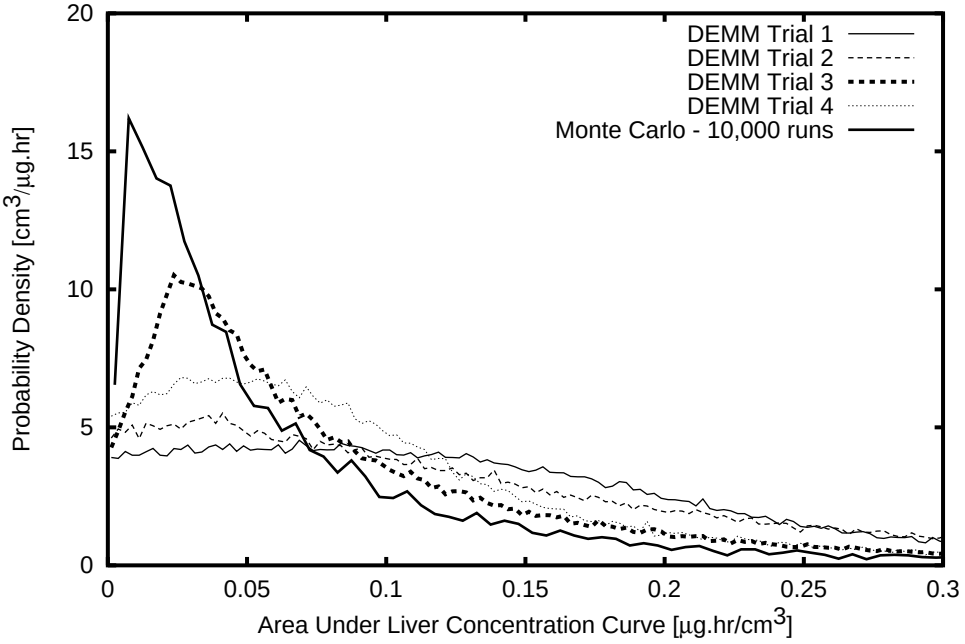
5.1.4 Results for the PBPK case study

Stage I (Six Uncertain Parameters): Evaluation of DEMM/PCM and ECM

The effect of six uncertain PBPK model parameters on the dose surrogate *pdf*s is studied in the first stage of uncertainty analysis. The number of model simulations used for DEMM/PCM and the ECM are equal, as both these methods differ only with respect to the selection of collocation points. Second and third order approximations are used for both DEMM/PCM and the ECM. The number of simulations required for both these methods for the case of six uncertain inputs are 28 and 84, for 2nd and 3rd order approximations, respectively. Additionally, 10,000 Monte Carlo simulations are used in the first stage for evaluation of the DEMM/PCM and ECM.



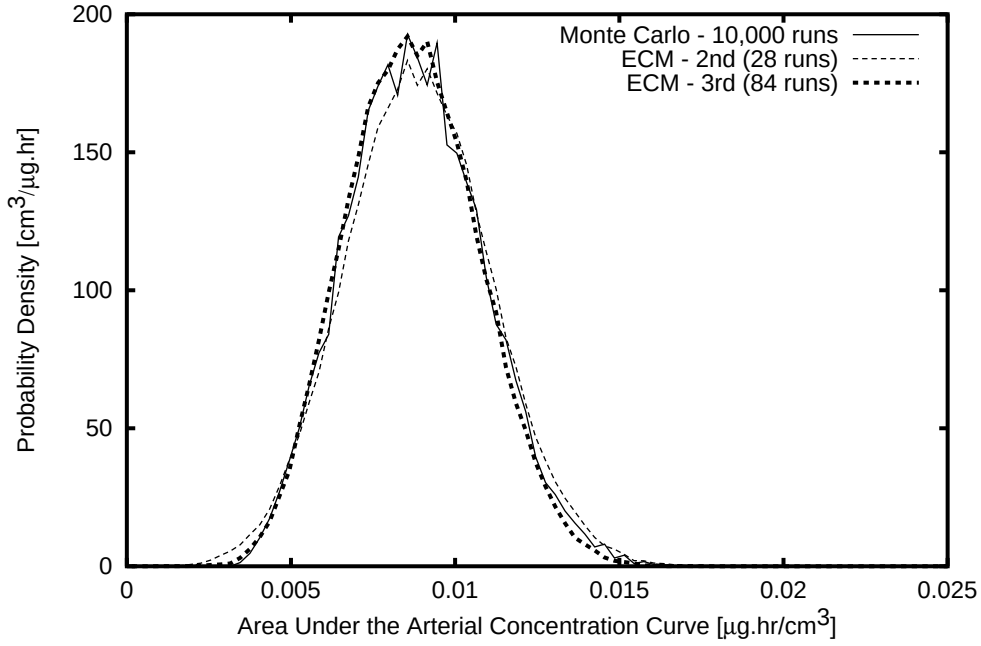
(a)



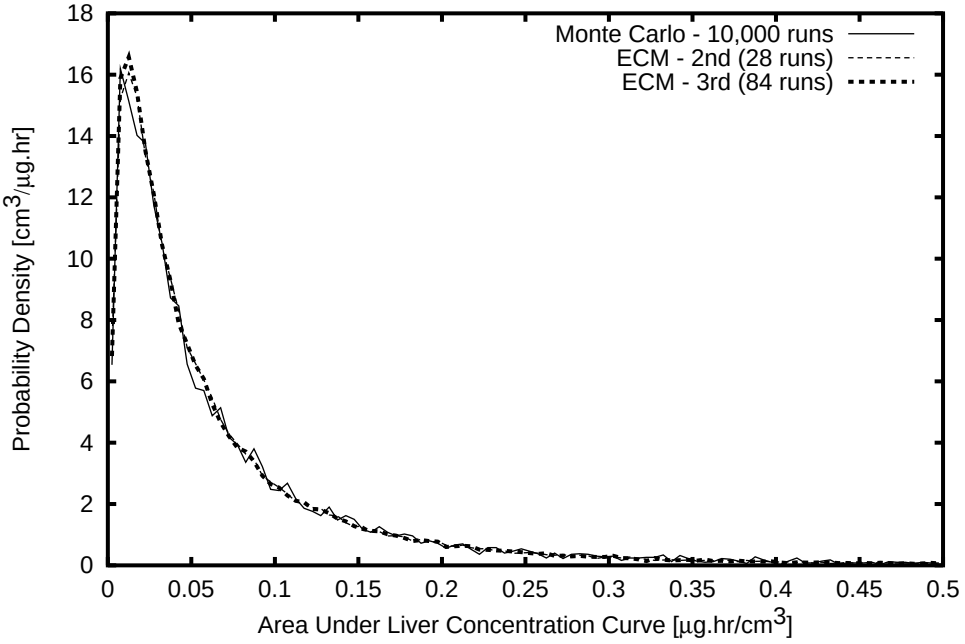
(b)

Figure 5.3: Evaluation of DEMM/PCM: uncertainty in the area under the PERC concentration-time curve in (a) arterial blood (AUCA) and in (b) liver (AUCL), over a period of one day, resulting from 6 uncertain parameters

Figures 5.2 and 5.3 present the *pdf*s estimated from four different sets of collocation points selected from the same set of original 4096 points (DEMM Trial 1, 2, 3 and 4), for the dose surrogates AUCA, AUCL and CML, respectively. They also present the *pdf*s estimated by standard Monte Carlo approach. From these figures, it can be seen that the DEMM/PCM approach cannot consistently guarantee convergence of the approximation to the true probability density. Since the DEMM/PCM method failed to produce consistent results for a simple model with independent input probability distributions, this method was not used in other case studies, which involved more complex models or input distributions.



(a)



(b)

Figure 5.4: Evaluation of ECM: uncertainty in the area under the PERC concentration-time curve in (a) arterial blood (AUCA) and in (b) liver (AUCL), over a period of one day, resulting from 6 uncertain parameters

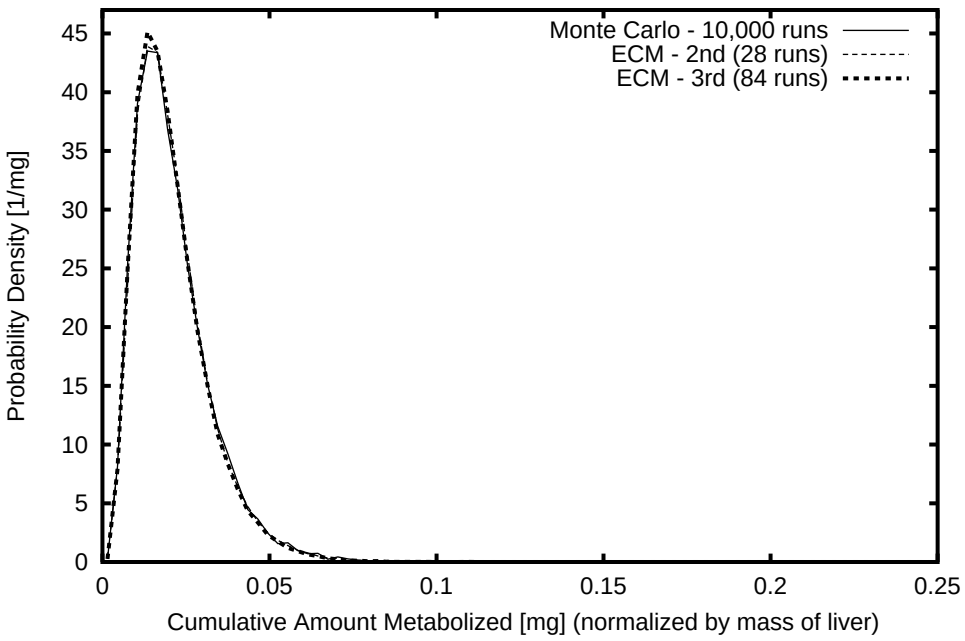


Figure 5.5: Evaluation of ECM: uncertainty in the cumulative amount of PERC metabolized, over a period of one day, resulting from 6 uncertain parameters

A comparison of the *pdf*s of the AUCL, AUCL, and CML dose surrogates, obtained using 2nd and 3rd order ECM and Monte Carlo method, for the six uncertain parameter case is shown in Figures 5.4 and 5.5. The *pdf*s estimated by both 2nd and 3rd order DEM/ECM appear to agree well with each other and with the *pdf*s estimated using Monte Carlo simulation. Since the differences between the 2nd order and 3rd order approximations do not appear to be significant, the approximation is judged to have converged, and uncertainty analysis has not been performed with higher order approximations.

Although the *pdf*s for all six input parameters considered in the first stage of the uncertainty analysis are log-normally distributed, the shapes of the dose surrogate *pdf*s vary considerably, from the near symmetrical *pdf* for AUCA to the highly skewed *pdf*s for AUCL and CML. In all the cases, however, the ECM method produced results very close to those of the Monte Carlo method, but required substantially fewer model runs.

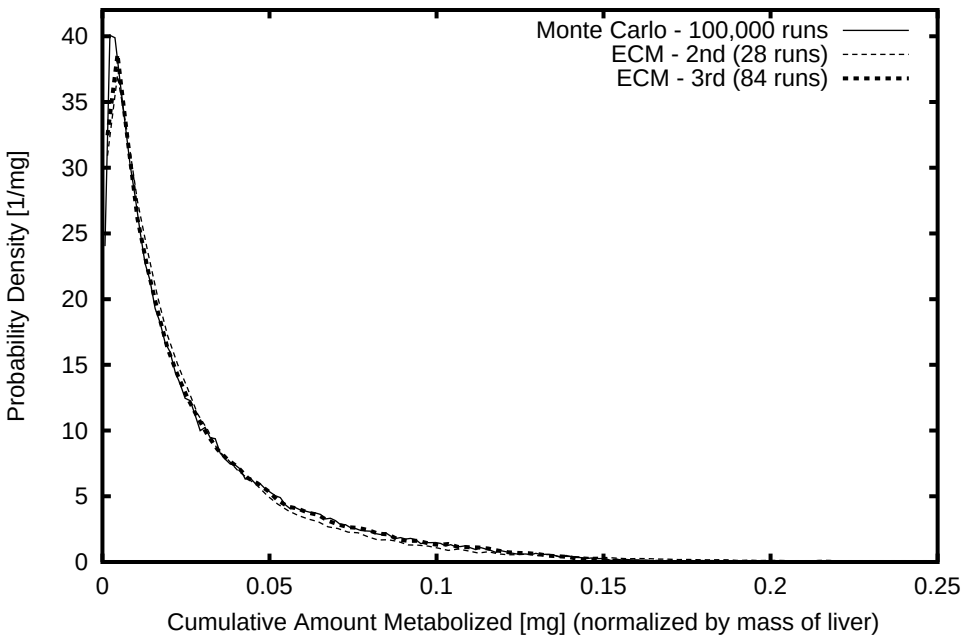
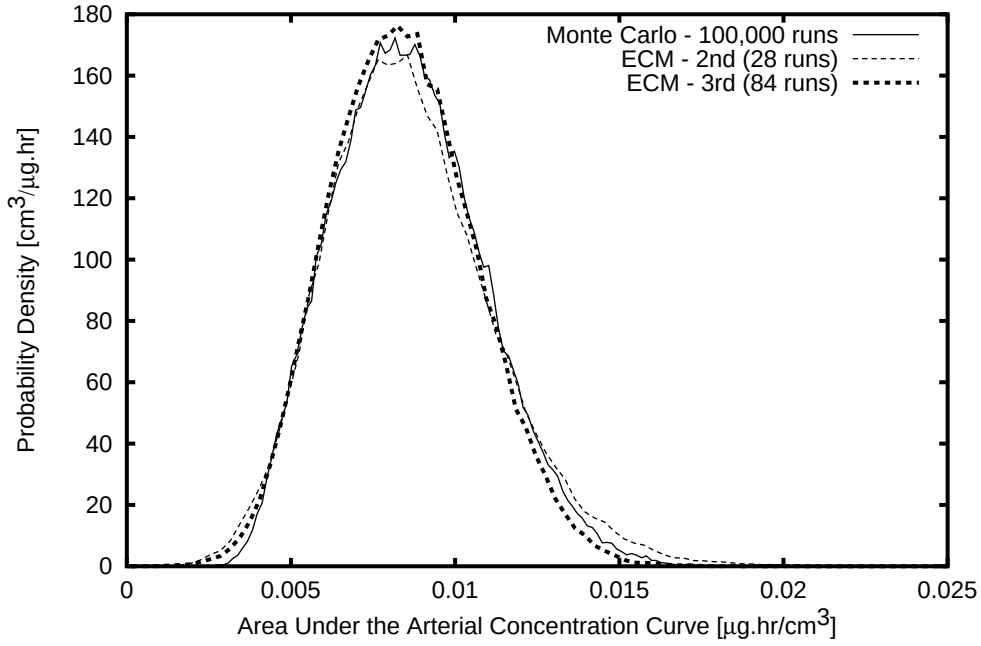


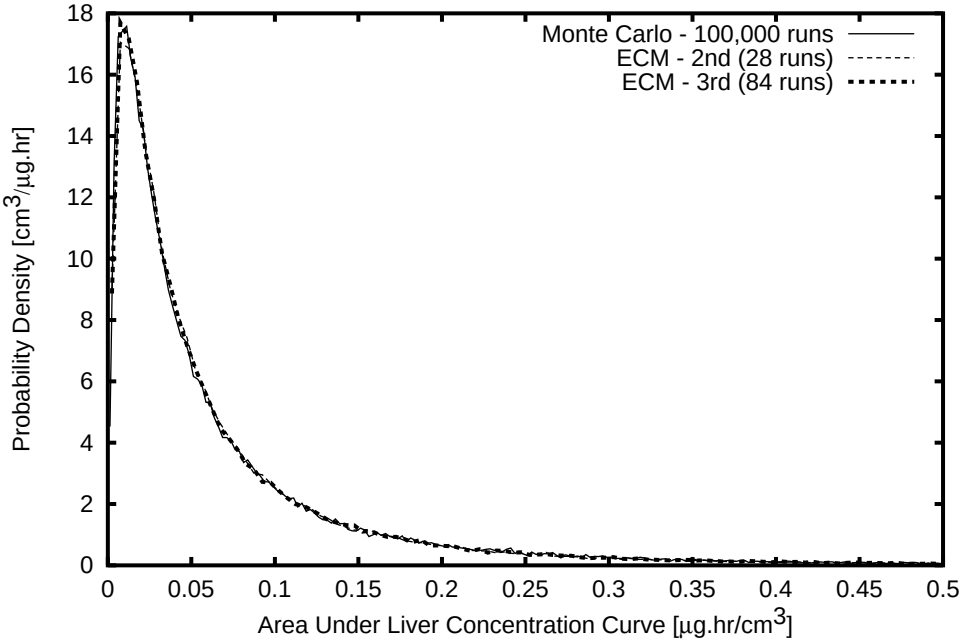
Figure 5.6: Evaluation of ECM: uncertainty in the cumulative amount of PERC metabolized, over a period of one day, resulting from 11 uncertain parameters

Stage II (Eleven Uncertain Parameters): Evaluation of ECM, Regression Based SRSM and LHS

In the second stage, the uncertainty in compartmental volume proportions are included in the analysis, resulting in eleven uncertain parameters. These five additional parameters are described by a set of mutually dependent Dirichlet distributed random variables. The number of simulations required for the 2nd and 3rd order DEM/ECM are 78, and 364, compared to 100,000 Monte Carlo simulations used to generate the dose surrogate *pdf*s. Figures 5.6 and 5.7 show the comparison of the *pdf*s of the CML, AUCA, and the AUCL dose surrogates. Again, The *pdf*s estimated by both 2nd and 3rd order ECM appear to agree well with each other and with the *pdf*s estimated using Monte Carlo simulation.



(a)



(b)

Figure 5.7: Evaluation of ECM: uncertainty in the area under the PERC concentration-time curve in (a) arterial blood (AUCA) and in (b) liver (AUCL), over a period of one day, resulting from 11 uncertain parameters

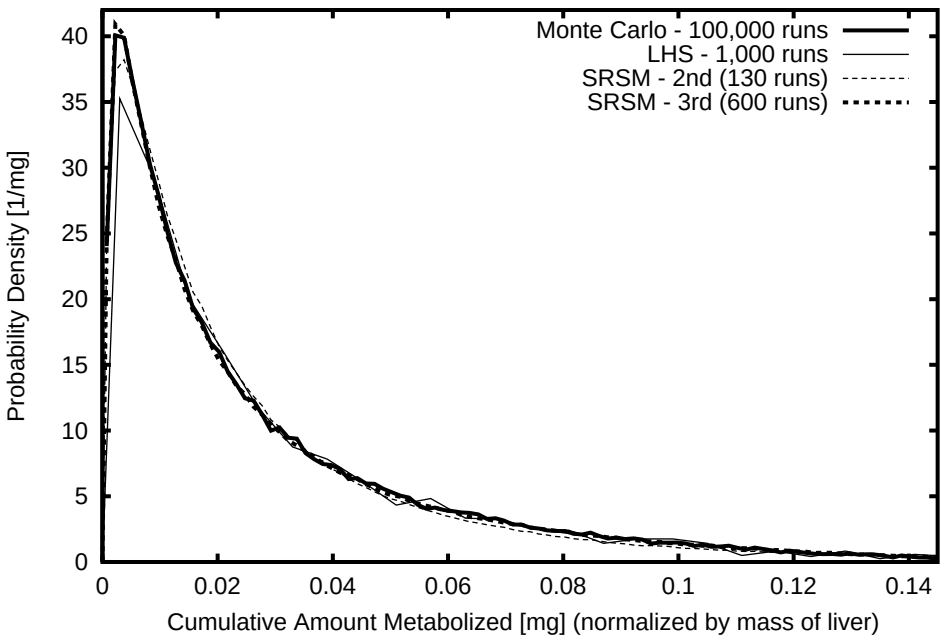
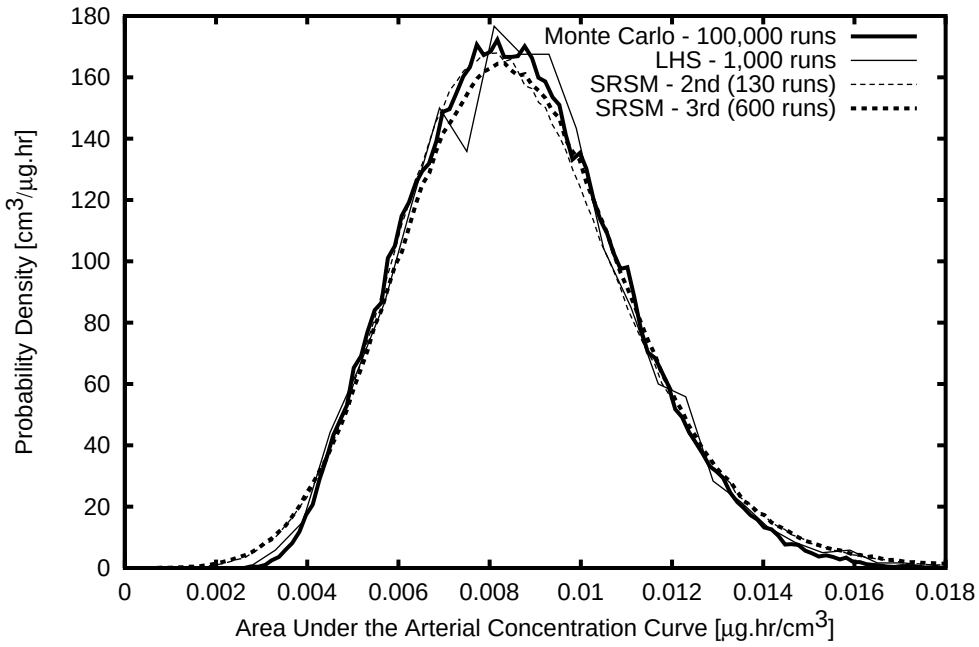
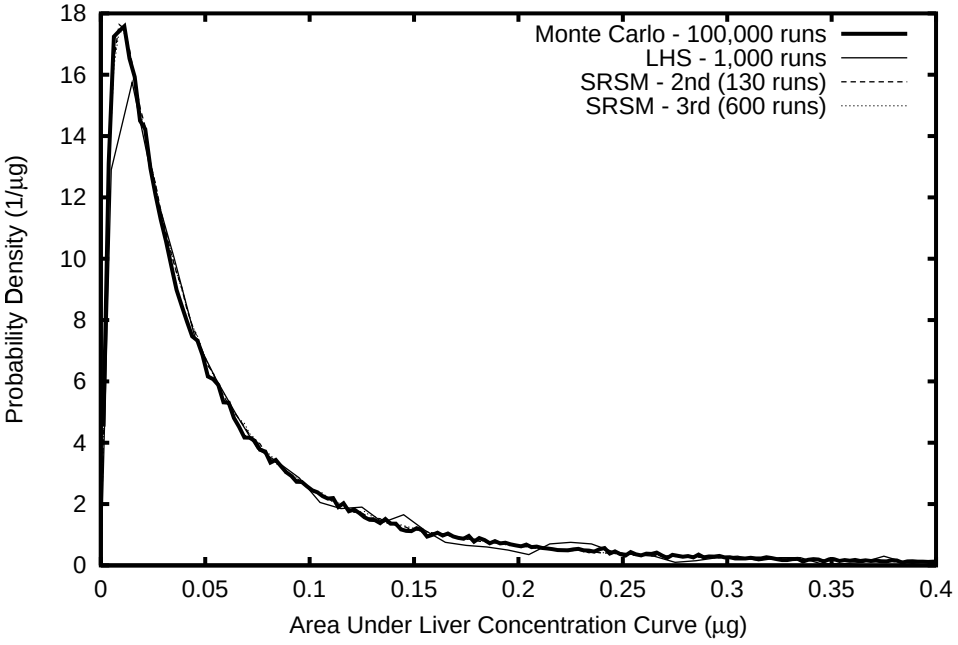


Figure 5.8: Evaluation of SRSM (regression based) and LHS: uncertainty in the the CML resulting from 11 uncertain parameters

The number of simulations used for the 2nd and 3rd order SRSM (regression based SRSM) are 130 and 600, respectively. On the other hand, 1,000 Latin Hypercube samples, and 100,000 Monte Carlo simulations used to generate the dose surrogate *pdf*s. For this case study, both the ECM and the regression based methods resulted in close approximations, indicating that for simple systems, the ECM method is preferable to the regression based SRSM. The *pdf*s estimated by both 2nd and 3rd order SRSM appear to agree well with each other and with the *pdf*s estimated using Monte Carlo simulation. Further, the SRSM approximations have a closer agreement with the Monte Carlo simulation results than the results from the Latin Hypercube sampling method. The results indicate that the SRSM is computationally significantly more efficient than the standard Monte Carlo and more accurate than the Latin Hypercube sampling method.



(a)



(b)

Figure 5.9: Evaluation of SRSM (regression based) and LHS: uncertainty in (a) AUCL and in (b) AUCA (Figure b) resulting from 11 uncertain parameters

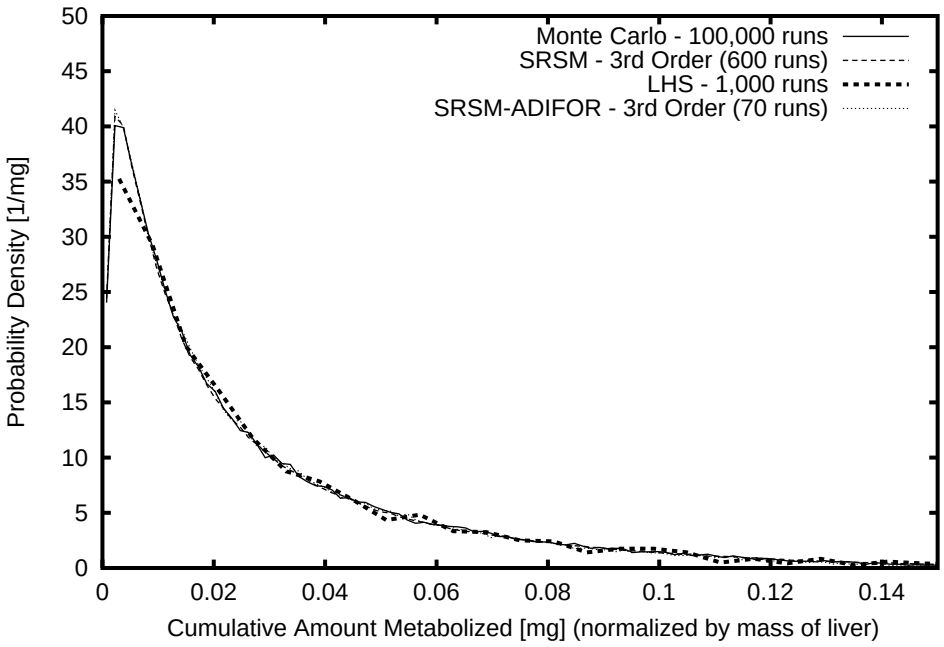
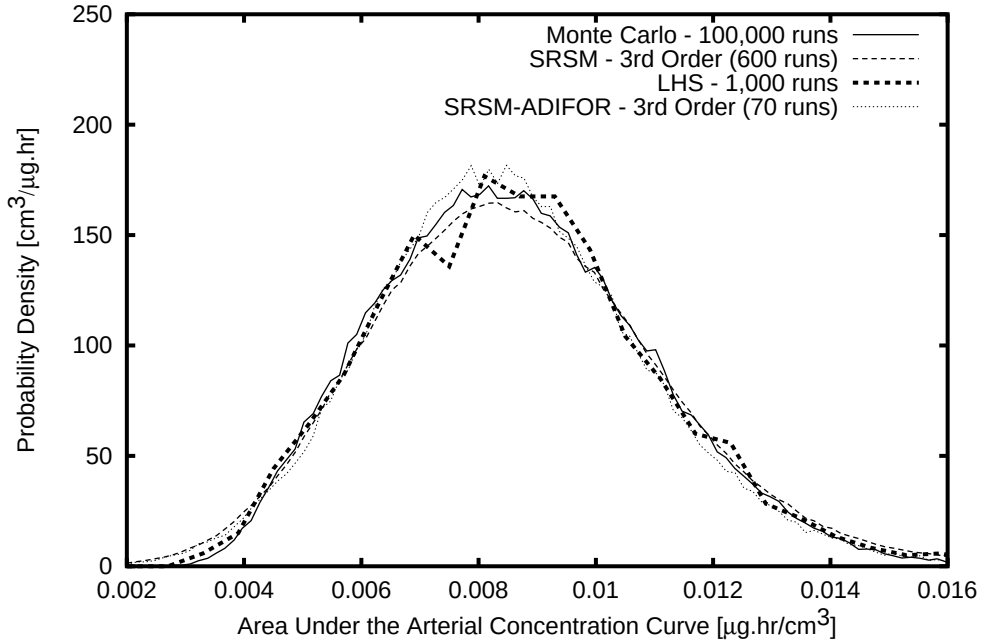


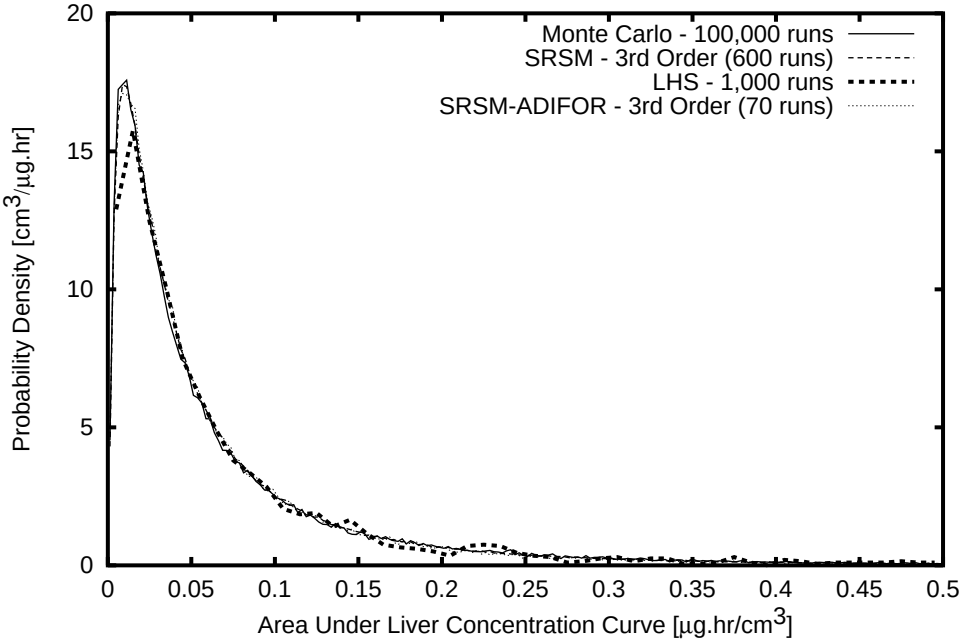
Figure 5.10: Evaluation of SRSM-ADIFOR: uncertainty in the cumulative amount of PERC metabolized, over a period of one day, resulting from 11 uncertain parameters

Evaluation of SRSM-ADIFOR

In order to apply the SRSM-ADIFOR, the PERC PBPK model has been re-implemented in standard FORTRAN and that model was used in conjunction with the ADIFOR system to obtain the derivative code. The derivative model was then run at a set of selected sample points for the inputs. The model outputs and derivatives obtained at those points were then used to approximate the *pdf*s of the three dose surrogates. The number of simulations used for the SRSM-ADIFOR approach for the third order was 70, compared to 600 simulations for the regression based SRSM.



(a)



(b)

Figure 5.11: Evaluation of SRSM-ADIFOR: uncertainty in the area under the PERC concentration-time curve in (a) arterial blood (AUCA) and in (b) liver (AUCL), over a period of one day, resulting from 11 uncertain parameters

5.2 Case Study II: A Two-Dimensional Photochemical Air Quality Model

Uncertainty Analysis of the Reactive Plume Model (RPM-IV)

The Reactive Plume Model, version IV (RPM-IV), is a standard regulatory model used for calculating pollutant concentrations and establishing causal relationships between ambient pollutant concentrations and the emissions from point sources such as industrial stacks [153, 195]. It uses either point source emission estimates or initial plume concentrations as inputs, and calculates downwind concentrations, as the plume expands.

RPM-IV is often applied for regulatory purposes to calculate the ozone (O_3) concentrations at locations downwind of industrial point sources, since high ozone concentrations in the ambient environment lead to adverse health effects. Ozone is primarily formed in the atmosphere through a series of complex chemical reactions involving oxides of nitrogen (NO_x) and volatile organic compounds (VOCs) in the presence of sunlight. Some of the major point sources of NO_x and VOCs are industrial units, such as the power plants (NO_x sources) and refineries (VOC sources). The application of RPM-IV helps in establishing a quantitative causal relationship between emissions and ambient pollutant concentrations, which is useful in assessing various control strategies for emission reductions.

However, there are significant uncertainties in developing estimates of the emissions from industrial sources. These uncertainties occur with respect to the amounts of emissions (e.g., total amounts of VOCs and NO_x), and with respect to their chemical compositions (or speciations, i.e., the fractions of various chemicals within these groups of compounds). These uncertainties arise due to a variety of reasons: for example, emission estimates are typically derived from hourly averages projected from annual or seasonal averages [38]. Since there could be a significant variation in the load and operating conditions of an industrial unit, emission estimates and chemical compositions for specific days under consideration may differ significantly from the averages, and thus result in significant uncertainties. Hence, an uncertainty analysis that takes into account the emission estimate uncertainties is useful for a better understanding of the

effects of control strategies for emission reductions.

5.2.1 Description of RPM-IV

A brief description of the RPM-IV model is presented here; additional details are presented in Chapter 6, which deals with the model uncertainty associated with RPM-IV, and which presents an improved, three-dimensional version of the model.

RPM-IV simulates mechanistically the complex nonlinear photochemistry and dispersion processes occurring in an expanding plume. The nonlinear atmospheric gas phase photochemistry is described by the Carbon Bond IV (CB-IV) mechanism [216], which consists of a set of 95 chemical reactions among 35 surrogate chemical species corresponding to organic bonds and functional groups. Details on the CB-IV mechanism are presented in Appendix C. This model follows the trajectory of an expanding, moving plume and simulates its evolution. In this model, the pollutant mass is initially divided into cells containing equal amounts of pollutants. As the plume expands, the individual cells expand in volume and pollutant mass is transferred across cell boundaries in two phases: (a) an "entrainment" phase, where the expanding cell boundaries entrain the pollutants from other cells, and (b) a "detrainment" phase, where the pollutants diffuse across cell boundaries, due to concentration gradients. Further, the pollutants in each cell undergo chemical transformation governed by the Carbon Bond-IV mechanism.

The equation describing the concentration changes within a cell is given by:

$$\frac{dc_j^i}{dt} = \left(\frac{dc_j^i}{dt}\right)_{\text{chem}} - \left(\frac{1}{w_j}\cdot\frac{dw_j}{ds}\right)uc_j^i - \left(\frac{1}{h_j}\cdot\frac{dh_j}{ds}\right)uc_j^i + F_j^i \tag{5.4}$$

where c_j^i denotes the concentration of chemical species i in cell j. F_j^i is the net flux into the cell, and is given by $F_j^i = E_j^i - D_j^i$, where E_j^i is the entrainment as the cell expands, and D_j^i is the detrainment due to the diffusion of the species across the boundaries of the cell. Expressions for entrainment and detrainment are given by the following equations:

$$E_j^i = \frac{1}{y_j - y_{j-1}}\left\{\left(\frac{dy_j}{dt}\right)c_{j+1}^i - \left(\frac{dy_{j-1}}{dt}\right)c_j^i\right\} \tag{5.5}$$

$$D_j^i = \frac{\partial}{\partial y}\left(K\frac{\partial c}{\partial y}\right) = \frac{2}{y_j - y_{j-1}}\left\{K_j\left(\frac{c_{j+1}^i - c_j^i}{y_{j+1} - y_{j-1}}\right) - K_{j-1}\left(\frac{c_j^i - c_{j-1}^i}{y_j - y_{j-2}}\right)\right\} \tag{5.6}$$

where y_j is the distance from the plume centerline to the far most side of the cell j, in the horizontal direction. An additional condition is imposed on the expansion of the cells: the cells are allowed to expand in a way such that the amount of an inert species remains constant within each cell, i.e., no net transfer of an inert species occurs from a cell. This condition results in the following equation for expansion of the cells, as the plume travels downwind:

$$E_j^I - D_j^I = 0$$

RPM-IV was selected for evaluation of the SRSM and SRSM-ADIFOR because it is sufficiently complex to represent a wide range of environmental models, and at the same time it is computationally feasible to perform a large number of Monte Carlo simulations with this model. Further, the complex nonlinear photochemistry of this model is employed by a number of photochemical models that are computationally very demanding. Thus, evaluation of the SRSM and SRSM-ADIFOR with RPM-IV could potentially serve as a preliminary test of applicability of this method to a wide range of complex photochemical models.

5.2.2 Uncertainty Analysis of RPM-IV

The present work studies the effect of uncertainty in emission estimates on predicted downwind secondary pollutant concentrations; here, the specific focus is on the uncertainties in the calculated ozone concentrations resulting from the uncertainties in amounts and the chemical composition of the emissions. This study uses emission estimates and field data measured near Marathon oil refinery at Robinson, Illinois, during June and July, 1977 by Sexton et al. [183]. Since there was no characterization of uncertainty, and since the focus of this work is to evaluate the applicability of SRSM-ADIFOR, only representative probability distributions to describe the uncertainties in emissions were assumed.

The following distributions for the emissions of VOCs and NO_x are used: (a) the amounts of VOCs and NO_x released are assumed to have normal distributions with a standard deviation of 20% of the mean value, and (b) the chemical compositions of

VOCs and NO_x are assumed to follow a Dirichlet distribution [201]. The Dirichlet distribution satisfies the condition that the sum of mole fractions is unity. According to this distribution the mole fraction of the ith compound, y_i, is given by:

$$y_i = \frac{x_i}{\sum_{i=1}^{n} x_i}, \tag{5.7}$$

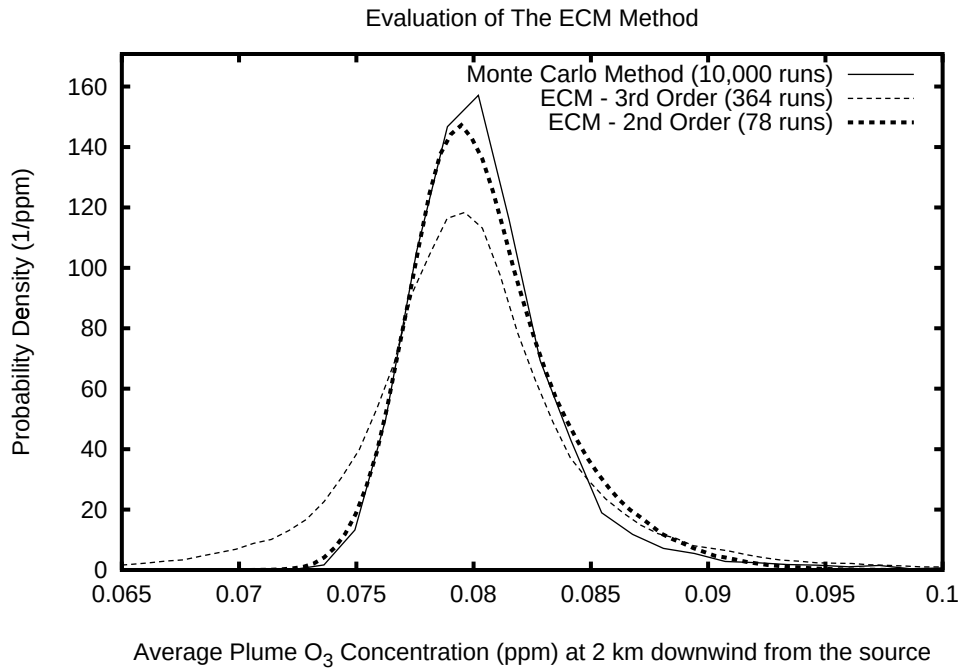
where x_i is an independent random variable, representing the the amount (in moles) of the ith compound; here a normal distribution is assumed for all x_is, with a nominal standard deviation of 40% of mean value. In this case study, the VOC group consists of five subgroups (paraffins, ethylene, olefins, toluene, and xylene), and the NO_x group consists of NO and NO_2. Thus, the total number of uncertain parameters for the model is nine (including two parameters representing the total amounts of VOCs and NO_x). Thus, a total of 9 uncertain parameters with 7 degrees of freedom were used in this case study; these were represented by 9 *srv*s.

The output metrics considered are the average ozone concentration in the plume for selected distances downwind from the source.

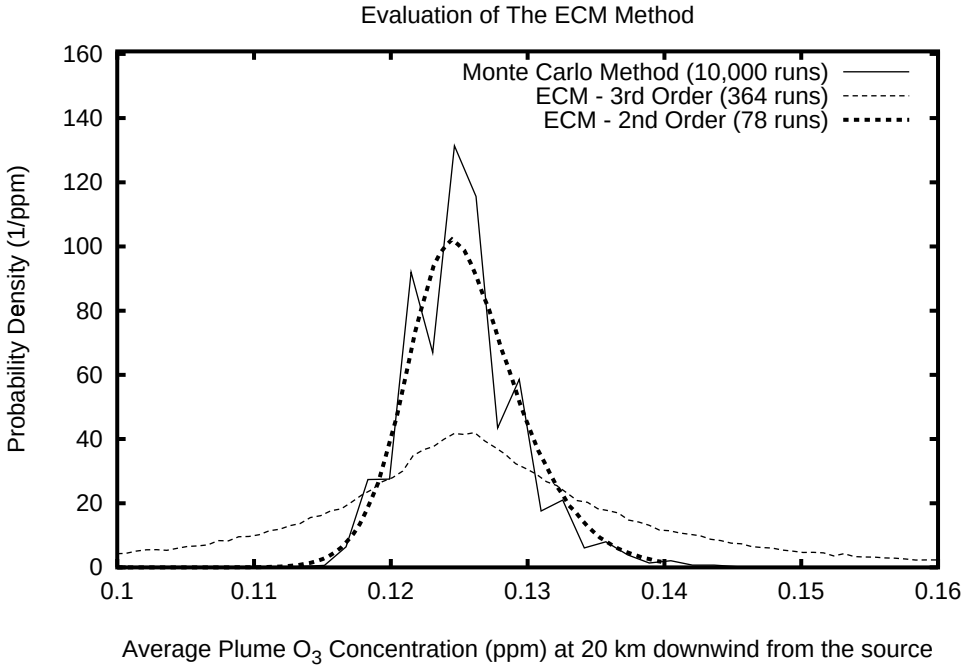
The original RPM-IV model was implemented in Fortran and obtained from the EPA Exposure Models Library [202]. The derivative code was obtained using the ADIFOR system on RPM-IV code. For Monte Carlo, LHS, and the SRSM, the model was run at selected sample points, whereas for the SRSM-ADIFOR, the derivative model was run.

5.2.3 Results and Discussion

From equations 3.6 and 3.6 for the number of unknown coefficients for a second and third order polynomial chaos expansions, for n=9, the number of coefficients to be determined is 78 and 364 for second and third order approximations, respectively. For a regression based method, 300 model simulations were selected for a second order approximation, and 600 model simulations were selected for a third order approximation, in order to facilitate regression with an adequate number of model outputs. The estimates of *pdf*s for these cases were evaluated with the results obtained from 10,000 Monte Carlo simulations and 10,000 Latin Hypercube sampling methods. For the case of the SRSM-ADIFOR, 80 simulations were used for a third order approximation.



(a)



(b)

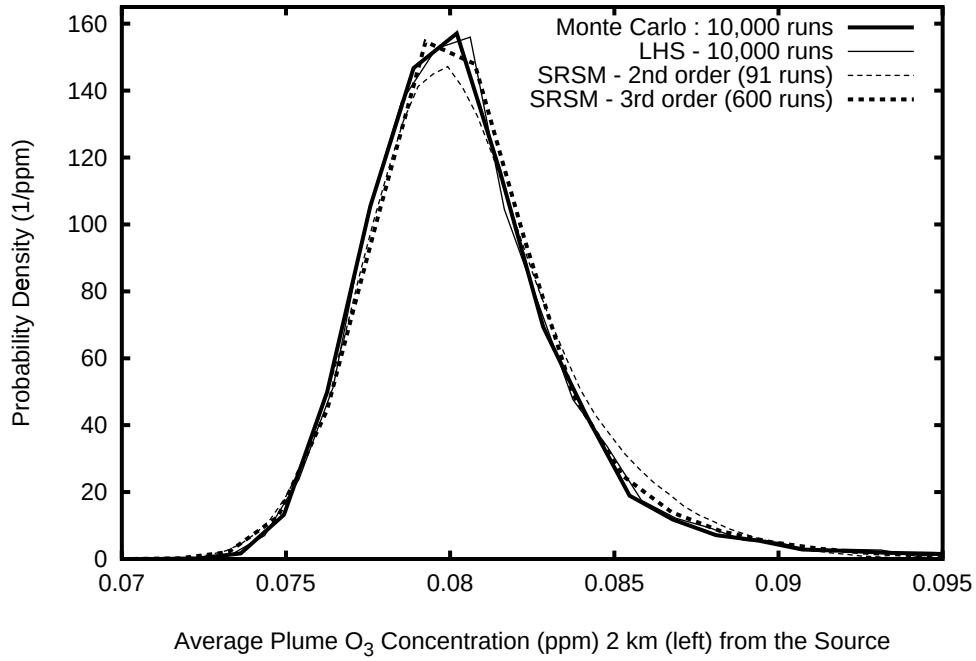
Figure 5.12: Evaluation of ECM: uncertainty in the predicted ozone concentrations at (a) 2 km and (b) 20 km downwind from the source

Two representative output metrics were considered here: (a) ozone concentration at a downwind distance of 2 km, representative of near-source transport and transformation, and (b) ozone concentration at a downwind distance of 20 km, representative of larger scale transport and transformation. The output *pdf*s were obtained first using second and third order approximations of the SRSM/ECM. Figure 5.12 shows the *pdf*s of ozone concentration at a downwind distance of 2 km and of 20 km, as estimated by the ECM, and the conventional Monte Carlo method.

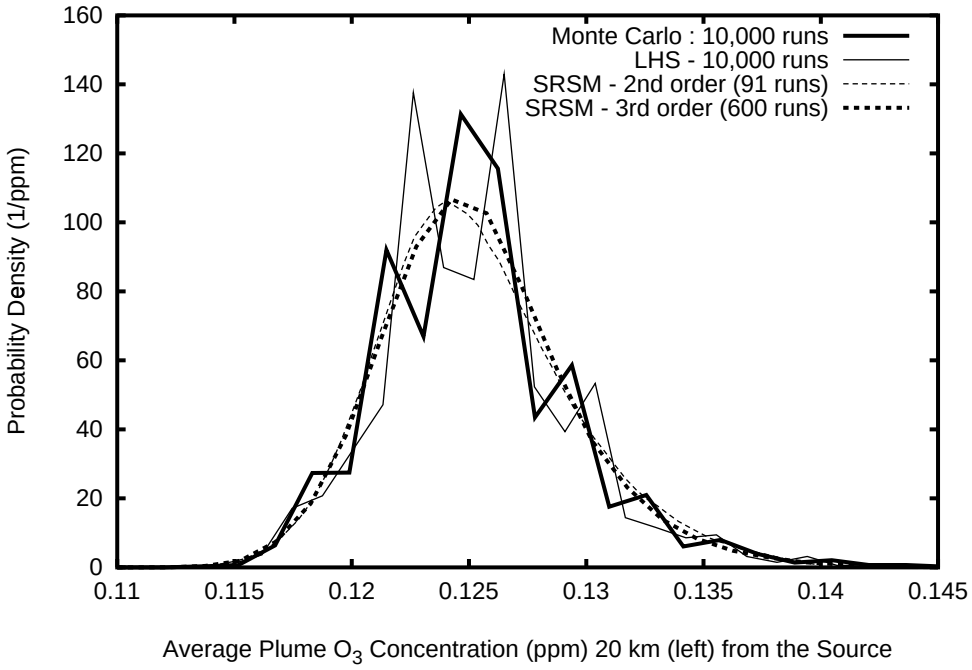
Figure 5.13 shows the *pdf*s of the predicted ozone concentrations at downwind distances of 2 km and 20 km, as estimated by the regression based SRSM, traditional Monte Carlo and Latin Hypercube Sampling.

As shown in Figures 5.12 and 5.13, although the regression based-method required significantly fewer runs than the Monte Carlo method, the results agree very closely with the Monte Carlo results. On the other hand, the the predictions of ECM become inaccurate as the order of approximation increases, indicating the lack of robustness in the collocation method. This behavior is more prominent for large downwind distance, indicating that the collocation method may not converge when used with highly non-linear models (the near-source behavior is expected to be less nonlinear than far-source behavior). On the other hand, a regression based method resulted in similar estimates for both second and third order approximations and was consistent for all ranges of downwind distances. The results indicate that, although the regression methods require a higher number of model simulations for uncertainty propagation, compared to the collocation methods, their robustness makes them a more viable tool for uncertainty analysis of complex environmental models.

For the evaluation of the SRSM-ADIFOR method with RPM-IV, Figure 5.14 shows the uncertainty estimates obtained. As shown in the figure, SRSM/ADIFOR gave closer estimates with only 80 model runs, while 10000 Latin Hypercube samples were not sufficient to achieve agreement with the results from the Monte Carlo methods.

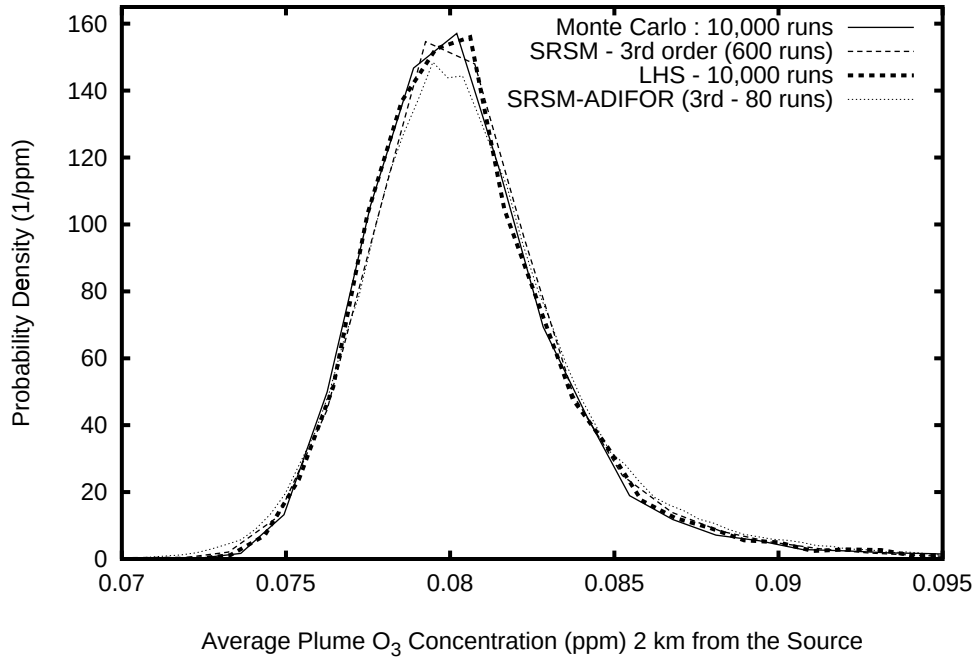


(a)

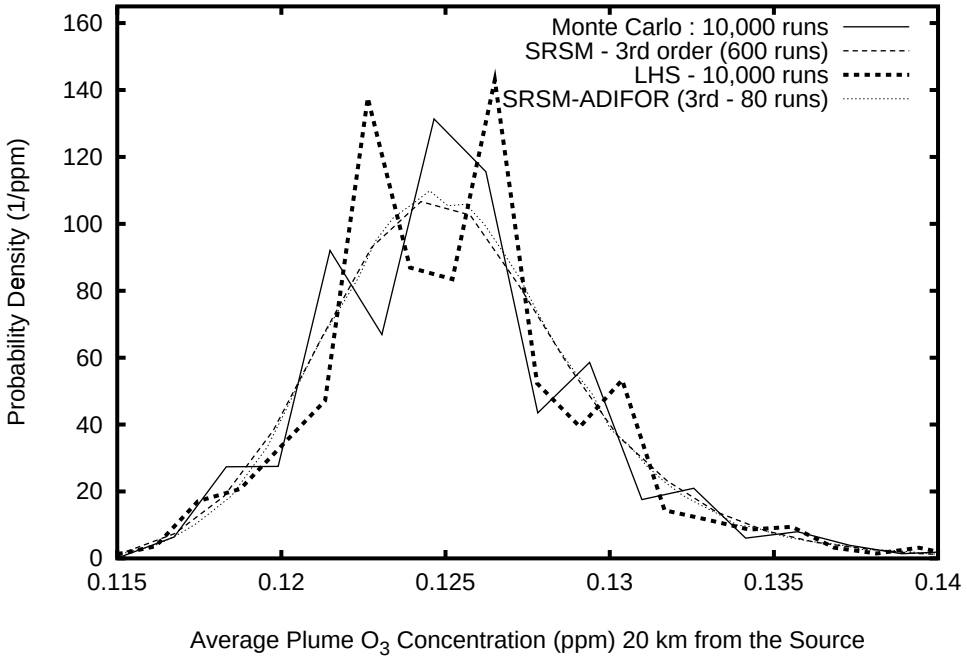


(b)

Figure 5.13: Evaluation of SRSM (regression based): uncertainty in the predicted ozone concentrations at (a) 2 km and (b) 20 km downwind from the source



(a)



(b)

Figure 5.14: Evaluation of SRSM-ADIFOR: uncertainty in the predicted ozone concentrations at (a) 2 km and (b) 20 km downwind from the source

5.3 Case Study III: A Three-Dimensional Urban/Regional Scale Photochemical Air Quality Model

Uncertainty Analysis of the Urban Airshed Model (UAM-IV) for the New Jersey/Philadelphia Domain

Photochemical modeling of ozone (O_3) levels in the atmosphere is complicated by the fact that ozone is a secondary pollutant; it is formed through complex chemical reactions involving oxides of nitrogen (NO_2 and NO), and volatile organic compounds (VOCs). The modeling process is further complicated by the presence of natural, model, and data uncertainties, as shown in Figure 5.3. Uncertainties in the input data and parameters (such as emission estimates and reaction rates) become especially significant when photochemical models are used to evaluate control strategies for the reduction of ozone levels [43].

Biogenic emissions have a significant impact on atmospheric ozone concentrations, depending on the availability of NO_x. Earlier studies on the effects of biogenic emissions on O_3 concentrations indicate that typically O_3 concentrations increase in response to increases in biogenic hydrocarbons. However, modeling studies suggest that in some extremely NO_x-limited areas, increases in biogenic emissions decrease O_3 concentrations [173]. Sensitivity tests in a case study show that local biogenic emissions are an important contributor to local O_3 production, relative to anthropogenic hydrocarbons (AHCs) [142]. Isoprene forms a main component of biogenic emissions, and affects the chemistry of the troposphere because its oxidation products are precursors for the photochemical production of ozone [11]. It is one of the most abundant phytogenic chemical species found in the ambient air [73], and accounts for a substantial portion of the atmospheric hydrocarbon load [88, 185].

However, biogenic emission estimates are laden with large uncertainties [173]. In addition, there are significant uncertainties associated with the chemical mechanisms employed in photochemical modeling, such as uncertainty in the chemical reaction rates and uncertainties arising from the lumping of chemical species [7]. The focus of the present work is on the comparative evaluation of the effects of uncertainty in the biogenic

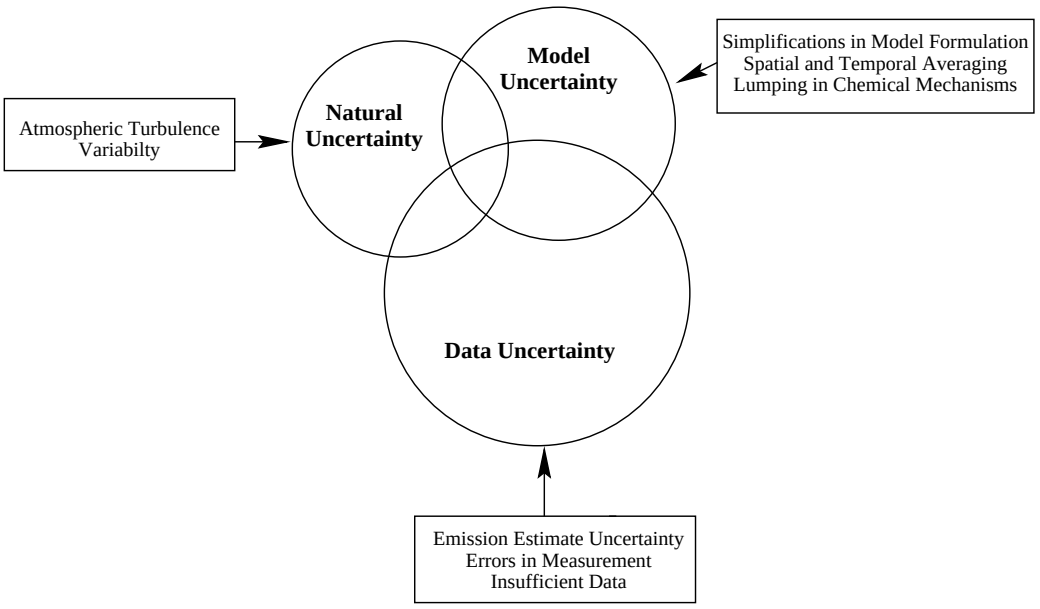


Figure 5.15: Origins and types of uncertainty present in Photochemical Air Quality Simulation Models

emission estimates and uncertainties in the reaction rate coefficients on the estimated ozone levels, with respect to isoprene (2-methyl-1,3-butadiene), since it is one of the major components of biogenic emissions. More specifically, the aim is to compare the effects of uncertainties associated with the emission rates of isoprene with the effects of uncertainties in the rates of reactions involving isoprene. Such uncertainty analysis could provide information that could be useful in identifying the area to focus resources: studying the reaction mechanisms in more detail versus improving the inventorying methods.

5.3.1 Uncertainties Associated with Biogenic Emission Estimates

There are significant uncertainties associated with biogenic emission estimates, and specifically with isoprene emission estimates. These uncertainties occur mainly due to the following factors:

- variability in the vegetation: the amounts of biogenic emissions differ significantly from one plant species to other, and there is a wide range of plant species in an urban/regional scale modeling area.
- variability in emission rates: the emission rates depend on factors such as sunlight,

temperature, nitrogen and water availability, and are associated with significant spatial and temporal variability.

Currently employed models and associated inventories for estimating biogenic emissions typically cannot treat the above variability in adequate detail. Hence significant uncertainties occur in the emission estimates.

The total U.S. emission estimates of biogenics are reported to range from 22 to 50 Tg/yr depending upon the formulation of different emission rate factors [135]. Some studies report the uncertainty in the biogenic emission estimates of an inventory for the United States to be of the order of 300% [136,173]. In another study, the uncertainties in the emission estimates of isoprene in Europe are reported to be of the order of 500% [188].

In addition, there are significant uncertainties associated with the description of isoprene chemistry. Recently reported studies show that different photochemical mechanisms with different treatments of isoprene chemistry give significantly different estimates of ozone levels [225]. The main sources of uncertainties in existing chemical mechanisms are: (a) lumping of various biogenic compounds, and (b) insufficient data to estimate reaction rate parameters associated with biogenic compounds. It is reported in the literature that some reaction rate parameters associated with isoprene chemistry are uncertain up to a range of 150% [7].

5.3.2 Uncertainty Analysis

Considering the importance of biogenic emissions, and the uncertainties involved in their estimates and reaction rates, it is important to study how these uncertainties propagate through the model and affect the predicted O_3 concentrations. For simple models, uncertainty propagation can be accomplished through conventional methods such as Monte Carlo methods. However, grid-based photochemical models are typically very demanding computationally, thus making conventional Monte Carlo methods prohibitively expensive.

Due to the lack of computationally efficient uncertainty analysis methods, a number

of studies reported in the literature with respect to uncertainty in photochemical models involved running the model for selected sets of parameter or input values, and thus obtaining some estimates on the "bounds" of uncertainty in model results. The main advantage of this approach is that it is simple and requires few model runs. However, this approach ("sensitivity/uncertainty" testing) does not provide insight into the distribution of the model results. Here, the regression based SRSM is applied to estimate the probability distributions of the estimated ozone concentrations.

As mentioned earlier, one goal of this approach is to characterize individual contributions of uncertainties in chemistry and emissions on the ambient ozone levels, and to identify important contributors to uncertainty, so that resources could be focused to reduce uncertainties in those factors. Another goal is to obtain an approximate estimate of the range of uncertainty in the predicted ozone concentration, resulting from the above mentioned input uncertainties; such an estimate could potentially assist in answering some regulatory questions.

5.3.3 Estimation of Uncertainties

The present work involves the study of the effects of uncertainty in the emission and reaction rates of isoprene, on the ozone levels in the atmosphere. This is pursued as follows:

- a single cell ("box") model is used as a screening tool to identify parameters that could potentially affect the results of the grid-based PAQSMs. The parameters considered include the the initial concentration of isoprene, and the rate constants for reactions involving isoprene in the CB-IV chemical mechanism.

- approximate estimates of uncertainty in the parameters thus identifies are obtained from the literature, and are then propagated through the Urban Airshed Model (UAM-IV) to study urban/regional scale effects of these uncertainties.

More specifically, a photochemical box model is used to perform preliminary screening based on the effects of uncertainty in isoprene reaction rate constants and initial concentrations of isoprene on the time-dependent concentrations of ozone. The results

from this exercise are utilized in identifying key factors contributing to uncertainty in the application of a grid-based PAQSM, specifically the UAM-IV.

A box-model consists of a single well mixed reactor cell, and the evolution of chemical species depends only on time dependent photochemical reaction rates. In the present work, a single cell option of the Reactive Plume Model (RPM-IV) is used for the screening runs. A single cell trajectory model, is equivalent to a box model, since there is no transfer of material occurring from outside the cell. The RPM-IV was used here since the same chemical mechanism, the Carbon-Bond IV (CB-IV) mechanism [216], is employed to describe the chemistry in both RPM-IV and UAM-IV. The CB-IV mechanism (see Appendix C) consists of a set of 95 chemical reactions among 35 surrogate chemical species corresponding to organic bonds and functional groups.

The Urban Airshed Model, version IV, [154] is a photochemical grid model that has been applied to several urban areas in the United States, Europe and the Far East [63, 78]. This model solves the atmospheric diffusion equation (ADE) [181] on a three dimensional grid covering the airshed of interest. Atmospheric chemistry is described by the CB-IV mechanism. This model requires meteorological inputs and the emission information for both anthropogenic and biogenic emissions. Further details of the UAM-IV model are presented in Appendix C.

This case study considers the modeling of the Philadelphia-New Jersey region for July 19 and 20, 1991, when a severe ozone episode occurred in this region. The domain and the modeling grid (at a horizontal resolution of 5 km$\times$5 km) for the case study are shown in Figure 5.16.

The meteorological inputs were developed using the Diagnostic Wind Model (DWM) [57], the anthropogenic emissions were obtained from State-provided inventories, processed using the Emission Processing System (EPS) [31]. The biogenic emissions were obtained from the EPA's Biogenic Emissions Inventory System (BEIS) model [31]. These inputs were obtained from the input data from Georgopoulos et al. [78]

The procedure used for uncertainty analysis consisted of the following steps:

- Performing systematic screening tests to study the sensitivity of the results of the

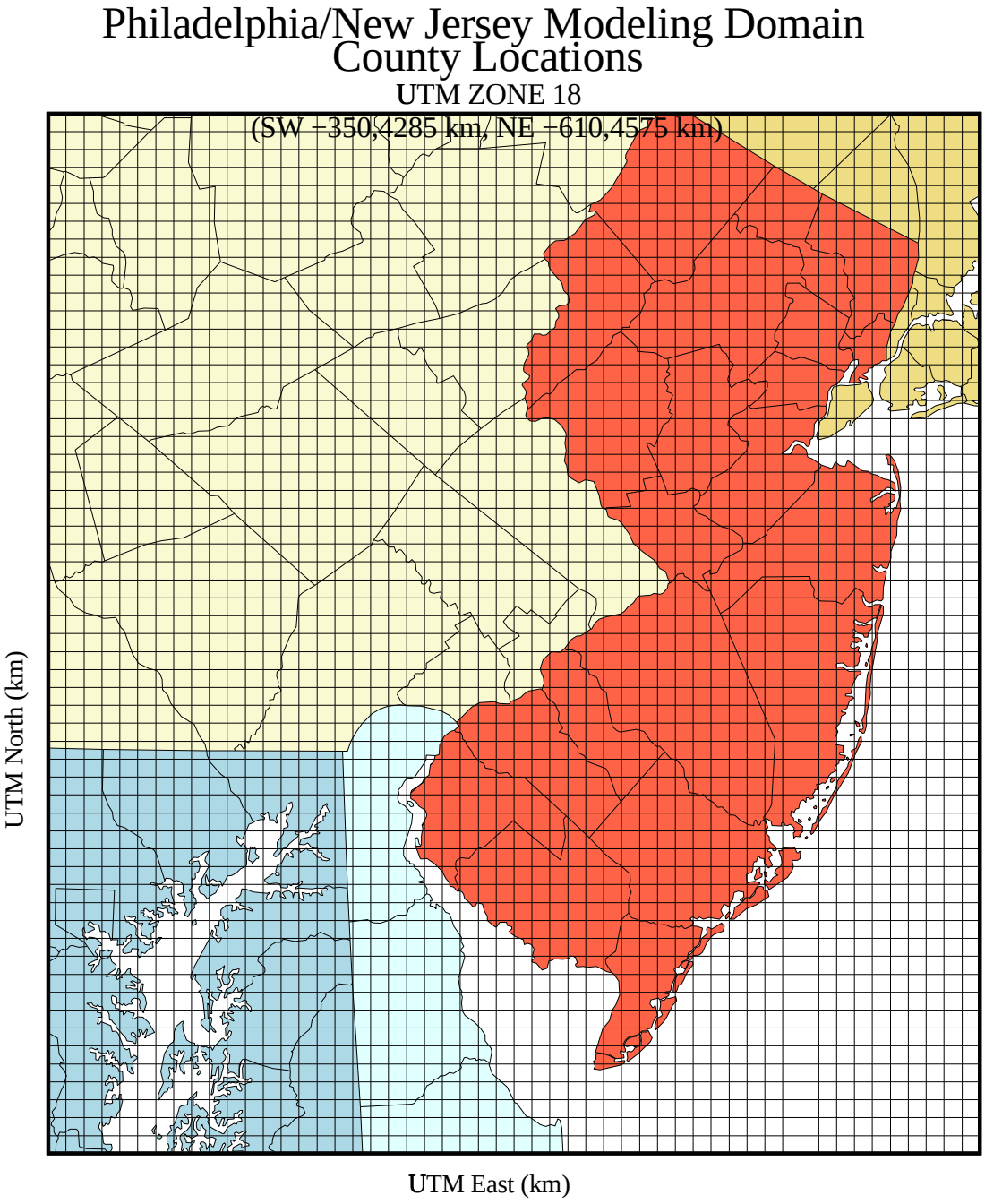


Figure 5.16: The Philadelphia/New Jersey modeling domain used for uncertainty analysis

box-model with respect to selected reaction rate constants of isoprene, and initial concentration of isoprene

- Identifying the factors that have clear effects of the predicted ozone concentration

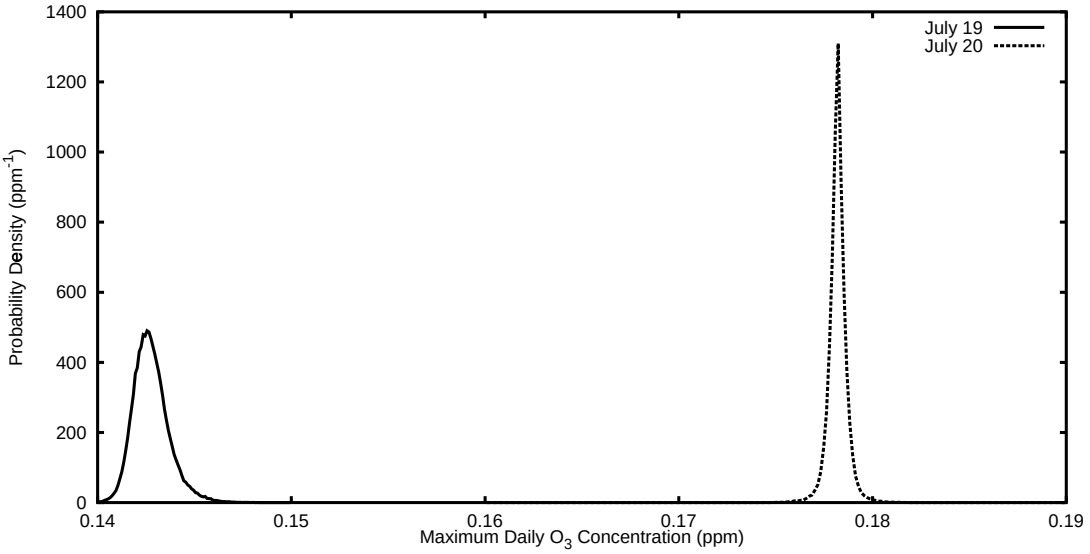


Figure 5.17: Probability distribution of the daily maximum ozone concentration

- Obtaining approximate uncertainty estimates for these factors from the literature [7]
- Performing uncertainty analysis using the Stochastic Response Surface Method, and studying the contributions of these uncertainties to the uncertainty in the ozone concentrations predicted by the UAM-IV
- Characterizing the relative importance of these parameters, and the overall uncertainty in predicted ozone concentrations for specific case studies.

5.3.4 Results of the Case Study

In the first phase (screening phase), nine parameters were considered, and the objective was to identify parameters that have a significant effect on the O_3 concentration predictions. The box model was used to to calculate the O_3 concentration. The nine parameters included initial concentrations of isoprene and monoterpenes, and the rates of seven reactions of isoprene with the O_3, NO_3, O and OH radicals. Of these, only three factors, namely the initial concentration of isoprene, and the reaction rates of the reactions of isoprene with NO_3 and with the OH radical were found to have an appreciable effect on the calculated O_3 concentrations.

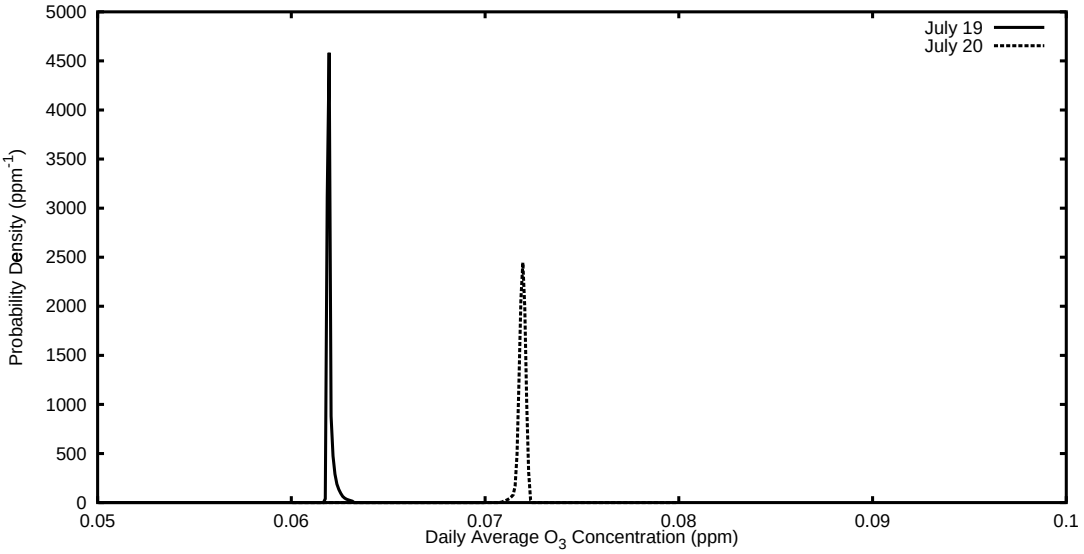


Figure 5.18: Probability distribution of the daily average ozone concentration

In the second phase, representative estimates of uncertainties in these parameters were used to study their effects on urban scale O_3 levels. Truncated lognormal probability distributions were used to represent uncertainties in these parameters. The value of a given parameter, x, could lie between x_m/f and $x_m f$, where x_m is the reported value and f is an uncertainty factor that is estimated *a priori*. This means that $\log(x)$ is normally distributed with a mean $\log(x_m)$, and a variance $\log(f)/2.5$. Here, 2.5 is used as a truncation factor to specify the probability distribution uniquely, and to limit the range of values x could have.

More specifically, the uncertainties in isoprene emission estimates, and the rate constants for the reactions of isoprene with NO_3 and with the OH radical, were all assumed to be represented by a truncated lognormal distributions with uncertainty factors of 1.5; this is a rough estimate of the uncertainties based on the articles by Roselle [173] and by Atkinson [7].

The application of SRSMs, for this case study involving 3 uncertain parameters, required 10 simulations for a second order approximation method, and 20 simulations for a third order approximation. The approximation converged for the third order - there was an almost negligible difference between the estimates of the second and third order approximations. The third order approximations were used to estimate the

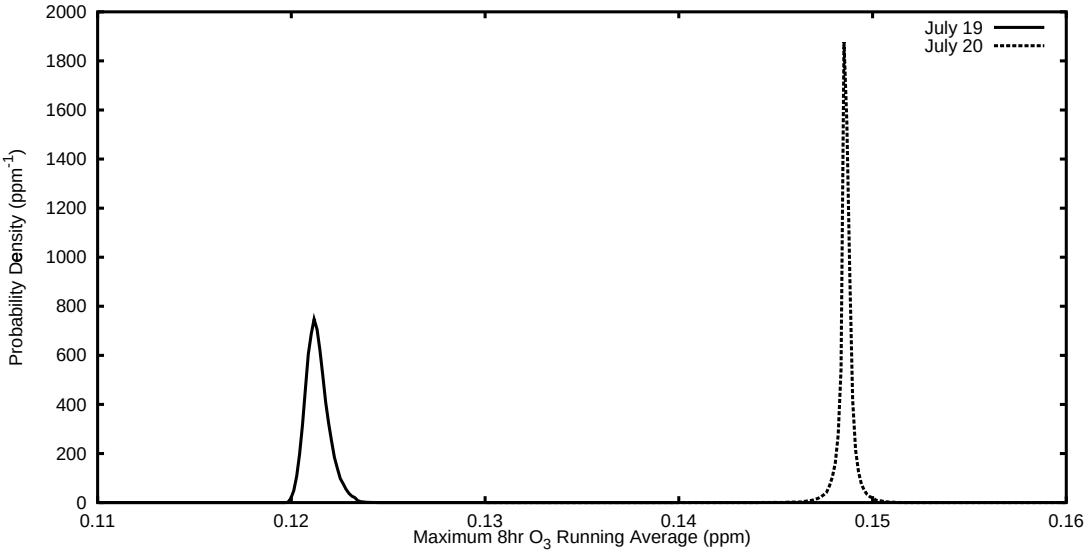


Figure 5.19: Probability distribution of the daily maximum 8 hr running average ozone concentration

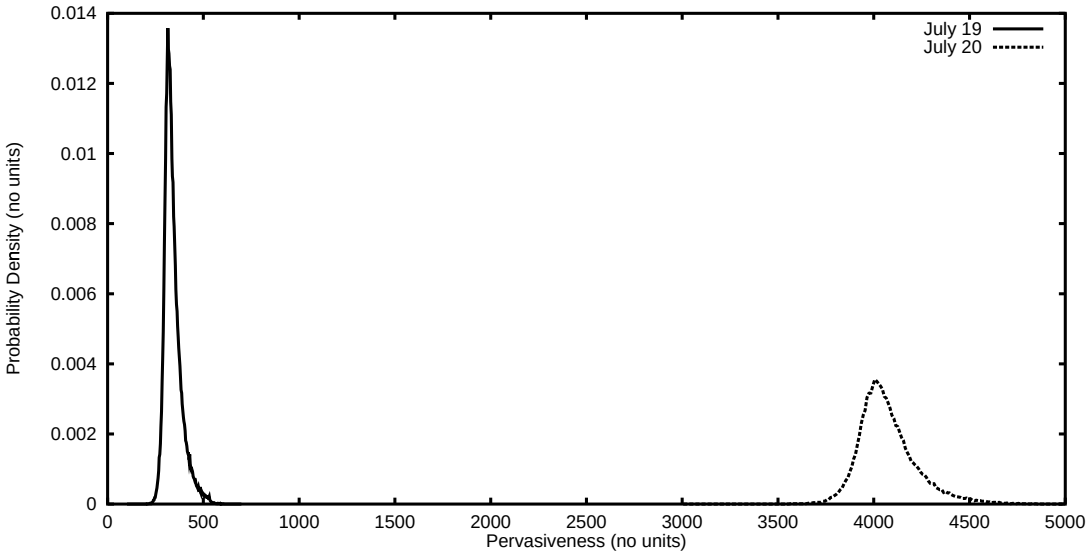


Figure 5.20: Probability distribution of the pervasiveness of the ozone episode

uncertainties in the output metrics considered.

The effect of uncertainty in these parameters, on four representative metrics of photochemical air quality O_3 levels through the study domain was studied. The four metrics were: (a) the daily maximum of hourly average O_3 concentration, (b) the daily average O_3 concentration, (c) the daily maximum eight hour running average (an important

metric in relation to the new ozone regulatory standard), and (d) the "pervasiveness", an alternative metric, defined as the number of grid cell hours that have ozone concentrations in excess of the hourly maximum O_3 standard of 120ppb [78, 81].

Figure 5.17 shows the probability distributions describing the uncertainty in the daily maximum of hourly average concentration for the episode days of July 19 and 20, 1991, for the Philadelphia-New Jersey domain. Figures 5.18 and 5.19 describe the uncertainty in the daily average concentration, and the daily maximum eight hour running average concentration, for the same case, whereas Figure 5.20 shows the uncertainty in the pervasiveness.

In summary, the following conclusions were reached:

- The daily average O_3 concentration was not affected significantly by the input uncertainties considered.
- The daily maximum eight hour running average was affected mildly by the input uncertainties.
- The daily maximum hourly average was also affected mildly.
- The pervasiveness was affected significantly by the uncertainty in the inputs considered.
- The probability distributions of output metrics differed significantly for the two episode days considered, implying that the uncertainty analysis method needs to be applied on a case by case basis.
- The variance in the concentration metrics for all cases was less than 1% of the mean value.
- The variance in pervasiveness was of the order of 5%-10% of the mean value.

Table 5.2 summarizes the uncertainty in the uncertainty estimates resulting from this analysis. The coefficients in the table indicate the uncertainty contributed to a metric, due to the uncertainty in the individual parameter. The coefficients indicate that the uncertainty contribution is very small relative to the mean value of the metric. Further, they indicate that the uncertainty in the isoprene emission rate contributes

Table 5.2: Contribution of uncertainties in the individual parameters to the overall uncertainty in the UAM-IV, as estimated by the SRSM

	Jul 19				Jul 20			
	Mean	OH	NO_3	ISOP	Mean	OH	NO_3	ISOP
Max (ppm)	0.143	0.0003	0.0000	0.0009	0.178	0.0003	-0.0001	0.0003
Ave (ppm)	0.062	0.0000	0.0000	0.0002	0.072	-0.0001	0.0000	0.0001
8hr-Ave (ppm)	0.121	0.0000	-0.0001	0.0005	0.149	-0.0003	0.0000	0.0001
Pervasiveness	341.71	6.32	-1.22	47.53	4075.0	31.30	-9.17	157.14

significantly to the overall uncertainty, and that the uncertainty in the reaction rate of the NO_3-isoprene reaction has an almost negligible effect. However, the uncertainty in the OH-isoprene reaction constant contributes to an extent comparable to the isoprene emission uncertainty. So, the above analysis indicates that the isoprene emission rates and the reaction rate constant of the isoprene-OH reaction should be studied in more detail.

This case study also demonstrates how the SRSM, while requiring a small number of simulations, can be used to estimate the uncertainties in the outputs of a complex model such as UAM-IV. The execution time, about five days (20 model runs $\times$ 6 hours per model run) is much less when compared with a Monte Carlo analysis involving, for example, 500 runs, requiring 125 days of computer time.

5.4 Case Study IV: A Ground Water Model with Discontinuous Probability Distributions

Uncertainty Analysis of Tritium Contamination in a Landfill using the EPACMTP Model

Characterization of exposure due to leakage of hazardous contaminants from a land disposal unit involves significant uncertainty due to inherent variability in the hydro-geologic properties of the site, and due to incomplete understanding of the processes involved in transport of contaminants. Since the exposure estimates are utilized in

making policy decisions, it is important that any exposure characterization should take into account the variability and uncertainty involved with the physical system, and with the modeling process.

Here, a case study is presented in relation to characterization of uncertainty in estimated tritium exposure at a receptor well. The main source for the contamination is a hypothetical landfill unit in the southern United States. The fate and transport model used in this work is the EPA's Composite Model for leachate Migration and Transformation Products (EPACMTP). In the following sections, a brief description of the model used, the estimates of the model parameters and the uncertainty analysis methods is given. The variability and uncertainty associated with the hydrogeologic parameters and with the physical properties of the landfill unit are considered to characterize the uncertainty in the calculated concentrations of tritium. Environmental metrics considered are the estimated maximum concentration of tritium in a "receptor" well and the estimated time of occurrence of the maximum concentration.

5.4.1 EPACMTP Model

The EPA's Composite Model for leachate Migration and Transformation Products (EPACMTP) provides estimates of potential human exposure to hazardous chemicals leaching from land disposal facilities [64]. EPACMTP simulates the subsurface fate and transport of contaminants released from land disposal sites, and predicts the associated groundwater exposure in a domestic drinking water receptor well. This model is an improvement over the EPA's Composite Model for Landfills (EPACML) [62]. EPACML accounts for the first-order decay and sorption of chemicals, but disregards the formation and transport of transformation products. In addition, EPACML can describe only uniform, unidirectional groundwater flow. On the other hand, EPACMTP can take into consideration: (i) chain decay reactions and transport of daughter and grand-daughter products, (ii) effects of water-table mounding on groundwater flow and contaminant migration, (iii) finite source as well as continuous source scenarios, and (iv) metals transport.

EPACMTP consists of two modules: an unsaturated zone module called Finite Element and semi-analytical Contaminant Transport in the Unsaturated Zone (FECTUZ), and a saturated zone module called Combined Analytical-Numerical SAturated Zone in 3-Dimensions (CANSAZ-3D). FECTUZ is a one-dimensional model that simulates vertically downward steady-state flow and contaminant transport through the unsaturated zone above an unconfined aquifer. CANSAZ-3D simulates 3-D steady-state groundwater flow and transient or steady state contaminant transport. EPACMTP currently uses a simplified 2-D version of the CANSAZ-3D, and the modules are optimized for computational efficiency. Appendix D provides detailed description of the formulation and implementation of FECTUZ and CANSAZ-3D modules.

5.4.2 Data Sources for the Application of EPACMTP

Data for the site characteristics, infiltration rates, the volume and area of the landfills, and the probability distributions for the hydrogeologic parameters are obtained from a review conducted by the Hydrogeologic Inc. [64]. In this review, a number of different sources were used for the development of this site-based approach. Four of these sets were selected to derive the regional characteristics of important parameters for each sampled site:

- the OPPI survey of waste management units (EPA,1986),
- the infiltration and recharge analysis performed by OSW,
- the USGS state-by-state inventory of groundwater resources, and
- the Hydrogeologic Database for Modeling (HGDB), developed from a survey of hazardous waste field sites in the U.S.

These datasets were used in conjunction with the soil mapping database provided by the Soil Conservation Service (SCS), the data sets from the National Oceanic and Atmospheric Administration, and simulations from the HELP model [180]. The data for the hypothetical scenario in this case study were adapted from a data set provided by Hydrogeologic Inc.

5.4.3 Implementation of the Monte Carlo Method in EPACMTP

The Monte Carlo method requires that for each input parameter that has associated uncertainty or variability, a probability distribution (or a frequency distribution) be provided. The method involves the repeated generation of pseudo-random values of the uncertain input variables (drawn from the known distribution and within the range of any imposed bounds) and the application of the model using these values to generate a set of model responses or outputs (for example, the receptor well concentration, C_{RW}). These responses are then statistically analyzed to yield the probability distribution of the model output. The various steps involved in the application of a Monte Carlo simulation are:

- selection of representative cumulative probability distribution functions for the relevant input variables,
- generation of a pseudo-random number from the selected distributions, representing a possible set of values (a realization) for the input variables,
- checking whether the realization satisfies the constraints or bounds on the distribution (and resampling till a reasonable realization is obtained),
- application of the model to compute the derived inputs and outputs,
- repeated application of the above steps for a specified number of iterations, and
- statistical analysis of the series of the output values generated.

5.4.4 Uncertainty Analysis

This case study consists of a site based landfill modeling for tritium contamination at a groundwater well resulting from a hypothetical landfill unit in the souther United States. The landfill is a finite source, of 0.35 km^2 area, with a leaching duration of 20 years, and a recharge rate of 0.381 m/yr. The receptor well is located at a distance of 1635 m from the landfill. The EPACMTP model is used to simulate the radioactive decay and transport of tritium through the saturated and unsaturated zone underlying the landfill. The data used here is adapted from data provided by Hydrogeologic Inc.

Table 5.3: Probability distributions used for the EPACMTP model parameters for the uncertainty analysis

Param	Dist.[a]	Distribution Parameters[b]	units	Description
K_s	LN	$\eta = 0.343$, $\sigma = 0.989$	cm/hr	Saturated hydraulic conductivity
K	U	$\min = 1500$, $\max = 2114$	m/yr	Hydraulic conductivity
α	LN	$\eta = 0.019$, $\sigma = 0.012$	cm^{-1}	Moisture retention parameter
β	SB	$\eta = 1.409$, $\sigma = 1.629$	-	Moisture retention parameter
θ_r	SB	$\eta = 0.068$, $\sigma = 0.071$	-	Residual water content
%OM	SB	$\eta = 0.105$, $\sigma = 5.88$	-	Percentage of organic matter
d	LN	$\eta = 0.09$, $\sigma = 4.0$	cm	Particle diameter
α_L	LN	$\eta = 10.0$, $\sigma = 5.0$	m	Longitudinal dispersivity
pH	LN	$\eta = 6.5$, $\sigma = 0.75$	-	Aquifer pH
f_{oc}	SB	$\eta = 0.000432$, $\sigma = 0.0456$	-	Fractional organic carbon content
m	N	$\eta = 800$, $\sigma = 150$	m	Radial distance of observation well
θ	U	$\min = 0°$, $\max = 90°$	-	Angle off-center
B	U	$\min = 30.0$, $\max = 35.0$	m	Aquifer thickness

[a]Distribution types : LN - Lognormal, U - Uniform, N - Normal, and SB - Johnson SB.

[b]Distribution parameters: η - mean, σ - standard deviation, $\min$ - minimum value, and $\max$ - maximum value

The probability distributions of the site-specific parameters for the land fill at the Old Burial Ground are given in Table 5.3, and the values of the constant model parameters for EPACMTP are given in Table 5.4.

5.4.5 Results and Discussion

Figure 5.21 shows the uncertainty associated with the estimation of maximum tritium concentration in the ground water, as a result of the leaching from the landfill unit.

Table 5.4: Deterministic Parameters in the EPACMTP model for the case study

Param	Value	Units	Description
A_w	3.5×10^5	m^2	Area of the source
X_w	304.8	m	Length of the source
Y_w	1143.0	m	Width of the source
I_R	0.381	m/yr	Areal recharge rate
T	17.5	°C	Aquifer temperature
λ	0.0564	1/yr	Radioactive decay coefficient
t_s	20	yr	Source leaching duration
α_T	8.0	m	Transverse dispersivity
α_v	160.0	m	Vertical dispersivity
θ_s	0.45	-	Saturated water content
D_w	10.67	m	Depth to water table
S	0.0086	-	Hydraulic gradient
R_s	1.0	-	Retardation factor
ρ_b	1.65	g/cm^3	Bulk density

Figure 5.22 shows the uncertainty associated with the estimated time of occurrence of the maximum tritium concentration in the ground water at the well. The figures show the *pdf*s estimated by the Monte Carlo simulations and by the SRSM. Results of the SRSM with 350 model runs are compared with those of Monte Carlo simulations with 1000, 3000 and 5000 model runs.

The results indicate that the SRSM shows close agreement with the Monte Carlo results, while requiring much fewer number of runs. Further, there are certain other advantages in using the SRSM, that are focused in the ongoing research:

- The relationships between inputs and outputs can be easily estimated, since both

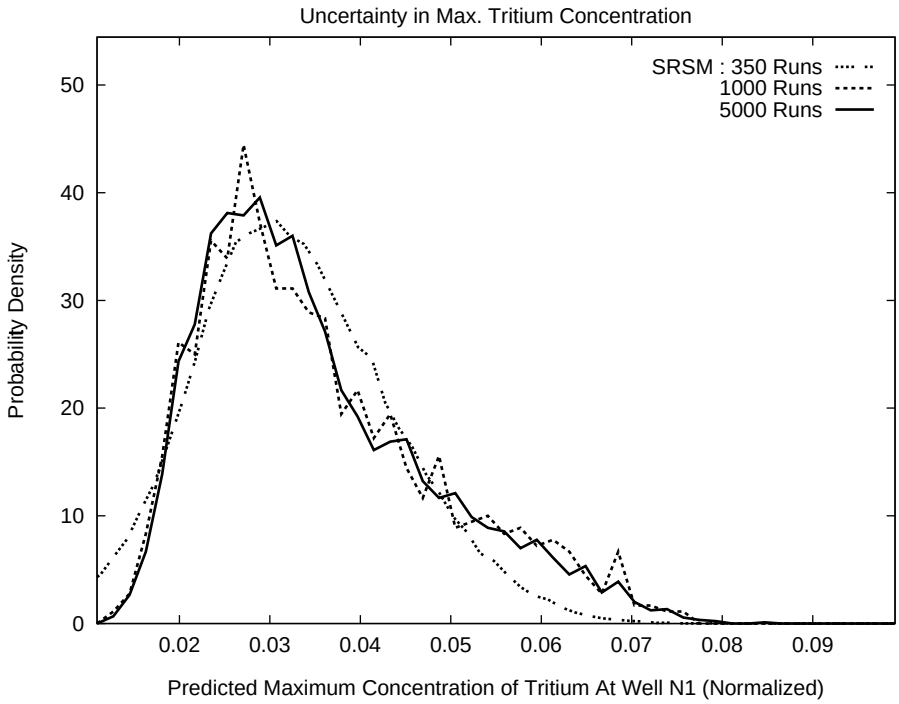


Figure 5.21: Uncertainty in estimated maximum tritium concentration in a receptor well

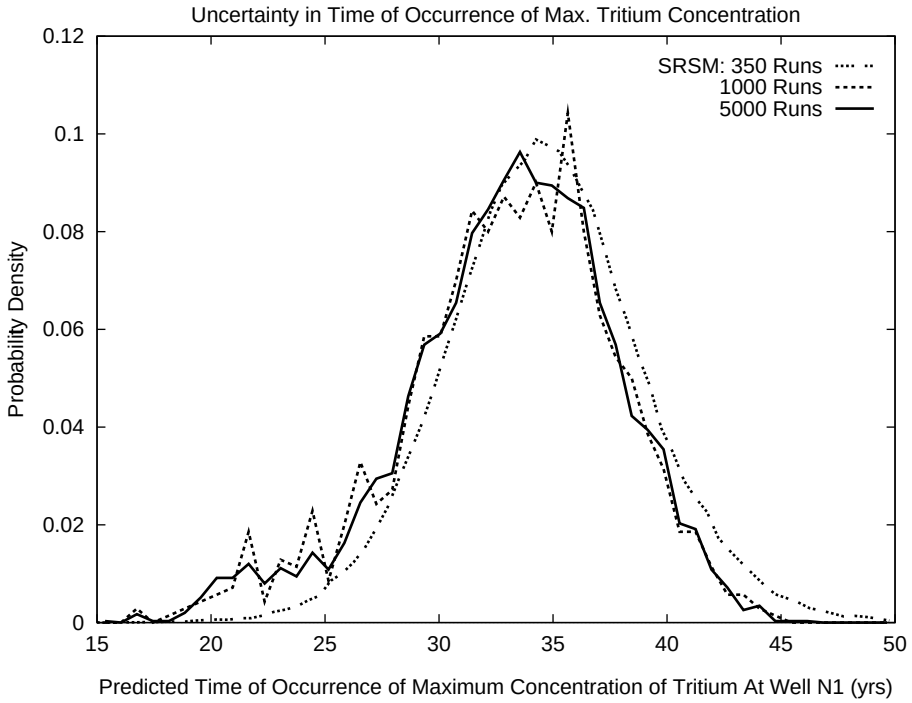


Figure 5.22: Uncertainty in estimated time of occurrence of maximum tritium concentration in a receptor well

the input and output *pdf*s are expressed as a series expansion in *srv*s. This implies that SRSM provides insight into the relative contribution of each uncertain

input each output, thus giving extensive sensitivity and uncertainty information. Conventional Monte Carlo methods do not provide such information.

- Correlation between outputs can be easily estimated since all outputs are expressed in terms of *srv*s. On the other hand, Monte Carlo methods require significantly large number of model simulations to obtain a reasonably good approximation of the correlation between outputs.
- The algebraic expressions for the outputs in terms of the *srv*s are smooth and continuous and could efficiently model the tails of the probability distributions of the outputs (which typically correspond to either high risk or worst case scenarios).

One of the limitations of the Stochastic Response Surface Method, that is a focus of the ongoing research, is as follows: the current implementation of the SRSM does not take into account constraints on the input probability distributions that make the distributions discontinuous. In short, this method assumes that the probability distributions are continuous. In the present work, such constraints are modeled as truncated probability distributions, and the results shown in Figures 5.21 and 5.22 for the SRSM are obtained in such a manner, indicating that they represent only an approximate estimate. Current work in progress involves the refinement of the SRSM so that all types of constraints can be propagated, and discontinuities in probability distributions can be fully addressed.

Chapter 6

CHARACTERIZATION AND REDUCTION OF MODEL UNCERTAINTY: AN ATMOSPHERIC PLUME MODEL STUDY

6.1 Introduction and Background

Uncertainty in the application of transport-transformation models arises not only due to uncertainty in model inputs or parameters (i.e., parametric or data uncertainty), but also due to uncertainty in model formulation (i.e., model uncertainty). As mentioned in Chapter 1, model uncertainty arises under several conditions, including the following: (a) when alternative sets of scientific or technical assumptions for developing a model exist (*Model Structure*), (b) when models are simplified for purposes of tractability (*Model Detail*), and (c) when a coarse numerical discretization is used to reduce the computation demands of the model (*Model Resolution*). Some of the major sources of model uncertainty in transport-transformation modeling are listed in Table 2.1.

When a model application involves both model and data uncertainties, it is important to identify the relative magnitudes of the uncertainties associated with data and model formulation. Such a comparison is useful in focusing resources where it is most appropriate (e.g., data gaps versus model refinement).

The approach followed here for characterizing model uncertainty here is based on the development of models corresponding to different model formulations subsequent comparison of the model results. The availability of the alternative models, ranging from simplified to more detailed, aids in evaluating the applicability of the low resolution models – if the results of the low resolution models agree closely with those of the high resolution models, then the low resolution models are preferable, since they typically

require fewer computational resources and lesser input data. Furthermore, the availability of the alternative models aids in the characterization of the model uncertainty associated with the application of those models – the bounds on model calculations corresponding to different model formulations provide an estimate of the uncertainty associated with the model formulation. For example, if the bounds are narrow the uncertainty associated with the model formulation can be considered to be relatively small.

This chapter focuses on the model uncertainties associated with photochemical air quality modeling applications. Specifically, the uncertainties associated with model resolution and assumptions of "well mixedness" are addressed here. A case study is presented involving an EPA regulatory atmospheric photochemical trajectory model, the Reactive Reactive Plume Model (RPM) [195], which is briefly described in Section 5.2, and is described in more detail in the following sections.

As mentioned in Section 5.2, the RPM describes the evolution of a photochemical plume from "point sources" of emissions, such as the stacks of refineries and of power plants. The currently used version of RPM, RPM-IV, lacks vertical resolution, as the model assumptions include uniform mixing in the vertical direction; the lack of resolution is highlighted in Figure 6.3. Here, two aspects of the model are studied: (a) the uncertainty associated with the assumption of uniform vertical concentration profile, and (b) the uncertainty associated with the choice of horizontal resolution. In this process, the results of the RPM at varying horizontal resolutions are studied first. Then a "three-dimensional version" of the RPM, termed RPM-3D, is developed by incorporating vertical resolution into the RPM-IV. The results from the RPM and the RPM-3D are subsequently compared with each other in order to characterize the uncertainty associated with the vertical resolution in the RPM.

The relevant background information on photochemical modeling, required for a detailed understanding of the RPM, is presented here. That is followed by a description of the formulation of RPM-IV and RPM-3D.

6.2 Photochemical Air Pollution Modeling

The main objective of Photochemical Air Quality Simulation Models (PAQSMs) is to establish causal relations among emission levels, meteorological conditions and ambient air quality. These are numerical models that simulate the transport and chemical transformation of pollutants in the atmosphere. Air quality can be modeled via *prognostic modeling*, that is based on the fundamental physiochemical principles governing air pollution, and as via *diagnostic modeling*, that is statistical description of observed data.

These models have been applied to local scale (e.g., plume models), to urban and regional scales (e.g., airshed models). The applications of PAQSMs range from simulation of transport and transformation of chemicals over a few kilometers over few hours (local scale) to over a few thousand kilometers over a few weeks (regional scale).

In the application of PAQSMS, one of the main pollutants of concern is ozone, as mentioned in Chapter 5.2 and Appendix C. Ozone is a "secondary pollutant" formed through nonlinear chemical reactions between precursor species, oxides of nitrogen (NO_x) and volatile organic compounds (VOCs) that are emitted by a wide variety of both anthropogenic and natural (biogenic) emissions sources. Certain meteorological conditions can significantly enhance the ambient concentrations of ozone in specific locales, leading to substantial health risks [78, 182]. The PAQSMs are employed to study the effectiveness of various alternative strategies for reductions in emissions with respect to reductions in ambient ozone levels.

6.2.1 Mathematical and Numerical Formulation of PAQSMs

Many PAQSMs are based on the fundamental description of atmospheric transport and chemical processes. The atmospheric diffusion equation, given by

$$\frac{\partial \langle c_i \rangle}{\partial t} + \overline{u}_j \frac{\partial \langle c_i \rangle}{\partial x_j} = \frac{\partial}{\partial x_j}\left(K_{jj} \frac{\partial \langle c_i \rangle}{\partial x_j} \right) + R_i \left(\langle c_1 \rangle, \ldots, \langle c_N \rangle \right) + S_i \left(\mathbf{x}, t \right) \tag{6.1}$$

is typically employed to describe the pollutant concentration in a selected control volume. This equation expresses the conservation of mass of each pollutant in a turbulent

fluid in which chemical reactions occur. In the above equation, $\langle \cdot \rangle$ represents the ensemble average, c_i is the concentration, R_i is the rate of formation of the ith species, K_{jj} is the turbulent dispersion coefficient (obtained by approximating turbulent transport with a gradient diffusion scheme), and S_i is the rate of emissions of the ith species into the control volume. The Einstein summation convention is used here with respect to the subscript j. A more detailed description of the atmospheric diffusion is given in Appendix C (Equation C.1).

The assumptions leading to the atmospheric diffusion equation above include: (a) first order closure approximation (i.e., turbulent fluxes are approximated by a gradient driven fluxes), (b) negligence of molecular diffusion compared to turbulent dispersion, (c) incompressibility of the atmosphere, and (d) approximation of ensemble average of reaction rates with the reaction rates for ensemble averages.

Air quality models can be broadly categorized into the following types:

Box models: These are the simplest of the numerical models for air quality modeling. The region to be modeled is treated as a single cell, or box, bounded by the ground at the bottom, and some upper limit to the mixing on the top, and the east-west and north-south boundaries on the sides. Here, the pollutant concentrations in a volume of air, a "box", are spatially homogeneous and instantaneously mixed. Under these conditions, the pollutant concentrations can be described by a balance among the rates they are transported in and out of the volume, their rates of emissions, and the rates at which pollutants react chemically or decay. Since these models lack spatial resolution, they cannot be used in situations where the meteorological or emissions patterns vary significantly across the modeling region.

Grid Models: Grid models employ a fixed Cartesian reference system and divide the modeling region into a two- or three-dimensional array of uniform grid cells. Horizontal dimensions of each cell usually measure on the order of kilometers, while vertical dimensions can vary from a few meters to a few hundred meters. One example of the grid models is the Urban Airshed Model (UAM), briefly described in Section 5.3 and in Appendix C.

Grid models as currently designed are generally not capable of resolving pollutant

concentrations at the microscale – that is, at scales smaller than the size of a grid cell. Thus, important local effects, such as high pollutant concentrations in the immediate vicinity of point sources, tend to become obscured by the averaging of pollutant concentrations over the grid cell volume.

There are practical and theoretical limitations to the minimum grid cell size. Increasing the number of cells increases computing and data acquisition effort and costs. In addition, choice of a grid cell implies that the input data, such as the wind flows, turbulence, and emissions, are resolved to that scale. In practice, most applications employ grid cell sizes of a few kilometers. Further, the assumptions involved in the formulation of the atmospheric diffusion equation make it applicable for gird models with horizontal resolution coarser than 2 km [137].

These models are used to assess the effectiveness of emissions reductions at a urban/regional scale. In essence, considering the domain-wide effects of emissions reductions. As these models assume uniform conditions within each grid cell, all the local scale processes are lumped into one average quantity.

Trajectory Models: The trajectory models use a moving coordinate approach to describe pollutant transport. Here, the atmospheric diffusion equation is solved in an air parcel of interest, which is assumed to travel solely with the horizontal wind. These models give results only on the path of traversed by the air parcel described, and do not permit construction of the spatial and temporal variation of concentrations over the entire region. However, their utility is in the calculation of extreme concentrations that are typically found in the plumes from point sources.

Since the numerical grids used in grid models are considerably large in size (on the length scales of a few kilometers), an understanding of the "sub-grid" effects is important in order to assess local scale effects. In order to describe the sub-grid phenomena, with specific emphasis on establishing causal relationships between emissions from point sources (such as industrial stacks) and the pollutant concentrations in the neighborhood of the point sources, plume models are employed. Plume models are trajectory models that describe short term behavior of the plume of emissions from a point source. The typical time scales range from 1 hour to 24 hours, with length scales

ranging from a few hundred meters to a few kilometers. Such detail cannot be achieved by increasing the resolution of grid-based models, since the model formulation of the grid-based models is not valid for finer resolutions (i.e., less than 2 km length scales). The formulation of the grid models and of the plume models is discussed extensively in the literature [78, 80, 144, 181, 223].

6.3 Formulation of The Reactive Plume Model (RPM)

The Reactive Plume Model, version IV (RPM-IV), is a standard regulatory model used for estimating pollutant concentrations in the atmosphere, resulting from the emissions from point sources such as industrial stacks [153, 195]. It uses either point source emission estimates or initial plume concentrations as inputs, and calculates downwind concentrations, as the plume expands. RPM-IV makes use of a diffusion/reaction equation similar in form to the atmospheric diffusion equation (Equation 6.1), but with dispersion parameters that evolve with time. This model considers a control volume that moves downwind along the plume trajectory, and solves the diffusion/reaction equation by considering the nonlinear photochemistry and diffusion of chemical species from one cell into another cell.

In this model, the plume is modeled by a set of "cells" consisting of equal masses of pollutants, by assuming a Gaussian distribution of the pollutant mass along the plume centerline. As the plume expands, the individual cells expand in volume correspondingly. The transfer of pollutant mass across cell boundaries is modeled in two phases: (a) an "entrainment" phase, where the expanding cell boundaries entrain the pollutants from other cells, and (b) a "detrainment" phase, where the pollutants diffuse across cell boundaries due to concentration gradients. Further, the cells are modeled to expand in a manner such that the amount of an inert species remains constant within each cell. The expansion of the boundaries, and the equations governing equations for pollutant concentrations within each cell are as described in the following section.

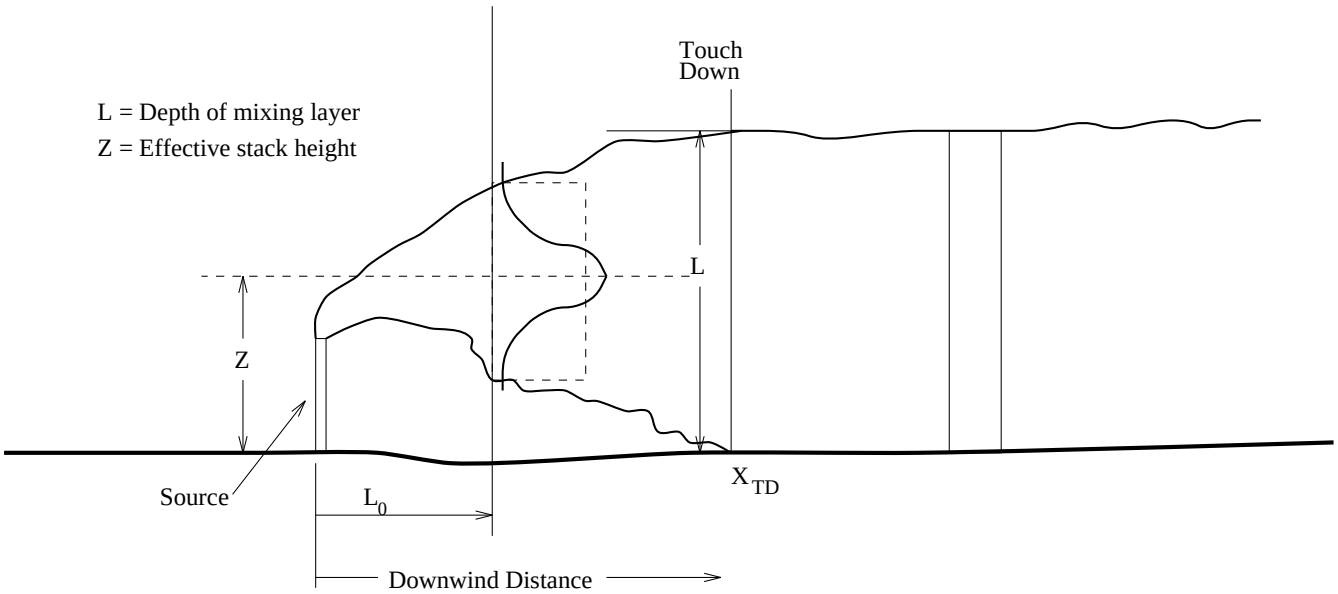


Figure 6.1: Schematic depiction of the evolution of an atmospheric plume (adapted from [195])

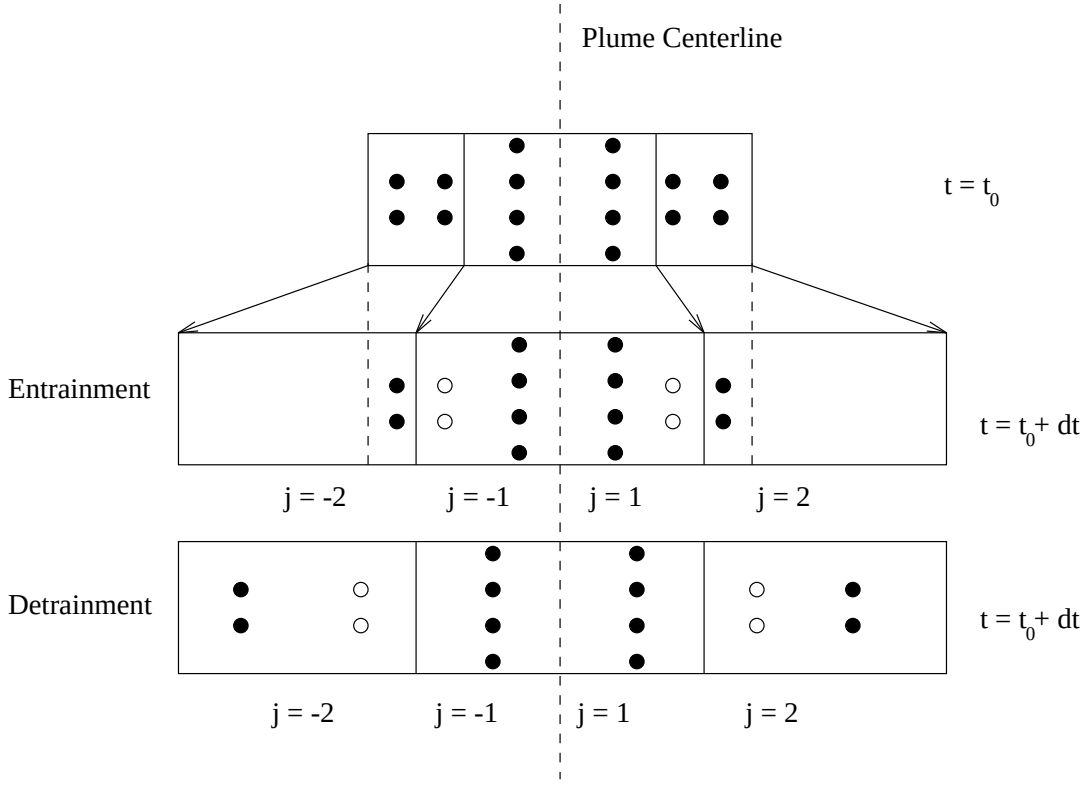


Figure 6.2: Schematic depiction of entrainment and detrainment steps simulated in the RPM (adapted from [195])

6.3.1 Initial Physical Dimensions of the Plume

The total width $W(s)$ and the total depth $H(s)$ of the photochemical plume can be estimated as functions of the horizontal and vertical dispersion parameters, as follows:

$$W(s) = f_x\,\sigma_x(s) \quad \text{and} \quad H(s) = f_y\,\sigma_z(s), \tag{6.2}$$

where s is the total distance traveled by the plume up to time t under consideration, and f_x, and f_z are selected as cutoff values of the plume boundaries. Based on the Gaussian distribution of initial plume mass [181], f_x and f_z indicate the amount of pollutant mass that is included in the model boundaries, as functions of standard normal quantiles. A typical value of $f_x = f_y = 4.0$ implies that 95.4% of the total pollutant mass is included within the plume boundaries.

For a plume described by M cells in the horizontal direction, with only one cell in the vertical, at the start of the simulation each cell contains equal amount of pollutant mass. Since the pollutant mass is assumed to follow a Gaussian distribution, the width of the cell j, w_j, can be calculated using the recursive equation

$$\frac{2}{M} = \text{erf}\left(\frac{w_j}{\sqrt{2}\sigma_y}\right) - \text{erf}\left(\frac{w_{j-1}}{\sqrt{2}\sigma_y}\right), \tag{6.3}$$

where w_j is the width of the jth cell [195].

6.3.2 RPM-IV Model Equations

Based on mass balance, the governing equation of the Reactive Plume Model is given by:

$$\frac{dc_j^i}{dt} = \underbrace{\left(\frac{dc_j^i}{dt}\right)_{\text{chem}}}_{A} - \overbrace{\left(\frac{1}{w_j}\cdot\frac{dw_j}{ds}\right)uc_j^i}^{B} - \underbrace{\left(\frac{1}{h_j}\cdot\frac{dh_j}{ds}\right)uc_j^i}_{C} + \overbrace{F_j^i}^{D} \tag{6.4}$$

Here, the term (A) describes the change in pollutant concentrations due to chemical transformation, term (B) the dilution as the jth cell expands in the horizontal, term (C) the dilution as the jth cell expands in the vertical, and term (D) the net flux of the ith contaminant into the cell.

In this equation, c_j^i denotes the concentration of chemical species i in cell j, u denotes the speed at which the cells travel downwind, and w_j and h_j denote the width and height of the jth, respectively. Additionally, F_j^i is the net flux into the cell, and is given by $F_j^i = E_j^i - D_j^i$, where E_j^i is the entrainment as the cell expands, and D_j^i is the detrainment due to the diffusion of the species across the boundaries of the cell. Figure 6.2 describes the process of entrainment and detrainment in detail. Expressions for entrainment and detrainment are given by the following equations:

$$E_j^i = \frac{1}{y_j - y_{j-1}} \left\{ \left(\frac{dy_j}{dt} \right) c_{j+1}^i - \left(\frac{dy_{j-1}}{dt} \right) c_j^i \right\} \tag{6.5}$$

$$D_j^i = \frac{\partial}{\partial y} \left(K \frac{\partial c}{\partial y} \right) = \frac{2}{y_j - y_{j-1}} \left\{ K_j \left(\frac{c_{j+1}^i - c_j^i}{y_{j+1} - y_{j-1}} \right) - K_{j-1} \left(\frac{c_j^i - c_{j-1}^i}{y_j - y_{j-2}} \right) \right\} \tag{6.6}$$

where y_j is the distance from the plume centerline to the far most side of the cell j, in the horizontal. An additional condition is imposed on the expansion of the cells: the cells are allowed to expand in a way such that the amount of an inert species remains constant within each cell, i.e., no net transfer of an inert species occurs from a cell. This condition results in the equation $E_j^I - D_j^I = 0$, for the expansion of the cells, as the plume travels downwind. This condition leads to a recursive relationship to compute K_j, as follows:

$$\left(\frac{dy_j}{dt} \right) C_{j+1}^I - \left(\frac{dy_{j-1}}{dt} \right) C_j^I = 2 \left\{ K_j \left(\frac{C_{j+1}^I - C_j^I}{y_{j+1} - y_{j-1}} \right) - K_{j-1} \left(\frac{C_j^I - C_{j-1}^I}{y_j - y_{j-2}} \right) \right\} \tag{6.7}$$

where C_j^I is the concentration of an inert species I in the jth cell. Further, the condition of zero concentration gradient at the plume centerline results in the equation:

$$K_1 = \left(\frac{C_2^I}{2} \frac{y_2 - y_0}{C_2^I - C_1^I} \right) \frac{dy_1}{dt} \tag{6.8}$$

The values of K_j can then be recursively obtained using the above two equations. Once K_js are computed, Equations 6.4, 6.5, and 6.6 can be used to compute the concentration in each cell at different times.

In RPM-IV, the nonlinear plume photochemistry is modeled through the carbon-bond mechanism, version IV (CB-IV) [216]. CB-IV describes the complex non-linear gas-phase atmospheric photochemistry through a set of 95 reactions among 35 surrogate

chemical species corresponding to organic bonds/functional groups [216]. This mechanism lumps similarly bonded carbon atoms, resulting in a condensed mechanism of that is used widely in the regulatory photochemical models. A brief description of the CB-IV mechanism, and the chemical reactions considered, is presented in Appendix C.

6.3.3 Limitations of the RPM-IV

Figure 6.1 presents a schematic depiction of the physical structure of a plume as it evolves from a stack, and travels downwind. The assumption of uniform vertical concentration in the RPM (hence, only one "well mixed" cell in the vertical) is valid only at distances much greater than X_{TD} from the point source. This poses a limitation, since the typical touchdown distance could vary between a few hundred meters to a few kilometers, depending on the meteorological conditions, the height of the stack, and the exit velocity at the plume source. When the value of X_{TD} is large, the well-mixedness assumption is not valid at the vicinity of the plume, thus lessening the main advantage of using a plume model to study the "sub-grid" scale, local phenomena. In such cases, if the RPM is used with the uniform vertical mixing assumption, significant "errors" could result in the model calculations.

In order to address this limitation of the RPM-IV, a corresponding "three-dimensional" version of the RPM, called RPM-3D, is developed as part of this work. This RPM-3D simulates the evolution of a plume by dividing the plume cross-section into rectangular regions consisting of equal initial pollutant mass; the plume cross section is divided into columns of cells, as in the case of RPM-IV, and each column is further subdivided into rows.

Figure 6.3 illustrates the physical structure of the plume cross section in the RPM-IV and RPM-3D. In the figure, R1 is the region containing the cell closest to the plume centerline, R2 contains the ground-level cell along the plume centerline in the horizontal, and R3 contains the centerline cells with respect to both the horizontal and the vertical. Here, R1 corresponds to the representation used by RPM-IV, whereas R2 and R3 correspond to representations used by RPM-3D. In typical cases, the extreme pollutant concentration occurs along the plume centerline, and the pollutant concentration

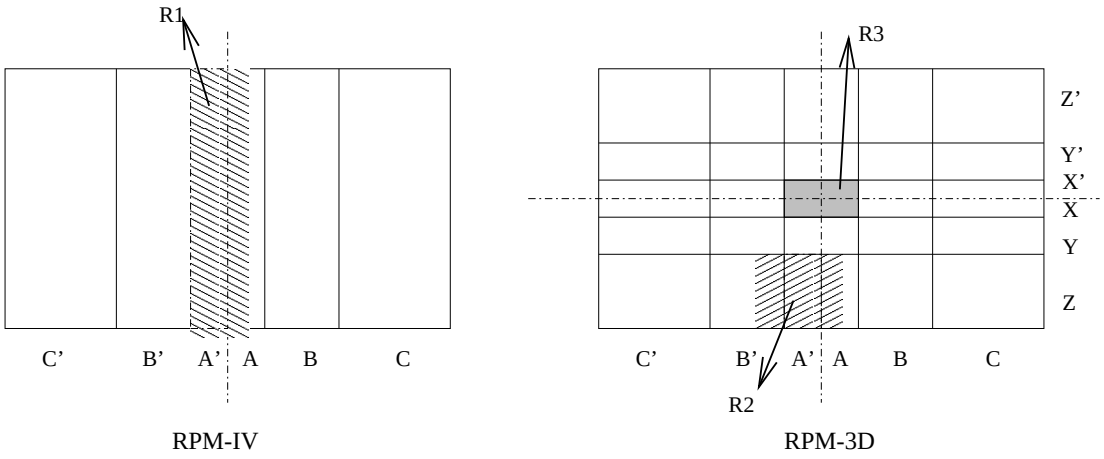


Figure 6.3: Discretization of the plume cross section by RPM-IV (left) and RPM-3D (right)

decreases as the distance of the cells from the centerline increases. Hence, the extreme pollutant concentration (e.g., concentration of ozone) calculated by the RPM-IV is an average over the region R1. This provides a concentration for the estimation of exposure to ozone. However, as shown in the figure, the region R2 corresponds to the most likely ground level maximum ozone concentration (typically covering the region from the ground to up to less than a hundred meters), whereas the region R3 corresponds to the maximum ozone level concentration at the plume centerline. Clearly, the RPM-3D provides estimates of pollutant concentrations at a detailed spatial resolution that is appropriate for estimating human exposure.

6.4 Formulation of the RPM-3D

The equations describing the transport of chemical species in RPM-IV are modified accordingly to accommodate the three dimensional description of the plume. Thus, Equation 6.4 is modified as follows:

$$\frac{dc^i_{j,k}}{dt} = \left(\frac{dc^i_{j,k}}{dt}\right)_{\text{chem}} - \left(\frac{1}{w_{j,k}} \cdot \frac{dw_{j,k}}{ds}\right) uc^i_{j,k} - \left(\frac{1}{h_{j,k}} \cdot \frac{dh_{j,k}}{ds}\right) uc^i_{j,k} + F^i_{Y,j,k} + F^i_{Z,j,k} \tag{6.9}$$

where $c^i_{j,k}$ is the concentration of the ith species in the cell (j,k); j denotes the cell

number in the horizontal direction, and k denotes the cell number in the vertical. $F^i_{Y,j,k}$ and $F^i_{Z,j,k}$ denote the fluxes along the horizontal and vertical directions respectively. The equations for the calculation of the fluxes, are similar to Equations 6.5 and 6.6, and are as follows:

$$
\begin{aligned}
F^i_{Y,j,k} &= E^i_{Y,j,k} - D^i_{Y,j,k} \\
F^i_{Z,j,k} &= E^i_{Z,j,k} - D^i_{Z,j,k} \\
E^i_{Y,j,k} &= \frac{1}{y_{j,k} - y_{j-1,k}} \left\{ \left(\frac{dy_{j,k}}{dt} \right) c^i_{j+1,k} - \left(\frac{dy_{j-1,k}}{dt} \right) c^i_{j,k} \right\} \\
D^i_{Y,j,k} &= \frac{2}{y_{j,k} - y_{j-1,k}} \left\{ K_{Y,j,k} \left(\frac{c^i_{j+1,k} - c^i_{j,k}}{y_{j+1,k} - y_{j-1,k}} \right) - K_{Y,j-1,k} \left(\frac{c^i_{j,k} - c^i_{j-1,k}}{y_{j,k} - y_{j-2,k}} \right) \right\} \\
E^i_{Z,j,k} &= \frac{1}{z_{j,k} - z_{j,k-1}} \left\{ \left(\frac{dz_{j,k}}{dt} \right) c^i_{j,k+1} - \left(\frac{dz_{j,k-1}}{dt} \right) c^i_{j,k} \right\} \\
D^i_{Z,j,k} &= \frac{2}{z_{j,k} - z_{j,k-1}} \left\{ K_{Z,j,k} \left(\frac{c^i_{j,k+1} - c^i_{j,k}}{z_{j,k+1} - z_{j,k-1}} \right) - K_{Z,j,k-1} \left(\frac{c^i_{j,k} - c^i_{j,k-1}}{z_{j,k} - z_{j,k-2}} \right) \right\}
\end{aligned}
\tag{6.10}
$$

where $y_{j,k}$, and $z_{j,k}$ denote the distances from the plume centerlines to the far most sides, in the horizontal and vertical directions respectively. $K_{Z,j,k}$ and $K_{Y,j,k}$ are solved using recursive relationships similar to Equation 6.7: the relationships are obtained from the following conditions: (a) $F^I_{Y,j,k}$ and $F^I_{Z,j,k}$ are equal to zero for an inert species I, (b) there is a zero concentration gradient at the plume centerline in both horizontal and vertical directions.

The RPM-IV uses the Gear algorithm [76] to solve the stiff set of equations describing the chemical reactions. In this method, the step size for each successive calculation is adjusted according to the local truncation errors. The same approach is followed in the RPM-3D.

6.5 Case Studies

The RPM-IV and RPM-3D are both applied to two significantly different case studies. An attempt is made to understand the extent of the uncertainty associated with the model assumptions and model resolution, by comparing the predictions of these two models.

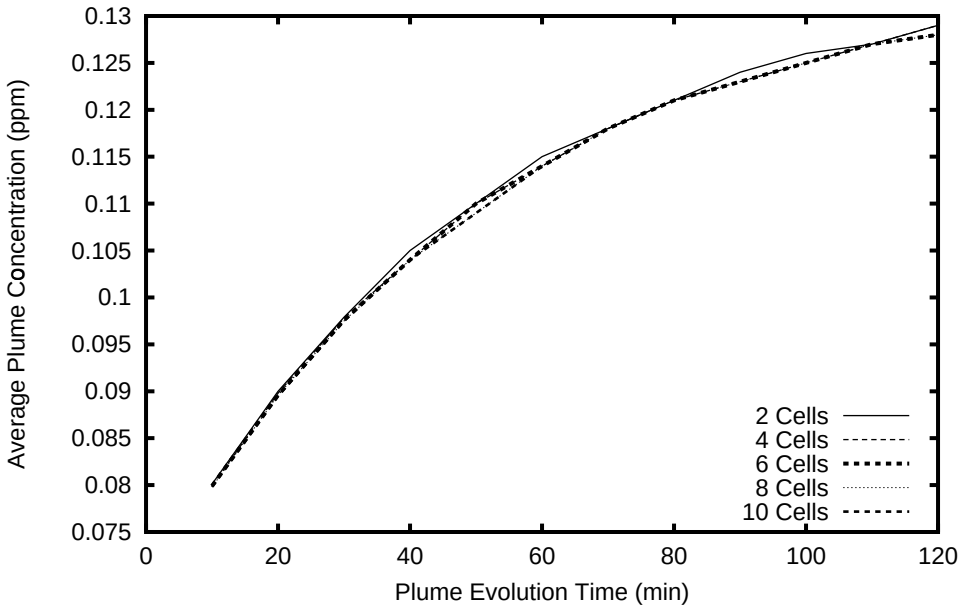


Figure 6.4: Dependence of plume average O_3 concentration on horizontal resolution of the RPM-IV in a VOC dominant regime

The first case study uses emission estimates and field data measured near Marathon Oil Refinery at Robinson, Illinois, during June and July, 1977 [183]. This case study involves a primarily VOC dominant source (i.e., the volatile organic compounds emissions are significantly higher than the NO/NO_2 emissions). The meteorological data and initial plume concentrations were obtained from the work of Georgopoulos and Roy [79], which was based on the measurements reported by Sexton et al. [183].

Figure 6.4 shows the average ozone concentrations in the plume, calculated by RPM-IV, as a function of the plume evolution time for different horizontal resolutions, ranging from two cells to ten cells in the horizontal. Similarly, Figure 6.5 shows the average plume concentrations calculated by RPM-3D for different vertical resolutions with four cells in the horizontal. The calculations from RPM-IV, which has no vertical resolution, and RPM-3D with vertical resolution corresponding to two cells, four cells, and six cells, are shown in the figure. The results indicate that the different vertical and horizontal resolutions produce similar estimates of the average plume concentrations. Hence, it is sufficient to use a low-resolution two-dimensional model, for studying the plume average concentrations in such cases.

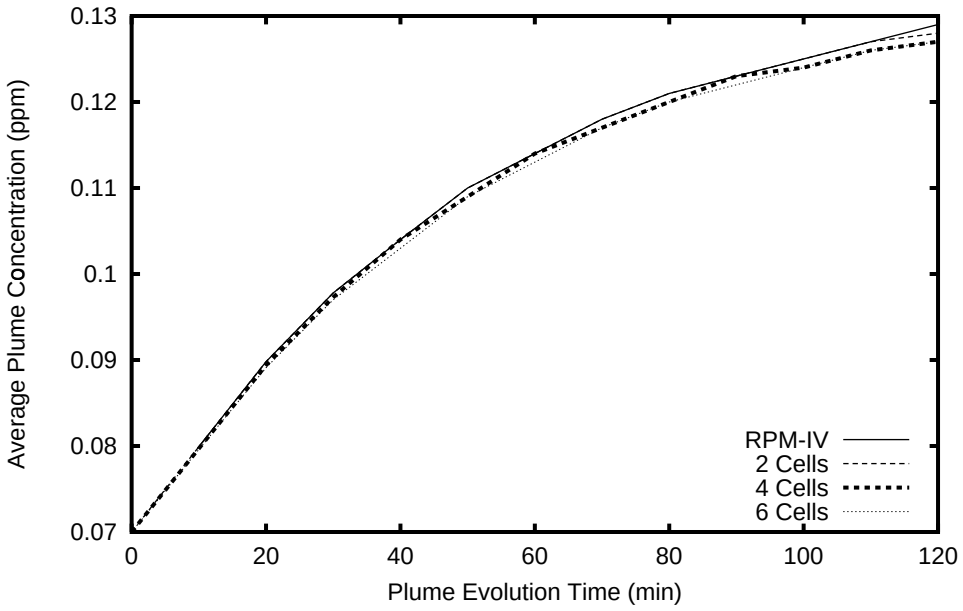


Figure 6.5: Dependence of plume average O_3 concentration on vertical resolution of the RPM-3D in a VOC dominant regime (4 horizontal cells)

Figure 6.6 shows the ozone concentrations within each horizontal cell calculated by RPM-IV when the plume cross section is divided into six cells. The plots indicate the average ozone concentrations in the different horizontal cells (indicated by "A", "B", and "C" in Figure 6.3 for RPM-IV). Similarly, Figure 6.7 shows the vertical profile of the ozone concentrations calculated by RPM-3D for the centerline horizontal cell when the plume is divided into four horizontal cells and each cell is in turn divided into six vertical cells. The plots indicate the average of the vertical cells along the horizontal centerline, and individual vertical cells at the horizontal line. The cells are illustrated in Figure 6.3. The horizontal centerline cell is given by "A", and the corresponding vertical slices are "AZ" (outermost), "AY" (middle), and "AX" (centerline). Examining the results, it is clear that in a NO_x dominant regime, this would clearly produce an under-estimation of the ground level ozone levels. Examining the results, it is clear that by ignoring the vertical structure of a plume, the RPM-IV approximates the highest possible ground level concentration by the average in the vertical. Hence, in a VOC dominant regime, this would clearly produce an over-estimation of the ground level ozone levels.

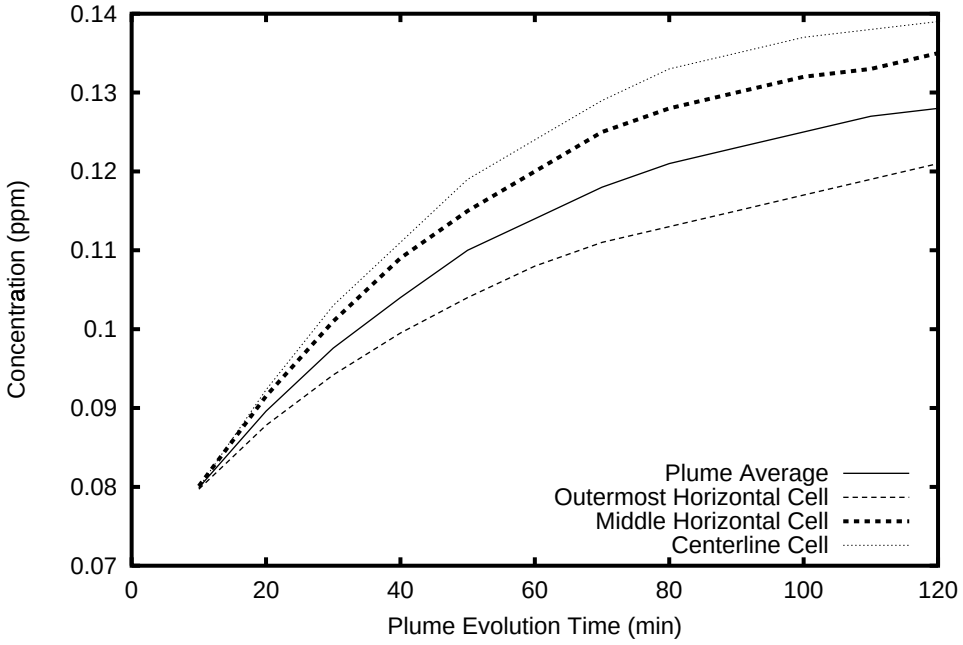


Figure 6.6: Horizontal profile of the O_3 concentration in the plume in a VOC dominant regime (6 horizontal cells, RPM-IV)

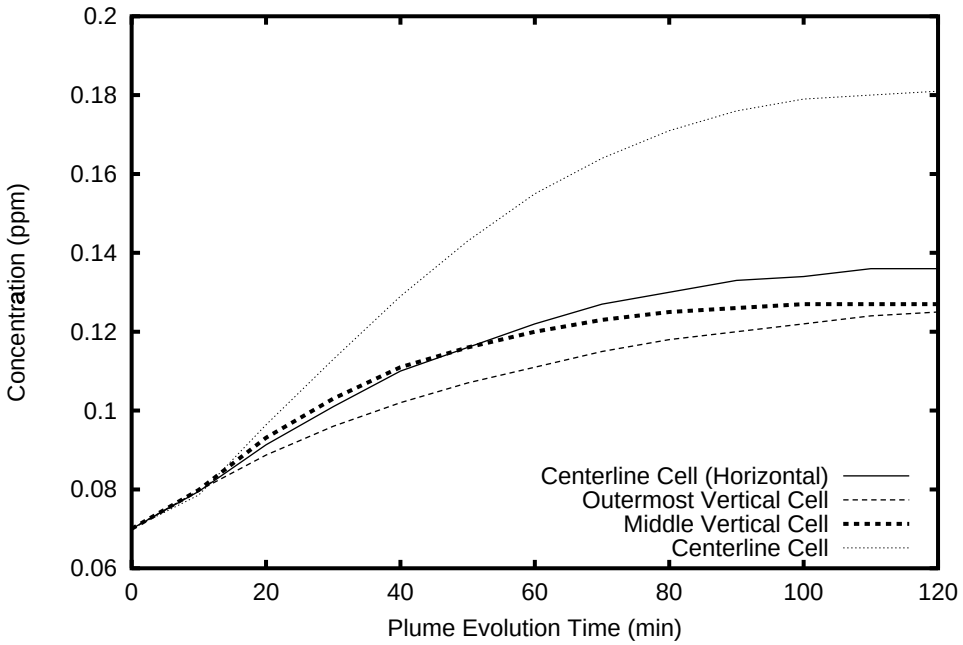


Figure 6.7: Vertical profile of the O_3 concentration at the plume centerline (w.r.t. horizontal) in a VOC dominant regime (4 horizontal cells and 6 vertical cells, RPM-3D)

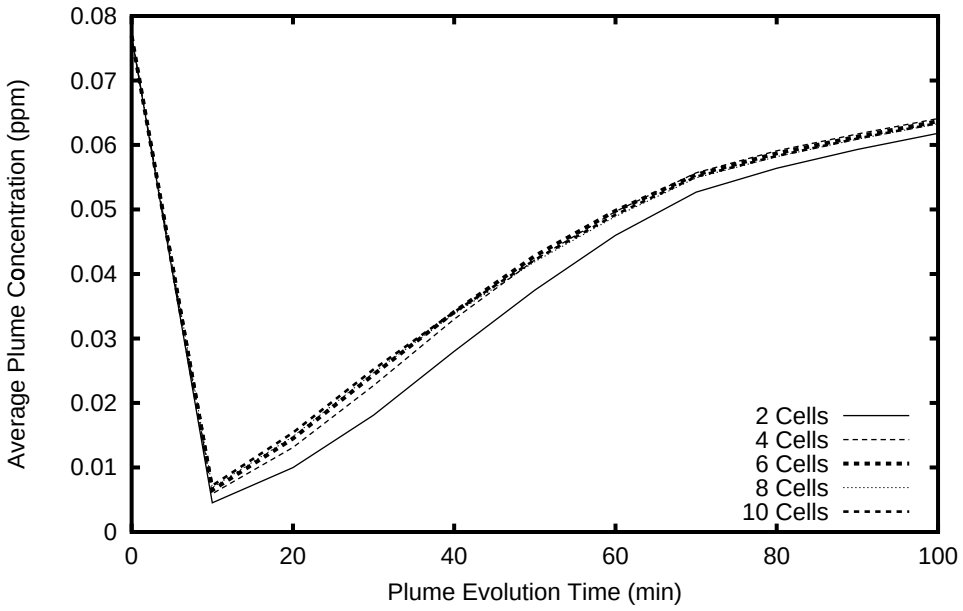


Figure 6.8: Dependence of plume average O_3 concentration on horizontal resolution of the RPM-IV in a NO_x dominant regime

The second case study considers a single point source in Mercer County, New Jersey. The simulation is performed for the period of July 7, 1988, 8:00 am to 10:00 am, corresponding to the occurrence of a severe ozone episode in the region. This case study involves a NO/NO_2 dominant regime. Ambient pollutant concentrations were obtained from the results of an Urban Airshed Model (UAM-IV) simulation for the Philadelphia/New Jersey modeling domain for the time period of July 6 to July 8, 1988, described by Georgopoulos et al. [81]. The meteorological inputs were obtained from the model inputs of the UAM-IV simulation for the modeling domain [81]. The model results are analyzed in the same manner as described in the earlier case study involving a VOC dominant regime.

Figure 6.8 shows the average ozone concentrations in the plume, calculated by RPM-IV, as a function of the plume evolution time for different horizontal resolutions, ranging from two cells to ten cells in the horizontal. Similarly, Figure 6.9 shows the average plume concentrations calculated by RPM-3D for different vertical resolutions with four cells in the horizontal. The calculations from RPM-IV, which has no vertical resolution,

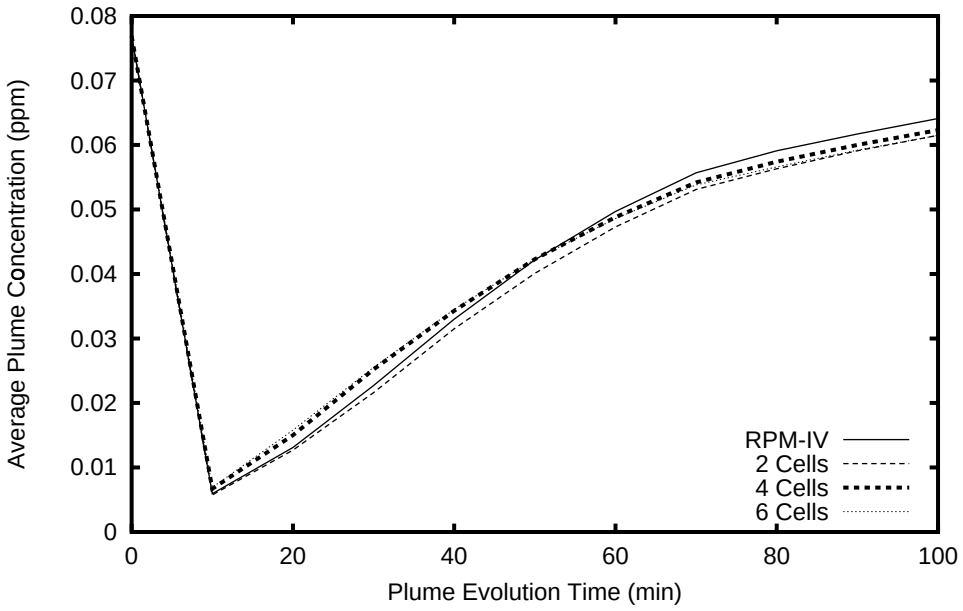


Figure 6.9: Dependence of plume average O_3 concentration on vertical resolution of the RPM-3D in a NO_x dominant regime

and RPM-3D with vertical resolution corresponding to two cells, four cells, and six cells, are shown in the figure. The results indicate that the different vertical and horizontal resolutions produce similar estimates of the average plume concentrations. Hence, it is sufficient to use a low-resolution two-dimensional model, for studying the plume average concentrations in such cases.

Figure 6.10 shows the ozone concentrations within each horizontal cell calculated by RPM-IV when the plume cross section is divided into six cells. Similarly, Figure 6.11 shows the vertical profile of the ozone concentrations calculated by RPM-3D for the centerline horizontal cell when the plume is divided into four horizontal cells and each cell is in turn divided into six vertical cells. Examining the results, it is clear that in a NO_x dominant regime, this would clearly produce an under-estimation of the ground level ozone levels.

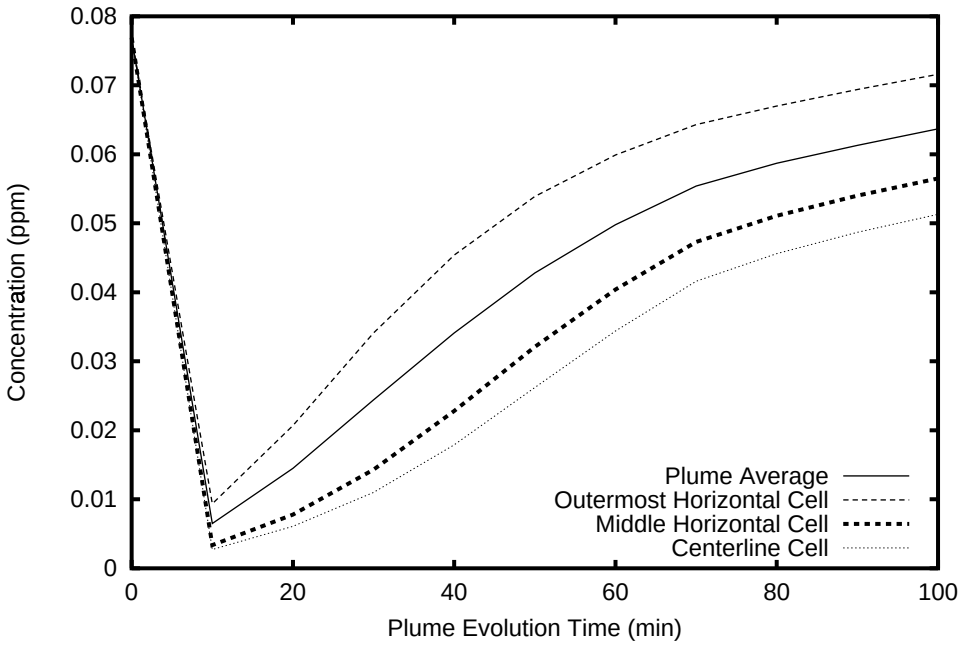


Figure 6.10: Horizontal profile of the O_3 concentration in the plume in a NO_x dominant regime (6 horizontal cells, RPM-IV)

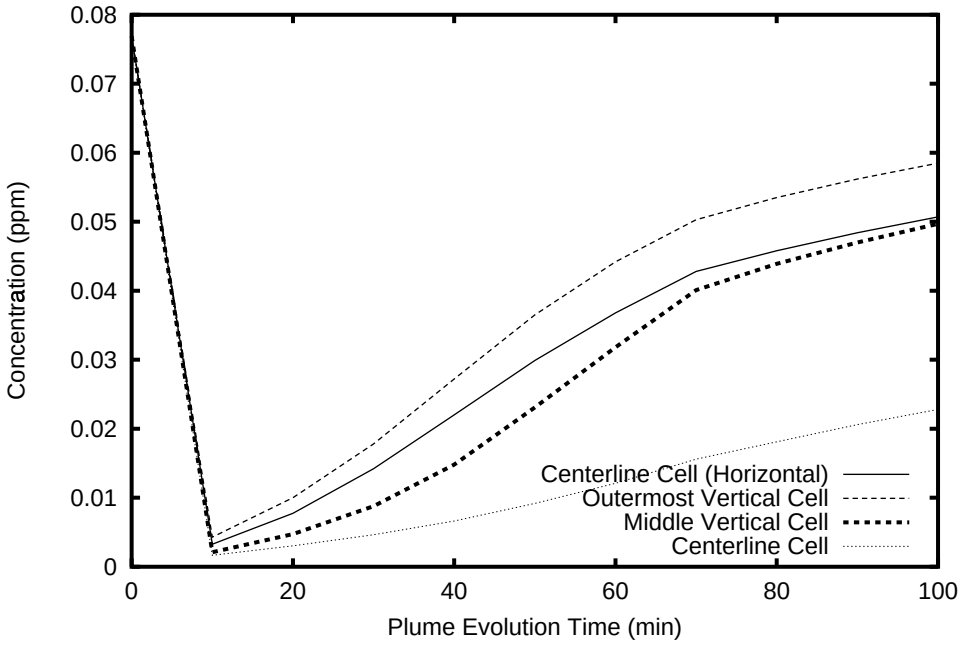


Figure 6.11: Vertical profile of the O_3 concentration at the plume centerline (w.r.t. horizontal) in a VOC limited (NO_x dominant) regime (4 horizontal cells and 6 vertical cells, RPM-3D)

6.6 Discussion

The comparison of calculations of the two- and three-dimensional versions of the RPM indicate that the choice of model resolution is dependent on the output metric(s) considered. The two case studies presented show that the two-dimensional version is preferable to the three-dimensional version when the average plume concentrations are calculated, since both the two- and three-dimensional versions give close estimates. However, when the output metrics require a higher resolution (for example, the ground level ozone concentrations are required when calculating human exposure to ozone), the three dimensional version is to be preferred. Further, the uncertainty associated with the calculation of average plume concentrations is relatively small.

When evaluating the uncertainties associated with model formulation, the approach suggested here, construction of a hierarchy of models with increasing detail, is recommended. The comparison of the model results can provide insight into what level of detail is sufficient to produce results similar to more detailed models. Such knowledge is very useful in building an "optimal model", one that produces outputs similar to more detailed models, but requires much less detailed inputs and much fewer computational resources. Further, results from models with varying detail provide an estimate of the range of model calculations, thus helping in characterizing the uncertainty associated with model formulation and detail.

Chapter 7
CONCLUSIONS AND DISCUSSION

The application of transport-transformation models involves significant uncertainties that may have implications on the confidence in model estimates and on their use in the assessment of possible alternative scenarios in a regulatory framework, for example, with respect to control strategy selection for decreasing pollutant levels. Hence, it is important to address these uncertainties. However, the main limitations in performing comprehensive uncertainty analyses of transport-transformation models are the associated cost and effort, and the increased data needs for characterizing uncertainties in inputs. Clearly, there is a need for methods that can reduce the cost and effort associated with uncertainty analysis of transport-transformation models.

7.1 Development and Application of the SRSM

One of the main contributions of this research is the development of computationally efficient methods for uncertainty propagation, specifically, the development of the Stochastic Response Surface Method (SRSM) (Chapter 3). The SRSM approximates the model inputs and the outputs through a series of "well behaved" standard random variables; the series expansions of the outputs contain unknown coefficients which are calculated by a method that uses the results of a limited number of model simulations. The SRSM is based on the Stochastic Finite Element Method [83] and the Deterministic Equivalent Modeling Method [198].

The SRSM has been implemented as a modular and readily portable stand-alone tool that can be used by researchers as a black-box tool, without requiring the mathematical details of the SRSM. In fact, the SRSM has been implemented as a web-based

tool, that can be accessed through a web browser, as shown towards the end of Chapter 3. This implementation includes modules for (a) the transformation of non-standard distributions into functions of the *srv*s, (b) the construction of polynomial chaos approximations for the model outputs, (c) the identification of sample points for model runs, (d) the estimation of the unknown coefficients in the approximations, and (e) the subsequent calculation of the statistical properties of the model outputs. Further, mathematical formulae have been developed and presented for the extension of this method for dealing with arbitrary empirical distributions (with correlated and uncorrelated inputs), and for the estimation of correlations between two model outputs, and between a model input and a model output.

The development and the application of the SRSM meets the need for fast and accurate estimates of the uncertainties associated with the model outputs, and also the need for identification of inputs that contribute to the output uncertainties the most. The main impact of this research is the facilitation of fast uncertainty analyses of complex, computationally intensive models, and the estimation of the relative contribution of each model input to the uncertainties in the outputs.

The SRSM has been evaluated for case studies involving the following models:

(a) a biological model (a Physiologically Based PharmacoKinetic (PBPK) model, i.e., a lumped system described by a set of stiff ODEs) describing the uptake and metabolism of perchloroethylene in humans,

(b) a two-dimensional atmospheric photochemical plume model that calculates pollutant concentrations downwind of point sources, the Reactive Plume Model (RPM),

(c) a three-dimensional urban/regional scale air quality model, the Urban Airshed Model (UAM), version IV, and

(d) a one-dimensional ground water model with discontinuous probability density functions, the EPA's Composite Model for leachate Migration and Transformation Products (EPACMTP).

In the first two cases, the performance of the SRSM was significantly superior to both the Monte Carlo and the Latin Hypercube Sampling methods - the SRSM application

required up to an order of magnitude fewer simulations for estimating the uncertainties in the model outputs. In the third case study, involving the Urban Airshed Model, it was impractical to perform hundreds of Monte Carlo simulations, whereas, the SRSM required just twenty (20) simulations to estimate the output uncertainties: this demonstrates the potential of the SRSM to address problems that are beyond the scope of conventional methods (in terms of computational and time limitations). In the last case study involving the ground water model, EPACMTP, the SRSM, as before, required significantly fewer model runs compared to the Monte Carlo method.

7.2 Development and Application of the SRSM-ADIFOR

In order to further improve the computational efficiency of the SRSM, this method was coupled with a state of the art automated sensitivity analysis method, ADIFOR (Automatic DIfferentiation of FORtran) [17]. ADIFOR produces fast and accurate estimates of first order partial derivatives of model outputs with respect to model inputs or any intermediate quantities in the model code. ADIFOR accomplishes this by rewriting the model code and producing code that calculates first order partial derivatives. Hence, the coupling of the SRSM and the ADIFOR, where the series expansions of the SRSM are used in conjunction with outputs and partial derivatives from the ADIFOR generated code, provides the potential to substantially reduce the computer demands.

The coupled method, SRSM-ADIFOR, was applied to two of the case studies used for evaluating SRSM: a human PBPK model, and the Reactive Plume Model (RPM). The case studies indicate substantial computer time savings; up to two orders of magnitude reductions in the required number of model simulations compared to "conventional" methods were achieved. This demonstrates the advantages of combining sensitivity analysis methods with uncertainty propagation methods. The application of the SRSM-ADIFOR involves processing the original model code using ADIFOR, and subsequent use of the derivative code for uncertainty analysis. For this reason, developing a "black-box tool" for using the combined SRSM-ADIFOR approach (as was done for the "stand-alone" SRSM) is not possible.

7.3 Consideration of Uncertainties Beyond Parametric/Data Uncertainty

In addition to the uncertainties associated with model inputs, there are often uncertainties associated with model formulation and structure. Model uncertainty arises mainly when competing models, ranging from simplified to more detailed models, exist. When a model application involves both model and data uncertainties, it is important to identify the relative magnitudes of the uncertainties associated with data and model formulation. Such a comparison is useful in focusing resources where it is most appropriate (e.g., filling data gaps versus refining a model).

The case study involving the development of a three dimensional version of the Reactive Plume Model (RPM-3D, Chapter 6) examines the issues associated with the characterization of uncertainties associated with model formulation. Lack of sufficient model resolution could in fact result in model predictions that may either significantly over-estimate or under-estimate the quantities that are important for decision making. The case study stresses the need for comprehensive hierarchical models, for the systematic evaluation of the simplified models.

7.4 Directions for Future Work

7.4.1 Improvements to the Stochastic Response Surface Method

In the present form, the SRSM includes modules for the transformation of probability distributions commonly used to represent uncertainties in the inputs of environmental and biological models. The transformations are listed in Table 3.1. However, this list is not extensive, since model inputs can follow one of several other distributions. In order to extend this list, the methods outlined for transforming random variables that follow empirical distributions (Section 3.2.3) and for transforming random variables that are correlated (Section 3.2.4) can be incorporated into the SRSM tool.

The SRSM calculates probability density functions (pdfs) to quantify the uncertainties in model outputs. Additionally, several other statistical metrics, such as the cumulative density function, moments of the output metrics, and percentile estimates, can be calculated. This can be accomplished by following the techniques outlined in

Section 3.5.

Finally, the methods for calculating correlations between outputs and between an input and an output, presented in Section 3.5, can also be incorporated into the SRSM tool. This would facilitate the identification of individual contributions of model inputs to the uncertainties in model outputs, in a more rigorous manner than that presented in relation to the case study in Chapter 5 (Table 5.2).

These improvements could further enhance the already wide range of applicability of the SRSM. Furthermore, the same improvements translate into the improvements in the SRSM-ADIFOR method, and even in the wed-based SRSM tool.

7.4.2 Further evaluation of the SRSM and the SRSM-ADIFOR

Further rigorous, statistically based evaluation of the SRSM and the SRSM-ADIFOR, that would be based not only on the *pdf*s of model outputs, but also on the moments, cumulative densities, percentiles and correlation coefficients, would aid in the wider acceptance of this method. Additionally, application of the SRSM and the SRSM-ADIFOR to more complex models could also aid in the identification of areas where these methods could be further improved.

7.4.3 Addressing uncertainty propagation under constraints

The case study for the evaluation of the SRSM with a groundwater model showed certain limitations in the application of the SRSM to models with discontinuous probability distributions. Transport-transformation models sometimes have constraints on the values the model inputs can assume; often, the constraints are based on the values of other model inputs. For instance, the sample value of random input $\boldsymbol{x}$ may affect the range from which another random input $\boldsymbol{y}$ is sampled. While many constraints are defined in terms of joint *pdf*s, in some cases, the constraints could follow a discontinuous pattern. For example, the diameter and porosity of a particle may not assume certain combinations in the EPACMTP model (Chapter 5).

The Monte Carlo method can address such constraints by following the rules listed below:

- generate sample point by randomly sampling all inputs from their respective distributions ignoring constraints,
- ignore the sample point if any constraints are not satisfied, and repeat the above step till an acceptable sample point is obtained,
- run the model at the accepted sample point,
- repeat the above procedure till required number of sample points are obtained, and
- statistically analyze the model outputs corresponding to all the sample points.

It must be noted that samples are drawn from the given *pdf*s, and they also obey the constraints. That is an advantage of using Monte Carlo methods in a brute-force manner.

In the SRSM, such an approach does not appear to be readily applicable, because the inputs are represented as algebraic functions of the *srv*s. This means that the inputs have *pdf*s that are continuous and well behaved (as opposed to the actual *pdf*s which are discontinuous). One possible approach that could be followed is suggested here:*

- express inputs in terms of *srv*s, ignoring constraints,
- approximate the outputs using the polynomial chaos expansions, and
- generate the sample points for the *srv*s without considering the constraints,
- estimate the unknown coefficients in the series expansion through regression on the model outputs at these sample points, and
- calculate the statistical properties of the model outputs from the *srv*s using the methods listed in Chapter 3.5 and in the process rejecting the combinations of *srv*s that do not obey the constraints.

One potential focus of the future efforts could be on the evaluation of this approach, in addition to the identification of other techniques to address uncertainty propagation under constraints.

*This approach has not been tested, and its applicability is still unknown.

7.4.4 Random processes and random fields

The SRSM, in its present form addresses only random variables, i.e., random quantities that do not vary with time or space. From the perspective of the uncertainties that occur in the nature, the random variables are analogous to points in a multi-dimensional space. Random processes and random fields are generalizations of random variables to multi-dimensional spaces.

A random process can be considered as a function of time in the random space, as

$$\boldsymbol{y}(t, \zeta)$$

where t denotes the time dimension, and ζ is used to denote the randomness. Here, a particular realization of $\boldsymbol{y}$ can be considered as a deterministic function $\boldsymbol{y}(t, \zeta_i)$, where ζ_i denotes one realization of the various possible functions that the random process $\boldsymbol{y}$ can assume. Further, at a given time, t_i, $\boldsymbol{y}$ reduces to a random variable. The relationship between random processes, random variables, deterministic functions, and deterministic variables can be summarized as follows [160]:

- $\boldsymbol{x}(t, \zeta)$ is a stochastic process, or a family of time functions,
- $\boldsymbol{x}(t_i, \zeta)$ is a random variable,
- $\boldsymbol{x}(t, \zeta_i)$ is a single time function, and
- $\boldsymbol{x}(t_i, \zeta_i)$ is a single number.

Similarly, random fields are random functions of spatial coordinates and time. Random processes and random fields are common in environmental and biological systems: for example, the flow rates or emissions from a point source are random processes, and emissions from area sources, wind patterns over a domain are examples of random fields. Further research could focus on extending the SRSM so that it can address uncertainty propagation involving model inputs defined by random processes and random fields.

7.4.5 Uncertainties Associated with Evaluation Data

Characterization of uncertainty associated with the evaluation data is an important component of a comprehensive uncertainty analysis. The uncertainties in the evaluation

data may sometimes in fact impact the selection of an appropriate model from a set of alternative models. Comparisons of parametric and evaluation data uncertainties can provide insight into where available resources must be focused (input data versus evaluation data). Various techniques (such as those used in the spatio-temporal analysis of environmental data, e.g., Vyas and Christakos [36, 210]) could be explored for their potential to improve the characterization of uncertainties in environmental databases that are used to evaluate transport-transformation models.

References

[1] A. E. Abdin and J. J. Kaluarachchi. Stochastic analysis of three-phase flow in heterogeneous porous media. 1. Spectra/perturbation approach. *Water Resources Research.*, 33(7):1549–1558, 1997.

[2] A. E. Abdin and J. J. Kaluarachchi. Stochastic analysis of three-phase flow in heterogeneous porous media. 2. Numerical simulations. *Water Resources Research.*, 33(7):1559–1566, 1997.

[3] G. Adomian. Applied stochastic processes. In G. Adomian, editor, *Stochastic System Analysis*, pages 1–17. Academic Press, New York, 1980.

[4] K. I. Al-Wali and P. J. Samson. Preliminary sensitivity analysis of urban airshed model simulations to temporal and spatial availability of boundary layer wind measurements. *Atmospheric Environment.*, 30(12-2):2027–2042, 1996.

[5] G. Alefeld and J. Herzberger. *"Introduction to Interval Computations"*. Academic Press, New York, 1983.

[6] K. E. Atkinson. *"An Introduction to Numerical Analysis, Second Edition"*. John Wiley & Sons, New York, 1988.

[7] R. Atkinson. Gas-phase trophospheric chemistry of organic compounds: A review. *Atmos. Environ.*, 24A(1):1–41, 1990.

[8] B. M. Ayyub and Kwan-Ling. Lai. Selective sampling in simulation-based reliability assessment. *International Journal of Pressure Vessels & Piping.*, 46(2):229–249, 1991.

[9] B. M. Ayyub and Kwan-Ling. Lai. Structural reliability assessment with ambiguity and vagueness in failure. *Naval Engineers Journal.*, 104(3):21–35, 1992.

[10] T. B. Bahder. *Mathematica for Scientists and Engineers.* Addison-Wesley, Reading, MA, USA, 1994.

[11] D. Baldocchi, A. Guenther, P. Harley, et al. The fluxes and air chemistry of isoprene above a deciduous hardwood forest. *Philosophical Transactions of the Royal Society of London Series. A-Physical Sciences & Engineering*, 315(1696):279–296, 1995.

[12] T. S. Barrett. *Probabilistic Design for Rotordynamic Simulations using Receptance-based Reanalysis.* PhD thesis, Texas A&M University, 1996.

[13] J. F. M. Barthelemy and L. E. Hall. Automatic differentiation as a tool in engineering design. *Structural Optimization.*, 9(2):76–82, 1995.

[14] L. R. Bauer and D. M. Hamby. Relative sensitivities of existing and novel model parameters in atmospheric tritium dose estimates. *Radiation Protection Dosimetry*, 37(4):253–260.

[15] J. C. Bezdek. Fuzzy models — what are they, and why? *Transactions on Fuzzy Systems*, 1(1):1–6, 1993.

[16] C. Bischof, A. Carle, P. Khademi, and A. Mauer. "The ADIFOR 2.0 System for the Automatic Differentiation of Fortran 77 Programs". Preprint MCS-P481-1194, Mathematics and Computer Science Division, Argonne National Laboratory, and CRPC-TR94491, Center for Research on Parallel Computation, Rice University, 1994.

[17] C. H. Bischof, P. Khademi, A. Mauer, and A. Carle. ADIFOR 2.0 - automatic differentiation of Fortran 77 programs. *IEEE Computational Science & Engineering.*, 3(3):18–32, 1996.

[18] C. H. Bischof, G. Pusch, and R. Knoesel. Sensitivity analysis of the MM5 weather model using automatic differentiation. *Computers in Physics*, 6(10):605–612, 1996.

[19] C. H. Bischof, L. Roh, and A. J. Maueroats. ADIC - an extensible automatic differentiation tool for ANSI-C. *Software-Practice & Experience.*, 27(12):1427–1456, 1997.

[20] P. Bjerager. On computation methods for structural reliability analysis. *Structural Safety*, 9(2):79–96, 1990.

[21] G. E. P. Box and N. R. Draper. *Empirical Model-Building and Response Surfaces.* John Wiley & Sons, New York, 1987.

[22] G. E. P. Box, Hunter W. G., and Hunter J. S. *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building.* John Wiley & Sons, New York, 1978.

[23] K. Breitung. Probability approximations by log likelihood maximization. *Journal of Engineering Mechanics-ASCE.*, 117(3):457–477, 1991.

[24] R. P. Broadwater, H. E. Shaalan, and W. J. Fabrycky. "Decision Evaluation with Interval Mathematics: A Power Distribution System Case Study". *IEEE Transactions on Power Delivery*, 9:59–65, 1994.

[25] W. Bryc. *The Normal Distribution.* LNS 100. Springer-Verlag, New York, 1995.

[26] W. Brzakala and W. Pula. Probabilistic analysis of foundation settlements. *Computers & Geotechnics.*, 18(4):291–309, 1996.

[27] C. G. Bucher and U. Bourgund. Fast and efficient response surface approach for structural reliability problems. *Structural Safety.*, 7(1):57–66, 1990.

[28] G. R. Camilo. *Food Web Analysis of Macrobenthic Riffle Communities (South Llano River, Texas).* PhD thesis, Texas Tech University, 1992.

[29] G. R. Carmichael, A. Sandu, and F. A. Potra. Sensitivity analysis for atmospheric chemistry models via automatic differentiation. *Atmospheric Environment.*, 31(3):475–489, 1997.

[30] R. F. Carsel and R. S. Parrish. Developing joint probability distributions of soil water retention characteristics. *Water Resour. Res.*, 29:755–770, 1988.

[31] M. C. Causley. "User's Guide for the Urban Airshed Model. Volume IV: User's Manual for the Emissions Preprocessor System". (U.S. EPA-450/4-90-007D), U.S. Environmental Protection Agency, Research Triangle Park, NC., 1990.

[32] Y. F. Chang and G. Corliss. ATOMFT - solving ODEs and DAEs using taylor series. *Computers & Mathematics with Applications.*, 28(10-12):209–233, 1994.

[33] H. M. L. Chaves and M. A. Nearing. Uncertainty analysis of the WEPP soil erosion model. *Transactions of the ASAE.*, 34(6):2437–2444, 1991.

[34] L. Chen, H. Rabitz, D. B. Considine, C. H. Jackman, and J. A. Shorter. Chemical reaction rate sensitivity and uncertainty in a two-dimensional middle atmospheric ozone model. *Journal of Geophysical Research-Atmospheres.*, 102(D13):16201–16214, 1997.

[35] S. Chinchalkar. Application of automatic differentiation to problems in engineering analysis. *Computer Methods in Applied Mechanics & Engineering.*, 118(1-2):197–207, 1994.

[36] G. Christakos and V. M. Vyas. "A composite space-time study of ozone distribution over the eastern United States". *Atmospheric Environment.*, page (In Press), 1998.

[37] R. A. Christensen. *Fitting Distributions to Data, Volume XI.* Entropy Ltd., 1994.

[38] K. C. Chun. "Uncertainty Data Base for Emissions-Estimation Parameters: Interim Report". "ANL/EES-TM-328", Argonne National Laboratory, Argonne, Illinois, 1987.

[39] L. Cizelj, B. Mavko, and H. Riesch-Oppermann. Application of first and second order reliability methods in the safety assessment of cracked steam generator tubing. *Nuclear Engineering & Design*, 147(3):359–368, 1994.

[40] D. C. Collins and R. Avissar. An evaluation with the fourier amplitude sensitivity test (FAST) of which land-surface parameters are of greatest importance in atmospheric modeling. *Journal of Climate.*, 7(5):681–703, 1994.

[41] M. E. Coltrin, R. J. Kee, and F. M. Rupley. Surface CHEMKIN: A general formalism and software for analyzing heterogeneous chemical kinetics at a gas-surface interface. *International Journal of Chemical Kinetics*, 23:1111, 1991.

[42] G. Corliss, A. Griewank, T. Robey, and S. Wright. Automatic differentiation applied to unsaturated flow — ADOL–C case study. Technical Memorandum ANL/MCS–TM–162, Mathematics and Computer Science Division, Argonne National Laboratory, April 1992.

[43] NRC (National Research Council). *Rethinking the Ozone Problem in Urban and Regional Air Pollution.* National Academy Press, Washington, DC, 1991.

[44] L. A. Cox. Reassessing benzene risks using internal doses and Monte-Carlo uncertainty analysis. *Environmental Health Perspectives.*, 104(Suppl 6):1413–1429, 1996.

[45] E. L. Crump, T. L. Jacobs, and P. A. Vesilind. "Fuzzy-set Approach for Optimizing Sludge Application Land Selection". *Journal of Urban Planning and Development*, 119:53–71, 1993.

[46] Decisioneering, Inc, Denver, CO. *Crystal Ball 4.0 for Risk Analysis.*

[47] F. R. deHoog, J. H. Knight, and A. N. Stokes. An improved method for numerical inversion of laplace transforms. *SIAM J. Sci. Stat. Comput.*, 3:357–366, 1982.

[48] R. G. Derwent. Treating uncertainty in models of the atmospheric chemistry of nitrogen compounds. *Atmospheric Environment*, 21(6):1445–1454, 1987.

[49] L. Devroye. *Non-Uniform Random Variate Generation.* Springer-Verlag, New York, 1986.

[50] M. Dobmann, M. Liepelt, and K. Schittkowski. Algorithm 746 - PCOM - A Fortran code for automatic differentiation. *ACM Transactions on Mathematical Software.*, 21(3):233–266, 1995.

[51] J. D. Doll and D. L. Freeman. "Randomly Exact Methods". *Science*, 234:1356–1360, 1986.

[52] W. M. Dong, W. L. Chiang, and F. S. Wong. Propagation of uncertainties in deterministic systems. *Computers & Structures.*, 26(3):415–423, 1987.

[53] C. H. Dou, W. Woldt, I. Bogardi, and M. Dahab. Steady state groundwater flow simulation with imprecise parameters. *Water Resources Research.*, 31(11):2709–2719, 1995.

[54] C. H. Dou, W. Woldt, I. Bogardi, and M. Dahab. Numerical solute transport simulation using fuzzy sets approach. *Journal of Contaminant Hydrology.*, 27(1-2):107–126, 1997.

[55] E. P. Doughtery and Rabitz H. A computational algorithm for the Green's function method of sensitivity analysis in chemical kinetics. *International Journal of Chemical Kinetics*, 11:1237–1249, 1979.

[56] E. P. Doughtery, J. T. Hwang, and Rabitz H. Further developments and applications of the Green's function method of sensitivity analysis in chemical kinetics. *International Journal of Chemical Kinetics*, 71(4):1794–1808, 1979.

[57] S. G. Douglas, R. C. Keller, and E. L. Carr. "User's Guide for the Urban Airshed Model. Volume III: User's Manual for the Diagnostic Wind Model". (U.S. EPA-450/4-90-007C), U.S. Environmental Protection Agency, Research Triangle Park, NC., 1990.

[58] Dow Chemical Company, Midland, MI. *SimuSolv Reference Guide – Volumes I and II.*

[59] N. R. Draper and H. Smith. *Applied Regression Analysis.* John Wiley & Sons, New York, 1981.

[60] A. M. Dunker. The decoupled direct method for calculating sensitivity coefficients in chemical kinetics. *Journal of Chemical Physics*, 81(5):2385–2393, 1984.

[61] Environmental Protection Agency (EPA). Health Assessment Document for Tatrachloroethylene (Perchloroethylene). External Review Draft. EPA-600/8-82-005B, 1983.

[62] U.S. EPA. "Background Document for EPA's Composite Model for Landfills (EPACML).". Technical report, U.S. EPA, Office of Solid Waste, Washington, D.C., 1990. Background Document and User's Guide.

[63] U.S. EPA. "Guide for Regulatory Application of the Urban Airshed Model M-91.". (EPA-450/4-91-013), U.S. Environmental Protection Agency, Research Triangle Park, NC., 1991.

[64] U.S. EPA. "EPA's Composite Model for leachate Migration and Transformation Products (EPACMTP).". Technical report, U.S. EPA, Office of Solid Waste, Washington, D.C., 1995. Background Document and User's Guide.

[65] G. W. Evans, W. Karwowski, and M. R. Wilhelm. "An Introduction to Fuzzy Set Methodologies for Industrial and Systems Engineering". In G. W. Evans, W. Karwowski, and M. R. Wilhelm, editors, *Applications of Fuzzy Set Methodologies in Industrial Engineering*, pages 3–11. Elsevier, New York, 1986.

[66] A. H. Falls, G. J. McRae, and J. H. Seinfeld. Sensitivity and uncertainty analysis of reaction mechanisms for photochemical air pollution. *International Journal of Chemical Kinetics*, 10:1137–1162, 1979.

[67] D. Farrar, B. Allen, K. Crump, and A. Shipp. Evaluation of uncertainty in input parameters to pharmacokinetic models and the resulting uncertainty in output. *Toxicology Letters*, 49:371–385, 1989.

[68] V. V. Fedorov. Analysis and design of simulation experiments for the approximation of models. IIASA, WP-83-71, 1983.

[69] G. Fichtner, H. J. Reinhart, and D. W. T. Rippin. "Design of Flexible Chemical Plants by the Application of Interval Mathematics". *Computers and Chemical Engineering*, 14:1311–1316, 1990.

[70] R. A. Fisher. *The Design of Experiments.* Hafner Press, New York, 1971.

[71] G. S. Fishman. *Monte Carlo : Concepts, Algorithms, and Applications.* Springer Verlag, New York, 1996.

[72] H. C. Frey. Quantitative analysis of uncertainty and variability in environmental policy making. Available from: Directorate for Science and Policy Programs, American Association for the Advancement of Science, 1333 H Street, NW, Washington, DC, September 1992.

[73] J. D. Fuentes, D. Wang, H. H. Neumann, et al. Ambient biogneic hydrocarbons and isoprene emissions from a mixed deciduous forest. *Journal of Atmospheric Chemitry*, 25(1):67–95, 1996.

[74] D. F. Gao, W. R. Stockwell, and J. B. Milford. Global uncertainty analysis of a regional-scale gas-phase chemical mechanism. *Journal of Geophysical Research-Atmospheres.*, 101(D4):9107–9119, 1996.

[75] C. W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences.* Springer-Verlag, New York, 1983.

[76] C. W. Gear. The automatic integration of ordinary differential equations. *Communications of the ACM*, 14:176–179, 1971.

[77] C.W. Gear. *Numerical Initial Value Problems in Ordinary Differential Equations.* Englewood Cliffs, 1971.

[78] P. G. Georgopoulos. "Regulatory Ozone Modeling: Status, Directions and Research Needs". *Environmental Health Perspectives*, 103 - Supplement 2:107–132, 1995.

[79] P. G. Georgopoulos and A. Roy. Review and evaluation of the Reactive Plume Model (RPM-II) modifications. Ozone Research Center Technical Report ORC-TR93-01, Environmental and Occupational Health Sciences Institute, Piscataway, NJ, 1993.

[80] P. G. Georgopoulos and J. H. Seinfeld. "Mathematical Modeling of Turbulent Reacting Plumes". A.R.B. Contract No. A0-044-32, California Institute of Technology, Pasadena, CA, 1986.

[81] P.G. Georgopoulos, S. Arunachalam, and S.W. Wang. Alternative Metrics for Assessing the Relative Effectiveness of NOx and VOC Emission Reductions in Controlling Ground-Level Ozone. *J. Air & Waste Manage. Assoc.*, 47:838–850, 1997.

[82] R. Ghanem and D. Ghiocel. Comparative analysis of FORM/SORM and polynomial chaos expansions for highly nonlinear systems. *Proceedings of Engineering Mechanics*, 1:535–538, 1996.

[83] R. G. Ghanem and P. D. Spanos. *Stochastic Finite Elements: A Spectral Approach.* Springer-Verlag, New York, 1991.

[84] M. Gonzalez. Analysis of the effect of microscale turbulence on atmospheric chemical reactions by means of the pdf approach. *Atmospheric Environment.*, 31(4):575–586, 1997.

[85] A. Griewank. On automatic differentiation. In M. Iri and K. Tanabe, editors, *Mathematical Programming: Recent Developments and Applications*, pages 83–108. Kluwer Academic Publishers, 1989.

[86] A. Griewank and G. F. Corliss, editors. *Automatic Differentiation of Algorithms: Theory, Implementation, and Application.* SIAM, Philadelphia, Penn., 1991.

[87] A. Griewank, D. Juedes, and J. Utke. Algorithm 755 - ADOL-C - a package for the automatic differentiation of algorithms written in C/C++. *ACM Transactions on Mathematical Software.*, 22(2):131–167, 1996.

[88] A. Guenther, P. Zimmerman, L. Klinger, et al. Estimates of regional natural volatile organic compound fluxes from enclosure and ambient measurements. *Journal of Geophysical Research-Atmospheres*, 101(D1):1345–1359, 1996.

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

[90] M. M. Hamed and P. B. Bedient. On the performance of computational methods for the assessment of risk from ground-water contamination. *Ground Water*, 35(4):638–646, 1997.

[91] M. M. Hamed, P. B. Bedient, and C. N. Dawson. Probabilistic modeling of aquifer heterogeneity using reliability methods. *Advances in Water Resources*, 19(5):277–295, 1996.

[92] M. M. Hamed, J. P. Conte, and P. B. Bedient. Probabilistic screening tool for ground-water contaminant assessment. *Journal of Environmental Engineering*, 121(11):767–775, 1995.

[93] D. Hattis, S. Tuler, L. Finkelstien, and L. Zhi-Quan. A Pharmakokinetic/Mechanism-Based Analysis of the Carcinogenic Risk of Perchloroethylene. No. CTPID 86-7, Center for Technology, Policy and Industrial Development, Massachusetts Institute of Technology, 1986.

[94] A. Heck. *"Introduction to Maple"*. Springer-Verlag, New York, 1992.

[95] B. Heller. *"MACSYMA for Statisticians"*. John Wiley & Sons, New York, 1991.

[96] J. C. Helton. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. pages 327–367, 1993.

[97] M. Hohenbichler, S. Gollwitzer, W. Kruse, and R. Rackwitz. New light on first- and second-order reliability methods. *Structural Safety.*, 4(4):267–284, 1987.

[98] J. E. Horwedel, R. J. Raridon, and R. Q. Wright. Sensitivity analysis of AIRDOS–EPA using ADGEN with matrix reduction algorithms. Technical Memorandum ORNL/TM 11373, Oak Ridge National Laboratory, Oak Ridge, Tenn., 1989.

[99] J. E. Horwedel, R. J. Raridon, and R. Q. Wright. Automated sensitivity analysis of an atmospheric dispersion model. *Atmospheric Environment.*, 26A(9):1643–1649, 1992.

[100] Jim E. Horwedel. GRESS: A preprocessor for sensitivity studies on Fortran programs. In Andreas Griewank and George F. Corliss, editors, *Automatic Differentiation of Algorithms: Theory, Implementation, and Application*, pages 243–250. SIAM, Philadelphia, Penn., 1991.

[101] P. Hovland, C. Bischof, D. Spiegelman, and M. Casella. Efficient derivative codes through automatic differentiation and interface contraction - an application in biostatistics. *SIAM Journal on Scientific Computing.*, 18(4):1056–1066, 1997.

[102] P. H. Howard, editor. *"Handbook of Fate and Exposure Data for Organic Chemicals"*, volume II. Lewis Publishers, 1990.

[103] P. J. Huber. *Robust statistical procedures.* SIAM, Philadelphia, 2nd edition, 1996.

[104] M. Hurme, M. Dohnal, and M. Jaervalaeinen. "Qualitative Reasoning in Chemical and Safety Engineering". *Computers and Chemical Engineering*, 17:441–446, 1993.

[105] D. Hwang, D. W. Byun, and M. T. Odman. An automatic differentiation technique for sensitivity analys is of numerical advection schemes in air quality models. *Atmospheric Environment.*, 31(6):879–888, 1997.

[106] J. T. Hwang, E. P. Dougherty, and H. Rabitz. The Green's function method of sensitivity analysis in chemical kinetics. *Journal of Chemical Physics*, 69(11):5180–5191, 1978.

[107] R. L. Iman and W. J. Conover. Small sample sensitivity analysis techniques for computer models, with an application to risk assessment. *Communications in Statistics, Part A. Theory and Methods*, 17:1749–1842, 1980.

[108] R. L. Iman and M. J. Shortencarier. A FORTRAN 77 program and user's guide for the generation of Latin Hypercube and random samples for use with computer models. NUREG/CR-3624, SAND83-2365, Sandia National Laboratories, Albuquerque, NM, 1984.

[109] M. Iri. Roles of automatic differentiation in nonlinear analysis and high-quality computation. *Nonlinear Analysis-Theory Methods & Applications.*, 30(7):4317–4328, 1997.

[110] J. C. Issac and R. K. Kapania. Aeroelastic sensitivity analysis of wings using automatic differentiation. *AIAA Journal.*, 35(3):519–525, 1997.

[111] Y. S. Jang, N. Sitar, and A. D. Kiureghian. Reliability analysis of contaminant transport in saturated porous media. *Water Resources Research*, 30(8):2435–2448, 1994.

[112] P. H. M. Janssen, P. S. C. Heuberger, and R. Sanders. UNCSAM: A tool for automating sensitivity and uncertainty analysis. *Environmental Software.*, 9(1):1–11, 1994.

[113] M. Jerosolimski and L. Levacher. New method for fast calculation of jacobian matrices: automatic differentiation for power system simulation. *IEEE Transactions on Power Systems.*, 9(2):700–706, 1994.

[114] N. L. Johnson and S. Kotz. *Distributions in Statistics: Continuous Multivariate Distributions.* John Wiley & Sons, New York, 1972.

[115] C. H. Juang, X. H. Huang, and D. J. Elton. Modelling and analysis of non-random uncertainties - fuzzy-set approach. *International Journal for Numerical & Analytical Methods in Geomechanics*, 16(5):335–350, 1992.

[116] M. H. Kalos and P. A. Whitlock. *Monte Carlo Methods : Basics.* John Wiley & Sons, New York, 1986.

[117] A. Karamchandani and C. A. Cornel. Sensitivity estimation within first and second order reliability methods. *Structural Safety*, 11(2):95–107, 1992.

[118] L. S. Katafygiotis and C. Papadimitriou. Dynamic response variability of structures with uncertain properties. *Earthquake Engineering & Structural Dynamics.*, 25(8):775–793, 1996.

[119] A. Kaufmann and M. M. Gupta. *Introduction to Fuzzy Arithmetic: Theory and Applications.* Van Nostrand Reinhold Company, New York, New York, 1985.

[120] R. B. Kearfott. Algorithm 763: INTERVAL_ARITHMETIC: A fortran 90 module for an interval data type. *ACM Transactions on Mathematical Software*, 22(4):385–392, December 1996.

[121] R. B. Kearfott, M. Dawande, K. Du, and C. Hu. Algorithm 737: INTLIB: A portable Fortran-77 elementary function library. *ACM Transactions on Mathematical Software*, 20(4):447–459, December 1994.

[122] R. B. Kearfott and V. Kreinovich, editors. *"Applications of Interval Computations"*, volume 3 of *Applied Optimization.* Kluwer Academic Publishers, 1996.

[123] A. I. Khuri and J. A. Cornell. *Response Surfaces: Design and Analyses.* Marcel Dekker, New York, 1987.

[124] T. W. Kim, S. H. Chang, and B. H. Lee. Comparative study on uncertainty and sensitivity analysis and application to LOCA model. *Reliability Engineering & System Safety.*, 21(1):1–26, 1988.

[125] G.J. Klir. "The Many Faces of Uncertainty". In B.M Ayyub and M.M. Gupta, editors, *Uncertainty Modeling and Analysis: Theory and Applications*, pages 3–19. Elsevier Science, 1994.

[126] J. G. Klir and B. Yuan. *Fuzzy Sets and Fuzzy Logic: Theory and Applications.* Prentice Hall, Englewood Cliffs, New Jersey, 1995.

[127] M. Koda, G. J. McRae, and J. H. Seinfeld. Automatic sensitivity analysis of kinetic mechanisms. *International Journal of Chemical Kinetics*, 11:427–444, 1979.

[128] J. Korelc. Automatic generation of finite-element code by simultaneous optimization of expressions. *Theoretical Computer Science.*, 187(1-2):231–248, 1997.

[129] H. U. Koyluoglu, A. S. Cakmak, and S. R. K. Neilson. "Interval algebra to deal with pattern loading and structural uncertainty". *Journal of Engineering Mechanics*, 121(11):1149–1157, 1995.

[130] A. Kraslawski. Review of applications of various types of uncertainty in chemical engineering. *Chemical Engineering & Processing.*, 26(3):185–191, 1989.

[131] A. Kraslawski, T. Koiranen, and L. Nystrom. "Concurrent Engineering: Robust Design in Fuzzy Environment". *Computers and Chemical Engineering*, 17S:S447–S452, 1993.

[132] V. Kreinovich, L. Anatoly, J. Rohn, and P. Kahl, editors. *"Computational Complexity and Feasibility of Data Processing and Interval Computations"*, volume 10 of *Applied Optimization*. Kluwer Academic Publishers, 1997.

[133] K. Kubota. PADRE2, a FORTRAN precompiler yielding error estimates and second derivatives. In Andreas Griewank and George F. Corliss, editors, *Automatic Differentiation of Algorithms: Theory, Implementation, and Application*, pages 251–262. SIAM, Philadelphia, Penn., 1991.

[134] S. Kutscher and J. Schulze. "Some Aspects of Uncertain Modelling Experiences in Applying Interval Mathematics to Practical Problems". In H. Bandemer, editor, *Modelling Uncertain Data*, pages 62–68. Akademie Verlag, Berlin, 1993.

[135] B. Lamb, D. Gay, H. Westberg, and T. Pierce. Biogenic hydrocarbon emission inventory for the U.S.A. using a simple forest canopy model. *Atmos. Environ.*, 27A(11):1673–1690, 1993.

[136] B. Lamb, A. Guenther, D. Gay, and H. Westberg. National inventory of biogenic hydrocarbon emissions. *Atmos. Environ.*, 21(8):1695–1705, 1987.

[137] R. G. Lamb. "Note on the Application of K-Theory to Diffusion Problems Involving Nonlinear Chemical Reaction". *Atmospheric Environment*, 7:257–263, 1973.

[138] R. G. Lamb, W. R. Shu, D. R. Durran, J. H. Seinfeld, and L. E. Reid. Continued research in mesoscale air pollution simulation modeling, volume VI: further studies in the modeling of microscale phenomena. Technical report, U.S. EPA, 1984. EPA-600/3-84-095B.

[139] R.G. Lamb. Continued research in mesoscale air pollution simulation modeling, volume iii: modeling of microscale phenomena. Technical report, U.S. EPA, 1976. EPA-60-0/4-76-016c.

[140] B. W. Lee and O. K. Lim. Application of the first-order perturbation method to optimal structural design. *Structural Engineering & Mechanics.*, 4(4):425–436, 1996.

[141] J. T. Lim, H. J. Gold, G. G. Wilkerson, and C. D. Raper. Monte carlo/response surface strategy for sensitivity analysis. application to a dynamic model of vegetative plant growth. *Applied Mathematical Modelling.*, 13(8):479–484, 1989.

[142] X. Lin, O. T. Melo, D. R. Hastie, et al. Case study of ozone production in a rural area of central ontario. *Atmos. Environ.*, 26A(2):311–324, 1992.

[143] T. S. Liou and H. D. Yeh. Conditional expectation for evaluation of risk groundwater flow and solute transport – one-dimensional analysis. *Journal of Hydrology.*, 199(3-4):378–402, 1997.

[144] M. K. Liu, G. E. Moore, and H. Y. Holman. "Survey of Plume Models for Atmospheric Application". Interin Report (EA-2243/1616-9), Systems Applications International, San Rafael, California, February 1982.

[145] Y. Q. Liu and R. Avissar. Sensitivity of shallow convective precipitation induced by land surface heterogeneities to dynamical and cloud microphysical parameters. *Journal of Geophysical Research-Atmospheres.*, 101(D3):7477–7497, 1996.

[146] W. L. Loh. On Latin Hypercube Sampling. *Annals of Statistics.*, 24(5):2058–2080, 1996.

[147] R. Lu, Y. Luo, and J. P. Conte. Reliability evaluation of reinforced concrete beam. *Structural Safety*, 14:277–298, 1994.

[148] G. J. McRae, W. R. Goodin, and J. H. Seinfeld. "Mathematical Modeling of Photochemical Air Pollution". Technical report, Environmental Quality Laboratory, California Institute of Technology, Pasadena, CA, 1982. (EQL report no. 18).

[149] G. J. McRae, J. W. Tilden, and J. H. Seinfeld. Global sensitivity analysis - a computational implementation of the fourier amplitude sensitivity test (FAST). *Computers and Chemical Engineering*, 6(1):15–25, 1982.

[150] C. Mischler, X. Joulia, E. Hassold, A. Galligo, and R. Esposito. Automatic differentiation applications to computer aided process engineering. *Computers & Chemical Engineering.*, 19(Suppl):S779–S784, 1995.

[151] R. E. Moore. *"Interval Analysis"*. Prentice-Hall, Englewood Cliffs, New Jersey, 1966.

[152] R. E. Moore. *"Methods and Applications of Interval Analysis"*. SIAM, Philadelphia, 1979.

[153] E. R. Morris, E. C. Chang, Shepard S. B., and M. P. Ligocki. "User's Guide to Version IV of the Reactive Plume Model (RPM-IV)". (SYSAPP-92/037), Systems Applications International, San Rafael, California, April 1992.

[154] E. R. Morris and T. C. Myers. "User's Guide for the Urban Airshed Model. Volume I: User's Guide for the UAM (CB-IV).". (EPA-450/4-90-007A), U.S. Environmental Protection Agency, 1990.

[155] A. Neumaier. *"Interval Methods for Systems of Equations"*. Cambridge University Press, Cambridge, England, 1990.

[156] E. M. Oblow, F. G. Pin, and R. Q. Wright. Sensitivity analysis using computer calculus: A nuclear waste application. *Nuclear Science & Engineering*, 94(1):46–65, 1986.

[157] H. Osnes. Stochastic analysis of velocity spatial variability in bounded rectangular heterogeneous aquifers. *Advances in Water Resources.*, 21(3):203–215, 1998.

[158] I. Ozaki and T. Terano. Applying an automatic differentiation technique to sensitivity analysis in design optimization problems. *Finite Elements in Analysis & Design.*, 14(2-3):143–151, 1993.

[159] W. W. Pan, M. A. Tatang, G. J. McRae, and R. G. Prinn. Uncertainty analysis of direct radiative forcing by anthropogenic sulfate aerosols. *Journal of Geophysical Research-Atmospheres*, 102(D18):21915–21924, 1997.

[160] A. Papoulis. *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, New York, 1991.

[161] F. G. Pin, B. A. Worley, et al. An automated sensitivity analysis procedure for the performance assessment of nuclear waste isolation systems. *Nuclear & Chemical Waste Management*, 6(3-4):255–263, 1986.

[162] PLG, Inc, Newport Beach, CA. *RISKMAN: Integrated Quantitative Risk Assessment Package.*

[163] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *"Numerical Recipes (Fortran Version)".* Cambridge University Press, New York, 1989.

[164] Primavera Systems, Inc, San Francisco, CA. *Monte Carlo 3.0.*

[165] G. J. Prokopakis. Decoupled direct method for the solution of ordinary boundary value problems. *Applied Mathematical Modelling.*, 17(9):499–503, 1993.

[166] M. L. Puri and D. A. Ralescu. Fuzzy random variables. *J. Math. Analysis and Applications*, 114:409–422, 1986.

[167] S. Quin and G. E. O. Widera. "Application of Fuzzy Relational Modeling to Industrial Product Quality Control". *Journal of Pressure Vessel Technology - Transactions of the ASME*, 118:121–124, 1996.

[168] K. Radhakrishnan and D. A. Bittker. GCKP86 - an efficient code for general chemical kinetics and sensitivity analysis computations. *Chemical and Physical Processes in Combustion*, 4:1–46, 1986.

[169] M. E. Reed and W. B. Whiting. Sensitivity and uncertainty of process designs to thermodynamic model parameters: a Monte Carlo approach. *Chemical Engineering Communications*, 124:39–48, 1993.

[170] R. H. Reitz and R. J. Nolan. Physiological pharmacokinetic modeling for perchloroethylene dose adjustment. Unpublished Data Submitted to EPA Science Advisory Board, 1986.

[171] L. C. Rich and D. R. Hill. Automatic differentiation in MATLAB. *Applied Numerical Mathematics.*, 9(1):33–43, 1992.

[172] Y. Rong, R. F. Wang, and R. Chou. Monte Carlo simulation for a groundwater mixing model in soil remediation of tetrachloroethylene. *Journal of Soil Contamination.*, 7(1):87–102, 1998.

[173] S. J. Roselle. Effects of biogenic emission uncertainties on regional photochemical modeling of control strategies. *Atmospheric Environment.*, 28(10):1757–1772, 1994.

[174] T. Ross. *Fuzzy Logic with Engineering Applications.* McGraw-Hill, 1995.

[175] N. Rostaing, S. Dalmas, and A. Galligo. Automatic differentiation in Odyssee. *Tellus Series A-Dynamic Meteorology & Oceanography.*, 45A(5):558–568, 1993.

[176] A. Roy. *Mechanistic Modeling of Transport and Metabolism in Physiological Systems.* PhD thesis, Rutgers, The State University of New Jersey, New Brunswick, NJ, 1997.

[177] R. Y. Rubinstein. *Simulation and the Monte Carlo Method.* John Wiley & Sons, New York, 1981.

[178] R. D. Scheffe and R. E. Morris. A Review of the Development and Application of the Urban Airshed Model. *Atmospheric Environment*, 27B:23–39, 1993.

[179] B. A. Schichtel and R. B. Husar. Regional simulation of atmospheric pollutants with the CAPITA Monte Carlo model. *Journal of the Air & Waste Management Association.*, 47(3):331–343, 1997.

[180] P. R. Schroeder et al. "The hydrologic evaluation of landfill performance model (HELP): Volume I - Users guide for version I and Volume II - Documentation for version I.". Technical report, U.S. EPA, U.S. EPA, Washington, D.C., 1984. EPA/530-SW-84-009.

[181] J. H. Seinfeld. *"Atmospheric Chemistry and Physics of Air Pollution".* John Wiley & Sons, 1986.

[182] J. H. Seinfeld. Ozone air quality models: A critical review. *Journal of Air Pollution Control Association*, 38(5):616–645, 1988.

[183] K. Sexton and H. Westberg. "Photochemical Ozone Formation from Petroleum Refinery Emissions". *Atmospheric Environment*, 17:467–475, 1983.

[184] H. E. Shaalan and R. P. Broadwater. "Using Interval Mathematics in Cost Benefit Analysis of Distribution Automation". *Electric Power Systems Research*, 27:145–152, 1993.

[185] T. D. Sharkey. Isoprene synthesis by plants and animals. *Endeavour*, 20(2):74–78, 1996.

[186] D. Shiriaev, A. Griewank, and J. Utke. *A User Guide to ADOL–F: Automatic Differentiation of Fortran Codes.* Institute of Scientific Computing, TU Dresden, Germany.

[187] T. W. Simon. Combining physiologically based pharmacokinetic modeling with Monte Carlo simulation to derive an acute inhalation guidance value for trichloroethylene. *Regulatory Toxicology & Pharmacology.*, 26(3):257–270, 1997.

[188] D. Simpson, A. Guenther, C. N. Hewitt, and R. Steinbrecher. Biogenic emissions in europe .1. estimates and uncertainties. *Journal of Geophysical Research-Atmospheres*, 100(D11):22875–22890, 1995.

[189] G. Sistla, S.T. Rao, and Godowitch J. Sensitivity analysis of a nested ozone air quality model. Proceedings of the AMS/AWMA Joint Conference on Applications of Air Pollution Modeling and Its Applications, New Orleans, LA, 1991.

[190] V. Sitaraman and R. B. Oza. Particle trajectory model of atmospheric dispersion on a parallel processing system. *Environmental Software.*, 11(4):229–234, 1996.

[191] M. Smithson. *Ignorance and Uncertainty: Emerging Paradigms.* Springer-Verlag, New York, 1988.

[192] I. M. Sobol. *A Primer for the Monte Carlo Method.* CRC Press, 1994.

[193] D. J. Spiegelhalter. "A Statistical View of Uncertainty in Expert Systems". In W. A. Gale, editor, *Artificial Intelligence and Statistics*, pages 17–55. Addison-Wesley, Reading, 1986.

[194] M. Stein. Large sample properties of simulations using Latin Hypercube Sampling. *Technometrics.*, 29(2):143–151, 1987.

[195] D. A. Stewart and M. Liu. "Development and Application of a Reactive Plume Model". *Atmospheric Environment*, 15:2377–2393, 1981.

[196] E. A. Sudicky. The Laplace Transform Galerkin Technique: A time-continuous finite element theory and application to mass transport in groundwater. *Water Resour. Res.*, 25(8):1833–1846, 1989.

[197] M. A. Tatang. "Combined Stochastic and Deterministic Approach in Solving Stochastic Differential Equations". Master's thesis, Carnegie Mellon University, 1992.

[198] M. A. Tatang. *"Direct Incorporation of Uncertainty in Chemical and Environmental Engineering Systems."*. PhD thesis, Massachusetts Institute of Technology, 1995.

[199] M. A. Tatang, W. W. Pan, R. G. Prinn, and G. J. McRae. An efficient method for parametric uncertainty analysis of numerical geophysical model. *Journal of Geophysical Research-Atmospheres*, 102(D18):21925–21932, 1997.

[200] R. Tomović and M. Vukobratović. *General Sensitivity Theory.* Elsevier, New York, 1972.

[201] C. P. Tsokos. *Probability Distributions: An Introduction to Probability Theory with Applications.* Wadsworth Publishing Company, Belmont, CA, 1972.

[202] U. S. Environmental Protection Agency. *Exposure Models Library (EML) and Integrated Model Evaluation System (IMES)*, March 1996. EPA/600/C-92/002, Developed for the Office of Research and Development, USEPA, by Versar, Inc., Columbia, MD.

[203] U. S. Environmental Protection Agency. *Summary Report for the Workshop on Monte Carlo Analysis. EPA/630/R-96/001*, September 1996.

[204] U. S. Environmental Protection Agency. *Policy for use of Probabilistic Analysis in Risk Assessment. EPA/630/R-97/001*, March 1997.

[205] M. Th. van Genuchten. A closed form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J.*, 44:892–898, 1980.

[206] M. Vanderperk. Effect of model structure on the accuracy and uncertainty of results from water quality models. *Hydrological Processes.*, 11(3):227–239, 1997.

[207] B. E. Vieux and S. Needham. Nonpoint-pollution model sensitivity to grid-cell size. *Journal of Water Resources Planning & Management-ASCE.*, 119(2):141–157, 1993.

[208] J. Villadsen and M. L. Michelsen. *Solution of Differential Equation Models by Polynomial Approximation.* Prentice-Hall, Englewood Cliffs, New Jersey, 1978.

[209] L. Vuilleumier, R. A. Harley, and N. J. Brown. First- and second-order sensitivity analysis of a photochemically reactive system (a Green's function approach). *Environmental Science & Technology.*, 31(4):1206–1217, 1997.

[210] V. M. Vyas and G. Christakos. Spatiotemporal analysis and mapping of sulfate deposition data over eastern usa. *Atmospheric Environment.*, 31(21):3623–3633, 1997.

[211] A. Walia. *Microenvironmental Modeling Integrated With Geographic Information Systems for Exposure Analysis.* PhD thesis, Rutgers University, Piscataway, NJ, January 1998.

[212] P. Walley. *"Statistical Reasoning with Imprecise Probabilities".* Chapman & Hall, London, 1991.

[213] S.W. Wang, P.G. Georgopoulos, G. Li, and H. Rabitz. Condensing Complex Atmospheric Chemistry Mechanisms - 1: The Direct Constrained Approximate Lumping (DCAL) Method Applied to Alkane Photochemistry. *Environ. Sci. Technol.*, In Press, 1998.

[214] G. Whiffen, C. Shoemaker, C. H. Bischof, A. Ross, and A. Carle. Application of automatic differentiation to groundwater transport models. In A. Peters, editor, *Computational Methods in Water Resources X*, pages 173–182. Kluwer Academic Publishers, 1994.

[215] W. B. Whiting, T. M. Tong, and M. E. Reed. Effect of uncertainties in thermodynamic data and model parameters on calculated process performance. *Industrial & Engineering Chemistry Research.*, 32(7):1367–1371, 1993.

[216] G. Z. Whitten, H. Hogo, and J. P. Killus. "The Carbon Bond Mechanism: A Condensed Kinetic Mechanism for Photochemical Smog" Analysis Techniques to a Photochemical Ozone Model". *Environmental Science and Technology*, 14:690, 1980.

[217] E. B. WIlson and M. M. Hilferty. The distribution of chi-square. *Proceedings of the National Academy of Sciences*, 17:684–688, 1931.

[218] W. Winston. *Simulation Modeling Using @RISK (User's Guide to @RISK software).* Palisade Corporation - PAL #54, Newfield, New York, 1996.

[219] K. L. Wood, K. N. Otto, and E. K. Antonsson. "Engineering Design Calculations with Fuzzy Parameters". *Fuzzy Sets and Systems*, 52:1–20, 1992.

[220] B. A. Worley, F. G. Pin, J. E. Horwedel, and E. M. Oblow. ADGEN - ADjoint GENerator for computer models. Technical Report ORNL/TM-11037, Oak Ridge National Laboratory, 1989.

[221] J. Yen, R. Langari, and L. A. Zadeh. *Industrial Applications of Fuzzy Logic and Intelligent Systems.* IEEE Press, Piscataway, New Jersey, 1995.

[222] L. A. Zadeh. "Fuzzy Sets". *Information and Control*, 8:338–253, 1965.

[223] P. Zannetti. *"Air Pollution Modeling".* Van Nostrand Reinhold, 1990.

[224] D. X. Zhang. Numerical solutions to statistical moment equations of groundwater flow in nonstationary, bounded, heterogeneous media. *Water Resources Research.*, 34(3):529–538, 1998.

[225] J. Zimmermann and D. Poppe. A supplement for the RADM2 chemical mechanism - the photooxidation of isoprene. *Atmos. Environ.*, 30(8):1255–1269, 1996.

Appendix A

PROBABILISTIC APPROACH FOR UNCERTAINTY ANALYSIS

Probabilistic approach is the most widely used technique for uncertainty analysis of mathematical models. There are a number of text books that describe the concepts and application of probabilistic analysis in detail. Tsokos [201] presents excellent introductory material for probabilistic analysis, Johnson [114] explains the multivariate random variables, and Papoulis [160] presents an excellent description on probability and random variables from a mathematical view point. Additionally, Gardiner [75] presents the applications of probabilistic analysis in modeling. This appendix attempts to merely summarize some basic information on probability and random variables, which can be found in more detail in the abovementioned texts.

In the probabilistic approach, uncertainties are characterized by the *probabilities* associated with *events*. An event corresponds to any of the possible states a physical system can assume, or any of the possible predictions of a model describing the system. In the study of environmental pollution, the situation where the contaminant concentration exceeds a regulatory level can be an event. Similarly, in case of mathematical modeling, the situation where the model outputs fall in a certain range, is also an event. In short, any possible outcome of a given problem, such as the tossing of a coin, experimental measurement of contaminant concentration,as and mathematical modeling of a physical system, is an event. The *probability* of an event can be interpreted in terms of the frequency of occurrence of that event. When a large number of *samples* or *experiments* are considered, the probability of an event is defined as the ratio of the number of times the event occurs to the total number of samples or experiments. A probability of 0 for an event means that the event will never occur, and a probability

of 1 indicates that the event will always occur. Examples of experiments or samples include repeated coin tossing, a large number of independent measurements of contaminant concentration, or a large number of simulations using a mathematical model with randomly varying parameters.

The following examples illustrate the concept of probability based on a large number of samples: (a) if a coin is tossed, the probability of heads turning up is 0.5 and that of the tails is 0.5. This means that if the coin is tossed a large number of times, heads will turn up roughly half the time and tails half the time. (b) the statement that the probability that a pollutant concentration c lies between c_1 and c_2 equals p means the following: from a large number of independent measurements of the concentration c, under identical conditions, the number of times the value of c lies between c_1 and c_2 is roughly equal to the fraction p of the total number of samples.

A.1 Random Variables

Random variables are mathematical quantities that are used to represent probabilistic uncertainty. They can describe the probabilities associated with each value, or each sub-range of values, an uncertain quantity can take. For example, if a random variable $\boldsymbol{c}$ represents the uncertainty in the concentration of a pollutant, the following questions can be answered by analyzing the random variable $\boldsymbol{c}$: Given a concentration c_0, what is the probability that $\boldsymbol{c}$ takes a value lower than c_0; or given two numbers c_1 and c_2, what is the probability that $\boldsymbol{c}$ takes values between c_1 and c_2?

Mathematically, a random variable $\boldsymbol{x}$ maps a probability space $\boldsymbol{\Omega}$ onto the real line. The random variable $\boldsymbol{x}$ can assume values from $-\infty$ to $+\infty$, and there is an associated probability for each value (or interval) that $\boldsymbol{x}$ takes. Based on the nature of the values that random variables can assume, they can be classified into three types:

(a) continuous random variables: these variables can assume continuous values from an interval. Examples include contaminant concentrations in the environment; emissions from an industrial source; and physical parameters in an exposure model, such as body weight or respiration rate. In these cases, one cannot define the probability that a random variable $\boldsymbol{x}$ is exactly equal to a value x_0, since

there are uncountably infinite number of possible values, and the answer for each point would be zero. Hence, in such cases, probabilities are defined on intervals (e.g., the probability that $\boldsymbol{x}$ lies between x_1 and x_2). Further, a *probability density* can be defined at each point in the interval; the probability density at a point is representative of the probability in the vicinity of that point.

(b) discrete random variables: these variables can assume discrete values from a set. Examples include the following: rolling of dice, where the outcome can have only integer values between 1 and 6; the number of days an air quality standard is violated in a year - this can assume integral values between 0 and 365; and the number of defective cars in a production line. These variables have probabilities associated with a countable number of values they can assume. Continuous and discrete random variables can be contrasted as follows: one can define the probability that the number of air quality exceedences to be exactly equal to a given number n, where as one cannot define the probability that the atmospheric concentration is exactly equal to a given concentration c.

(c) mixed random variables: these variables can assume continuous as well as discrete values. For example, the sum of a discrete and a continuous random variable results in a mixed random variable. They may have the properties of continuous random variables in certain ranges, and may have properties of discrete random variables in others.

A major part of uncertainty analysis in environmental modeling and risk characterization involves uncertainties in continuous quantities. Examples include the uncertainties in measured or predicted concentrations, and the uncertainties in the estimated time of exposure to a contaminant. The present work focuses mainly on uncertainties described by continuous random variables.

A.1.1 Continuous Random Variables

Continuous random variables are characterized through the following functions:

(a) Cumulative density function, $F_{\boldsymbol{x}}(x)$. This denotes the probability that the random variable $\boldsymbol{x}$ has a value less than or equal to x. This function is also known as the *cumulative distribution*. The probability that $\boldsymbol{x}$ has a value greater than x is given by 1-$F_{\boldsymbol{x}}(x)$. Further, the probability that $\boldsymbol{x}$ takes on values between x_1 and x_2 for $x_2 \geq x_1$, is given by:

$$\Pr\{x_1 \leq \boldsymbol{x} \leq x_2\} = F_{\boldsymbol{x}}(x_2) - F_{\boldsymbol{x}}(x_1)$$

The important characteristics of the cumulative density function are:

$$F_{\boldsymbol{x}}(x_2) \geq F_{\boldsymbol{x}}(x_1) \text{ iff } x_2 \geq x_1 \quad \text{i.e., } F_{\boldsymbol{x}}(x) \text{ is monotonous}$$

$$0 \leq F_{\boldsymbol{x}}(x) \leq 1, \text{ for } -\infty < x < \infty$$

$$F_{\boldsymbol{x}}(-\infty) = 0 \text{ and } F_{\boldsymbol{x}}(\infty) = 1$$

In population risk characterization, the corresponding cumulative density function can be considered analogous to the fraction of the population that is at risk with respect to a given risk criterion.

(b) Probability density function, $f_{\boldsymbol{x}}(x)$. This function is also known as the *probability distribution*. This is the derivative of the cumulative density function.

$$f_{\boldsymbol{x}}(x) = \frac{dF_{\boldsymbol{x}}(x)}{dx}$$

The main properties of the probability density function are:

$$f_{\boldsymbol{x}}(x) \geq 0, \text{ for } -\infty < x < \infty$$

$$f_{\boldsymbol{x}}(-\infty) = f_{\boldsymbol{x}}(\infty) = 0$$

$$F_{\boldsymbol{x}}(x) = \int_{-\infty}^{x} f_{\boldsymbol{u}}(u)du$$

$$\Pr\{x_1 \leq \boldsymbol{x} \leq x_2\} = F_{\boldsymbol{x}}(x_2) - F_{\boldsymbol{x}}(x_1) = \int_{x_1}^{x_2} f_x(x)dx$$

(c) Expected value of a function, $g(\boldsymbol{x})$, of a random variable. This is defined as

$$\mathrm{E}\{g(\boldsymbol{x})\} = \int_{-\infty}^{\infty} g(x) f_{\boldsymbol{x}}(x)dx$$

A probability density function or a cumulative density function fully characterizes a random variable. In the following sections, the probability density function $f_{\boldsymbol{x}}(x)$ and the random variable $\boldsymbol{x}$ are interchangeably used, as they both represent the same. Further, the terms "distribution" and "random variable" are also interchangeably used.

Even though the density functions of a random variable provide all the information about that random variable, they do not provide information on uncertainty at a quick glance, especially when the distribution functions consist of a complex algebraic expressions. In such cases, the *moments* of a random variable serve as useful metrics that provide a significant amount of information about a distribution.

A.1.2 Moments of a Random Variable

The moments of a random variable provide concise information about a random variable. The following are the moments that describe a random variable.

(a) The *expected value* or *mean* of a random variable: this denotes the average value for the distribution of the random variable. If a large number of samples from the distribution are considered, the expected value of the distribution is equal to the arithmetic mean of the sample values.

Mathematically, the mean η of a random variable $\boldsymbol{x}$ is given by

$$\eta = \mathrm{E}\{\boldsymbol{x}\} = \int_{-\infty}^{\infty} x f_{\boldsymbol{x}}(x) dx$$

It should be noted that the mean does not necessarily represent a realistic value from the distribution. For example, if a coin is tossed a large number of times, and if heads is assigned a value 1, and tails a value 0, the mean is 1/2, which is not a possible outcome in the coin tossing problem.

(b) The *variance* or dispersion of a distribution: this indicates the spread of the distribution with respect to the mean value. A lower value of variance indicates that the distribution is concentrated close to the mean value, and a higher value indicates that the distribution is spread out over a wider range of possible values.

The variance σ^2 of $\boldsymbol{x}$ is given by:

$$\sigma^2 = \mathrm{E}\{(\boldsymbol{x} - \eta)^2\} = \int_{-\infty}^{\infty} (x - \eta)^2 f_{\boldsymbol{x}}(x) dx$$

The square root of variance is called the *standard deviation* (σ).

(c) The *skewness* of a distribution indicates the asymmetry of the distribution around its mean, characterizing the shape of the distribution. It is given by

$$\gamma_1 = \frac{1}{\sigma^3}\mathrm{E}\{(\boldsymbol{x} - \eta)^3\} = \frac{1}{\sigma^3}\int_{-\infty}^{\infty} (x - \eta)^3 f_{\boldsymbol{x}}(x) dx$$

A positive value of skewness indicates that the distribution is skewed towards values greater than the mean (i.e., skewed towards the right side) and a negative value indicates that the distribution is skewed towards the left side.

(d) The *kurtosis* of a distribution indicates the flatness of the distribution with respect to the normal distribution. It is given by

$$\gamma_2 = \frac{1}{\sigma^4}\mathrm{E}\{(\boldsymbol{x} - \eta)^4\} = \frac{1}{\sigma^4}\int_{-\infty}^{\infty} (x - \eta)^4 f_{\boldsymbol{x}}(x) dx$$

A value of kurtosis higher than 3 indicates that the distribution is flatter compared to the normal distribution, and a smaller value indicates a higher peak (relative to the normal distribution) around the mean value.

(e) Higher order moments: The higher order moments of a random variable are defined as follows:

The kth moment m_k of $\boldsymbol{x}$ is given by

$$m_k = \mathrm{E}\{\boldsymbol{x}^k\} = \int_{-\infty}^{\infty} x^k f_{\boldsymbol{x}}(x) dx$$

The kth central moment μ_k of a random variable $\boldsymbol{x}$ is defined as

$$\mu_k = \mathrm{E}\{(\boldsymbol{x} - \eta)^k\} = \int_{-\infty}^{\infty} (x - \eta)^k f_{\boldsymbol{x}}(x) dx$$

Clearly, $m_0 = 1$, $m_1 = \eta$, $\mu_0 = 1$, $\mu_1 = 0$, and $\mu_2 = \sigma^2$. Further, the central moments μ_k's and the moments m_k's are related as follows:

$$\mu_k = \sum_{r=0}^{k} {}^kC_r (-1)^r \eta^r m_{k-r} \quad \text{and}$$

$$m_k = \sum_{r=0}^{k} {}^kC_r \eta^r \mu_{k-r}$$

where

$$ {}^{k}C_{r} = \frac{k.(k-1)\ldots(k-r-2)(k-r-1)}{1.2\ldots r} = \frac{k!}{(k-r)!\ r!} $$

A.1.3 Median, Mode and Percentiles of a distribution

In addition to the information provided by the moments of a distribution, some other metrics such as the *median* and the *mode* provide useful information. The median of a distribution is the value for the 50th percentile of the distribution (i.e., the probability that a random variable takes a value below the median is 0.5). The mode of a distribution is the value at which the probability density is the highest. For example, for the normal distribution $\mathrm{N}(\mu, \sigma)$, with zero mean, the mean, median, and mode are equal to μ, whereas for the lognormal distribution, $\mathrm{LN}(\mu, \sigma)$, the mean is $e^{(\mu+\sigma^2/2)}$, the median is e^{μ}, and the mode is $e^{(\mu-\sigma^2)}$. Additionally, other percentiles of the distribution may sometimes be desired. For example, the 95th percentile indicate the values above which 5% of the samples occur.

A.2 Jointly Distributed Random Variables

When the behavior of a system depends on more than one random input, some of which may be interdependent, the relationships among these random inputs need to be considered for uncertainty analysis. This is especially useful in situations where the knowledge of one variable provides information regarding other variables, thus restricting the range of possible values the other variables can assume. Such relationships can be described by joint probability distributions.

To illustrate the concept of probability analysis involving many random variables, the case for two random variables is presented first, and the results are extended for the general case involving a set of random variables.

Given two random variables $\boldsymbol{x}$ and $\boldsymbol{y}$, the probabilities

$$\Pr\{\boldsymbol{x} \leq x\} = F_{\boldsymbol{x}}(x) \text{ and } \Pr\{\boldsymbol{y} \leq y\} = F_{\boldsymbol{y}}(y)$$

can be defined. Further, considering the event that $\boldsymbol{x} \leq x$ and $\boldsymbol{y} \leq y$, the *joint distribution function* can be defined as follows:

$$F_{\boldsymbol{xy}}(x, y) = \Pr\{\boldsymbol{x} \leq x, \boldsymbol{y} \leq y\}$$

The *joint probability density function* $f_{\boldsymbol{xy}}(x, y)$ is defined as:

$$f_{\boldsymbol{xy}}(x, y) = \frac{\partial^2 F_{\boldsymbol{xy}}(x, y)}{\partial x \partial y}$$

$F_{\boldsymbol{x}}(x)$ is called the *marginal distribution function* of $\boldsymbol{x}$, and $F_{\boldsymbol{y}}(y)$ is the marginal distribution of $\boldsymbol{y}$. Similarly, $f_{\boldsymbol{x}}(x)$ is called the *marginal density function* of $\boldsymbol{x}$, and $f_{\boldsymbol{y}}(y)$ is the marginal density function of $\boldsymbol{y}$. The following relationships hold for the joint distribution function:

$$F_{\boldsymbol{xy}}(x, \infty) = F_{\boldsymbol{x}}(x) \quad , \quad F_{\boldsymbol{xy}}(\infty, y) = F_{\boldsymbol{y}}(y)$$

$$F_{\boldsymbol{xy}}(\infty, \infty) = 1 \quad , \; F_{\boldsymbol{xy}}(-\infty, y) = 0 \quad , \quad F_{\boldsymbol{xy}}(x, -\infty) = 0$$

$$\int_{-\infty}^{\infty} f_{\boldsymbol{xy}}(x, y) dy = f_{\boldsymbol{x}}(x) \text{ and } \int_{-\infty}^{\infty} f_{\boldsymbol{xy}}(x, y) dx = f_{\boldsymbol{y}}(y)$$

Further, the probability that the random variables $\boldsymbol{x}$ and $\boldsymbol{y}$ lie in intervals given by $[x_1, x_2]$ and $[y_1, y_2]$ is given by:

$$\Pr\{x_1 \leq \boldsymbol{x} \leq x_2 \text{ and } y_1 \leq \boldsymbol{y} \leq y_2\} = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f_{\boldsymbol{xy}}(x, y) dy dx$$

In general, when n random variables are considered, the joint distribution function of random variables $\boldsymbol{x_1, x_2, \ldots, x_n}$ is given by:

$$\Pr\{\boldsymbol{x_1} \leq x_1, \boldsymbol{x_2} \leq x_2, \ldots, \boldsymbol{x_n} \leq x_n\} = F_{\boldsymbol{x_1, x_2, \ldots, x_n}}(x_1, x_2, \ldots, x_n)$$

The probability density function is given by:

$$f_{\boldsymbol{x_1, x_2, \ldots, x_n}}(x_1, x_2, \ldots, x_n) = \frac{\partial^n F_{\boldsymbol{x_1, x_2, \ldots, x_n}}(x_1, x_2, \ldots, x_n)}{\partial x_1 \partial x_2 \ldots \partial x_n}$$

The marginal distribution function of random variable $\boldsymbol{x_i}$ is given by:

$$F_{\boldsymbol{x_i}}(x_i) = F_{\boldsymbol{x_1, x_2, \ldots, x_n}}(\infty, \infty, \ldots, x_i, \infty, \ldots, \infty)$$

The marginal density function of $\boldsymbol{x_i}$ is given by:

$$f_{\boldsymbol{x_i}}(x_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\boldsymbol{x_1, x_2, \ldots, x_n}}(x_1, x_2, \ldots, x_n) dx_1 dx_2 \cdots dx_{i-1} dx_{i+1} \cdots dx_n$$

Further, the expected value of a function $g(\boldsymbol{x_1, x_2, \ldots, x_n})$ is given by:

$$\mathrm{E}\{g(\boldsymbol{x_1, x_2, \ldots, x_n})\} =$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) f_{\boldsymbol{x_1, x_2, \ldots, x_n}}(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n$$

A.2.1 Moments of Jointly Distributed Random Variables

For two jointly distributed random variables $\boldsymbol{x}$ and $\boldsymbol{y}$, the i, jth joint moment is defined as follows:

$$m_{i,j} = \mathrm{E}\{\boldsymbol{x}^i \boldsymbol{y}^j\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^i y^j f_{\boldsymbol{xy}}(x, y) dx dy$$

Similarly, for the case of n random variables, $\boldsymbol{x_1}$,$\boldsymbol{x_2}$, ...,$\boldsymbol{x_n}$, the $i_1, i_2, \ldots, i_n$th moment is defined as:

$$m_{i_1, i_2, \ldots, i_n} = \mathrm{E}\{{\boldsymbol{x_1}}^{i_1} {\boldsymbol{x_2}}^{i_2} \ldots {\boldsymbol{x_n}}^{i_n}\} =$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} {x_1}^{i_1} {x_2}^{i_2} \ldots {x_n}^{i_n} f_{\boldsymbol{x_1, x_2, \ldots, x_n}}(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n$$

The *covariance* of two random variables, widely used in probabilistic analysis, is defined as

$$\mu_{1,1} = \mathrm{E}\{(\boldsymbol{x} - \eta_x)(\boldsymbol{y} - \eta_y)\}$$

The *correlation coefficient* of $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined as

$$r_{\boldsymbol{xy}} = \frac{\mu_{1,1}}{\sigma_x \sigma_y}$$

where $\eta_x, \eta_y, \sigma_x, \sigma_y$ are means and standard deviations of $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively.

The *covariance matrix*, $\boldsymbol{\Sigma}$, of n random variables is defined as a symmetric matrix with the following elements:

$$\Sigma_{i,j} = \mathrm{E}\{(\boldsymbol{x_i} - \eta_{x_i})(\boldsymbol{x_j} - \eta_{x_j})\}$$

The uncertainty in model parameters are very often provided in terms of the covariance matrix.

A.2.2 Dependence of Random Variables

Two random variables $\boldsymbol{x}$ and $\boldsymbol{y}$ are *uncorrelated* if

$$\mathrm{E}\{\boldsymbol{xy}\} = \mathrm{E}\{\boldsymbol{x}\}\mathrm{E}\{\boldsymbol{y}\}$$

They are *orthogonal* if

$$\mathrm{E}\{\boldsymbol{xy}\} = 0$$

and they are *independent* if

$$f_{\boldsymbol{xy}}(x, y) = f_{\boldsymbol{x}}(x) f_{\boldsymbol{y}}(y)$$

Similarly, n random variables, $\boldsymbol{x_1, x_2, \ldots, x_n}$, are uncorrelated if

$$\mathrm{E}\{\boldsymbol{x_i x_j}\} = \mathrm{E}\{\boldsymbol{x_i}\}\mathrm{E}\{\boldsymbol{x_j}\} \text{ for all } i \neq j$$

They are orthogonal if

$$\mathrm{E}\{\boldsymbol{x_i x_j}\} = 0 \text{ for all } i \neq j$$

and they are independent if

$$f_{\boldsymbol{x_1, x_2, \ldots, x_n}}(x_1, x_2, \ldots, x_n) = f_{\boldsymbol{x_1}}(x_1) f_{\boldsymbol{x_2}}(x_2) \ldots f_{\boldsymbol{x_n}}(x_n)$$

The interested reader can find additional material on probability concepts from the excellent texts on this subject [75, 114, 160, 201].1

Appendix B
BASIC PBPK MODEL EQUATIONS

A Physiologically Based PharmacoKinetic (PBPK) model for humans descibes the body as a set of interconnected compartments, or continuous stirred tank reactors (CSTRs). Each compartment can describe either an organ or a tissue. The set of differential equations of the PBPK model are derived through the mass balance across various compartmens (see Figure 5.1, on Page 79), as follows.

A mass balance around the equilibrium lung compartment results in:

$$c_{\text{arterial}} = \frac{Q_{\text{cardiac}} c_{\text{venous}} + Q_{\text{alveolar}} C_{\text{air(inhaled)}}}{Q_{\text{cardiac}} + Q_{\text{alveolar}} / P_{\text{blood/air}}}\,, \tag{B.1}$$

where,

$$c_{\text{venous}} = \frac{1}{Q_{\text{cardiac}}} \sum_{j=1}^{n} Q_j c_j \tag{B.2}$$

and Q_j and c_j are volumetric blood flow and concentration with respect to compartment j.

The mass balance on all compartments j in the PBPK model, other than the viable skin and stratum corneum compartments, is given by:

$$V_j \frac{dc_j}{dt} = Q_j \left(c_{\text{arterial}} - \frac{c_j}{P_{j/\text{blood}}} \right) - R_j. \tag{B.3}$$

The rate of metabolism, R_j, in any given compartment is given by:

$$R_j = \frac{V\text{max}_j c_{v,j}}{K\text{m}_j + c_{v,j}} \tag{B.4}$$

where $V\text{max}_j$ and $K\text{m}_j$ are Michaelis-Menten constants.

Roy [176] presents a thorough description of PBPK models for humans.

Appendix C

URBAN AIRSHED MODEL (UAM-IV) AND CARBON BOND MECHANISM (CB-IV)*

C.1 The Urban Airshed Model – UAM-IV

C.1.1 Conceptual Overview

The Urban Airshed Model is a three-dimensional photochemical grid model designed to calculate the concentrations of both inert and chemically reactive pollutants by simulating the physical and chemical processes in the atmosphere that affect pollutant concentrations [178]. The basis for the UAM is the atmospheric diffusion equation (ADE). This equation represents a mass balance in which all of the relevant emissions, transport, diffusion, chemical reactions, and removal processes are expressed as follows:

$$\frac{\partial C_i}{\partial t} + \frac{\partial (UC_i)}{\partial x} + \frac{\partial (VC_i)}{\partial y} + \frac{\partial (WC_i)}{\partial z}$$

$$= \frac{\partial}{\partial x}\left(K_H \frac{\partial C_i}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_H \frac{\partial C_i}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_V \frac{\partial C_i}{\partial z}\right) + \mathcal{R}_i + \mathcal{S} + \mathcal{L}_i \qquad \text{(C.1)}$$

where,

*adapted with permission from Wang [213] and from Walia [211]. A detailed description of the UAM can be found in the UAM-IV Users Guide by Morris and Myers [154]

$$
\begin{aligned}
i &= \text{chemical species } i \text{ (where } i = 1, 2, \ldots, i, \ldots, N) \\
C_i &= \text{pollutant concentration of species } i \\
u, v, w &= \text{horizontal and vertical wind speed components} = f(x, y, z, t) \\
K_H, K_V &= \text{horizontal and vertical turbulent diffusion coefficients} = f(x, y, z, t) \\
\mathcal{R}_i &= \text{rate of formation of } i \text{ by chemical reactions} = f(C_1, C_2, \ldots C_i, \ldots, C_N) \\
\mathcal{S} &= \text{emission rate of all precursors } P_1, P_2, \ldots, P_p, \\
&= S(x, y, z, t, P_1, P_2, \ldots, P_i, \ldots, P_p) \\
\mathcal{L}_i &= \text{net rate of removal of pollutant } i \text{ by surface uptake processes} \\
&= f(x, y, z, t)
\end{aligned}
$$

This model is typically applied to an 8- to 72- hour period to model air quality "episodes" – periods during which adverse meteorological conditions result in elevated pollutant concentrations of the chemical species of interest. UAM is mainly used to study the *photochemical air quality*, which pertains to air quality with respect to ambient ozone concentrations. High ozone concentrations in the ambient environment lead to adverse health effects [78, 181]. Ozone is primarily formed in the atmosphere through a complex chemical mechanism involving oxides of nitrogen (NO_x) and volatile organic compounds (VOCs) in the presence of sunlight.

The major factors that affect photochemical air quality include [223]:

- The spatial and temporal distribution of emissions of NO_x and VOC (both anthropogenic and biogenic),
- The composition of the emitted VOC and NO_x,
- The spatial and temporal variations in the wind fields,
- The dynamics of the atmospheric boundary layer, including stability and the level of mixing,
- The chemical reactions involving VOC, NO_x, and other important species, (the Carbon Bond Mechanism-IV described in detail in Table C.2 is implemented in the UAM),
- The diurnal variations of solar radiation and temperature,

- The loss of ozone and ozone precursors by dry deposition, and
- The ambient background of VOC, NO_x, and other species at the immediate outside of the boundary of the region of study, and above the region of study.

The UAM is typically run with a horizontal resolution of 2-5 km and a vertical resolution of 4-6 layers up to about 2 km above ground level. The region to be simulated is divided up into a three-dimensional grid covering the region of interest. Horizontal grid cells are rectangular with constant lengths in the x- and y-directions. Vertical layer thicknesses are defined by the user based on the diffusion break, the top of the region, the number of layers below and above the diffusion break and the minimum layer thickness. The diffusion break corresponds to the top of a mixed layer, either an unstable convective layer during the day (i.e. the mixing height) or a shallow mechanically mixed layer at night. The region top is usually defined at or slightly above the maximum daily diffusion break.

The UAM solves the atmospheric diffusion equation (ADE) using the method of fractional steps. The master or advection time step, determined by numerical stability criteria (grid size vs domain-maximum wind speed), is typically on the order of 5 min. In each advection time step, the terms in the equation that represent the different atmospheric processes (e.g. advection, chemistry or vertical diffusion) are solved separately using the most efficient numerical integration technique for the given process. The advection time step must be divided into multiple chemistry and vertical diffusion time steps to maintain numerical stability.

C.1.2 Treatment of Physical/Chemical Processes

Advective pollutant transport

Pollutants are transported primarily by advection, that is by the mean or bulk motion of the air. Advection in the UAM is treated by specifying horizontal wind fields (i.e. u and v wind components in each grid cell) for each vertical layer. The vertical wind velocity in the UAM terrain-following coordinate system can then be calculated from the conservation of mass equation. Proper specification of the hourly and three-dimensional

varying winds is one of the key steps to successful application of the UAM. The winds influence how different emissions are mixed together, advected downwind and diluted.

Turbulent diffusion

Turbulent diffusion (dispersion) is assumed to be proportional to the rate of change of concentration in space (i.e. the concentration gradient). The proportionality factor is termed the eddy diffusivity coefficient (K_x, K_y and K_z) in the ADE). Because it has been difficult to obtain precise measurements of the eddy diffusivity coefficients, theoretical estimates have been used. Control theory techniques are employed in conjunction with the results of a planetary boundary layer model to generate optimal diffusivity coefficients in the UAM [138, 139].

Atmospheric Chemistry

The UAM employs the Carbon Bond Mechanism (CBM-IV) for solving chemical kinetics of atmospheric chemistry. As implemented in the UAM, the CBM-IV contains 86 reactions and 35 species. The differential equations that describe the CBM-IV are a "stiff" system, that is, the equations contain wide variations in time (reaction rate) constants. Solving these equations with a "stiff" numerical integration scheme, such as the one developed by Gear [77], would result in prohibitively expensive computer time. Thus, the solution of the CBM-IV in the UAM uses quasi-steady-state assumptions (QSSA) for the low-mass fast-reacting species (i.e. the stiff species) and the more computationally efficient Crank-Nicholson algorithm for the remainder of the state species.

Surface Removal Processes

Many types of pollutants, including nitrogen oxides, ozone, carbon monoxide and sulfur compounds, are removed from the surface layer by such features as vegetation through the process of dry deposition. In the UAM, dry deposition is assumed to occur in a two-step process: the transfer of pollutants through the atmosphere to the surface and the uptake of the pollutants by vegetation and other materials at the surface. This process involves a resistance to mass transport and a resistance to surface removal. The transport resistance is estimated from theoretical considerations of turbulent transfer in

the atmospheric boundary layer. The surface resistance is obtained from experimental data on the uptake of pollutants by various surface features.

C.1.3 Applications of the UAM

Because the UAM accounts for spatial and temporal variations as well as differences in the reactivity/speciation of emissions, it is ideally suited for evaluating the effects of emission control scenarios on urban air quality. This is accomplished by first replicating a historical ozone episode to establish a base case simulation. Model inputs are prepared from observed meteorological, emission, and air quality data for a particular day or days. The model is then applied with these inputs and the results are evaluated to determine its performance.

Once the model results have been evaluated and determined to perform within prescribed levels, the same meteorological inputs and a projected emission inventory can be used to simulate possible future emission scenarios. That is, the model will calculate hourly ozone patterns likely to occur under the same meteorological conditions as the base case.

The first regulatory use and practical applications of the UAM were carried out for the Denver area on behalf of the Colorado Division of Highways and EPA's Region VII in 1978. The UAM was used to evaluate whether various transportation plans and programs were consistent with the SIP and to evaluate the effects on Denver's air quality of urban growth and development that might result from the construction of proposed wastewater treatment facilities.

In the late 1970s, EPA's OAQPS initiated a program to examine the applicability and practicality of the UAM in routine ozone attainment demonstrations required by the SIP process. Data collection, emission inventory development, model performance evaluation and application were major elements of this nation-wide program. Building off the St. Louis UAM applications and an extensive series of UAM sensitivity studies designed to provide guidance concerning the types and amounts of data required to support the UAM application, data for an application of the UAM, supported by OAQPS, were collected in Tulsa, Philadelphia/New Jersey, Baltimore/Washington and

New York.

Routine use of the UAM is an emerging practice. Since its first use for air quality planning in 1978, the UAM has been or is being used by the EPA, the Exxon Corporation, British Leyland, the TNO in the Netherlands, Pacific Gas and Electric, Southern California Edison Company, Arizona Department of Health, the South Coast (California) Air Quality Management District, the New York Department of Environmental Protection, the New Jersey Department of Environmental Protection, the California Air Resources Board, and the cities of Taipei and Kaohsiung in Taiwan.

C.2 Carbon Bond Mechanism (CB-IV)

Chemistry mechanisms included in most Photochemical Air Quality Simulation Models (PAQSMS) use reduced versions of known chemical mechanisms. Most chemical mechanisms are too big to be implemented efficiently in PAQSMS, as is, due to high computational overhead involved with their solution at each time-step. These mechanisms are lumped, or approximated by a smaller number of reactions, by one of several strategies: mathematical lumping, molecular lumping, or structural lumping. The Carbon Bond-IV Mechanism was developed mainly for urban smog and regional atmospheric modeling [154]. This photochemistry mechanism is implemented in the Urban Airshed Model (UAM) [154], and the Reactive Plume Model (RPM) [153, 195].

This mechanism is a hybrid of explicit chemistry, surrogate approximations, and lumped or generalized chemistry designed to simulate the features of urban smog chemistry. Explicit chemistry is used for the inorganic and Carbonyl species and the chemistries of Ethene, Isoprene, and Formaldehyde. The Ethene chemistry, for example, is treated explicitly because it reacts much slower than other alkenes, it is a large fraction of common hydrocarbon emissions, and that it yields a high fraction of Formaldehyde. Isoprene is an Alkene but it is treated explicitly also because its reactions with free radicals are much faster, and it forms a prominent part of biogenic emissions in rural areas. Many peroxy radicals have been lumped into a single XO_2 universal Peroxy radical. The lumping method of carbon-bonds is used mainly for paraffins and olefins. Molecular surrogates Toluene and Xylene are used for higher

aromatic compounds. For instance, a complex molecule with both aromatic and alkene structures might be represented with a combination of TOL, OLE, and PAR surrogates. Also, some of the rates and the stoichiometries of some incorporated reactions depend upon the atmospheric composition of reacting hydrocarbons.

The chemical species and surrogates that are explicitly represented in CBM-IV are listed in Table C.1. The hierarchical relationship between the major species in the carbon bond mechanism is depicted in Figure C.1. The most complex species, in terms of their oxidation products (paraffins, olefins, isoprene and aromatic hydrocarbons), are placed at the highest levels. The simplest species, in terms of molecular complexity (NOx, HOx, CO and formaldehyde), are placed at the lowest levels. Table C.2 presents the chemical reactions that constitute the CB-IV mechanism.

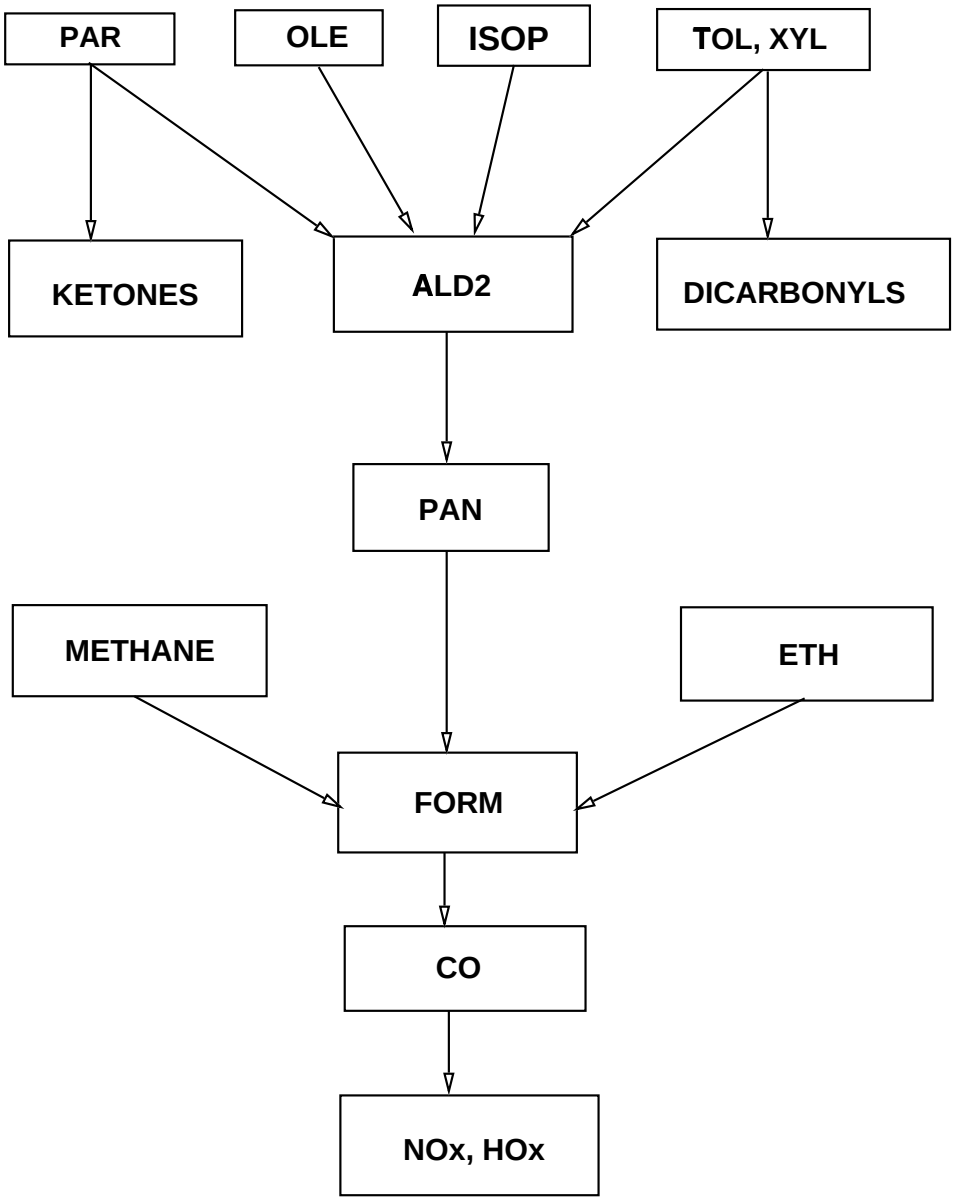


Figure C.1: Major species in the CBM and their hierarchical relationship (adapted from Wang [213])

Table C.1: Chemical Species in the CBM-IV mechanism

Species Name	representation
Nitric oxide	NO
Nitrogen dioxide	NO2
Nitrogen trioxide (nitrate radical)	NO3
Dinitrogen pentoxide	N2O5
Nitrous acid	HONO
Nitric acid	HNO3
Peroxy-nitric acid (HO_2NO_2)	PNA
Oxygen atom (singlet)	O1D
Oxygen atom (triplet)	O
Hydroxyl radical	OH
Water	H2O
Ozone	O3
Hydroperoxy radical	HO2
Hydrogen peroxide	H2O2
Carbon monoxide	CO
Formaldehyde ($CH_2=O$)	FORM
High molecular weight aldehydes	ALD2
Peroxyacyl radical ($CH_3C(O)OO\cdot$)	C2O3
Peroxyacyl nitrate ($CH_3C(O)OONO_2$)	PAN
Paraffin carbon bond (C-C)	PAR
Secondary organic oxy radical	ROR
Olefinic carbon bond	OLE
Ethene ($CH_2=CH_2$)	ETH
Toluene (C_6H_5-CH_3)	TOL
Cresol and higher molecular weight phenols	CRES
Toluene-hydroxyl radical adduct	TO2
Methylphenoxy radical	CRO
High molecular weight aromatic oxidation ring fragment	OPEN
Xylene (C_6H_4-$(CH_3)_2$)	XYL
Methylglyoxal ($CH_3C(O)C(O)H$)	MGLY
Isoprene	ISOP
NO-to-NO_2 operation	XO2
NO-to-nitrate operation	XO2N
Total	33

Table C.2: Chemical Reactions in the Carbon Bond IV Mechanism

Reaction No.	Reaction		
R1	$NO_2 + h\nu$	$\longrightarrow$	$NO + O$
R2	O	$\xrightarrow{O_2, M}$	O_3
R3	$O_3 + NO$	$\longrightarrow$	NO_2
R4	$O + NO_2$	$\longrightarrow$	NO
R5	$O + NO_2$	$\xrightarrow{M}$	NO_3
R6	$O + NO$	$\xrightarrow{M}$	NO_2
R7	$NO_2 + O_3$	$\longrightarrow$	NO_3
R8	$O_3 + h\nu$	$\longrightarrow$	O
R9	$O_3 + h\nu$	$\longrightarrow$	O^1D
R10	O^1D	$\xrightarrow{M}$	O
R11	$O^1D + H_2O$	$\longrightarrow$	$2\ OH$
R12	$O_3 + OH$	$\longrightarrow$	HO_2
R13	$O_3 + HO_2$	$\longrightarrow$	OH
R14	$NO_3 + h\nu$	$\longrightarrow$	$0.89\ NO_2 + 0.89\ O + 0.11\ NO$
R15	$NO_3 + NO$	$\longrightarrow$	$2\ NO_2$
R16	$NO_3 + NO_2$	$\longrightarrow$	$NO + NO_2$
R17	$NO_3 + NO_2$	$\xrightarrow{M}$	N_2O_5
R18	$N_2O_5 + H_2O$	$\longrightarrow$	$2\ HNO_3$
R19	N_2O_5	$\xrightarrow{M}$	$NO_3 + NO_2$
R20	$NO + NO$	$\xrightarrow{O_2}$	$2\ NO_2$
R21	$NO + NO_2\ H_2O$	$\longrightarrow$	$2\ HNO_2$
R22	$NO + OH$	$\xrightarrow{M}$	HNO_2
R23	$HNO_2 + h\nu$	$\longrightarrow$	$NO + OH$
R24	$OH + HNO_2$	$\longrightarrow$	NO_2
R25	$HNO_2 + HNO_2$	$\longrightarrow$	$NO + NO_2$
R26	$NO_2 + OH$	$\xrightarrow{M}$	HNO_3
R27	$OH + HNO_3$	$\xrightarrow{M}$	NO_3
R28	$HO_2 + NO$	$\longrightarrow$	$OH + NO_2$
R29	$HO_2 + NO_2$	$\xrightarrow{M}$	PNA
R30	PNA	$\xrightarrow{M}$	$HO_2 + NO_2$

CB-IV reactions – continued on next page

CB-IV reactions – continued from previous page

Reaction No.	Reaction		
R31	OH + PNA	$\longrightarrow$	NO_2
R32	HO_2 + HO_2	$\longrightarrow$	H_2O_2
R33	HO_2 + HO_2 H_2O	$\longrightarrow$	H_2O_2
R34	H_2O_2 + hν	$\longrightarrow$	2 OH
R35	OH + H_2O_2	$\longrightarrow$	HO_2
R36	OH + CO	$\xrightarrow{O_2}$	HO_2
R37	FORM + OH	$\xrightarrow{O_2}$	HO_2 + CO
R38	FORM + hν	$\xrightarrow{2O_2}$	2 HO_2 + CO
R39	FORM + hν	$\longrightarrow$	CO
R40	FORM + O	$\longrightarrow$	OH + HO_2 + CO
R41	FORM + NO_3	$\xrightarrow{O_2}$	HNO_3 + HO_2 + CO
R42	ALD2 + O	$\xrightarrow{O_2}$	C_2O_3 + OH
R43	ALD2 + OH	$\longrightarrow$	C_2O_3
R44	ALD2 + NO_3	$\xrightarrow{O_2}$	C_2O_3 + HNO_3
R45	ALD2 + hν	$\longrightarrow$	FORM + 2 HO_2 + CO + XO2
R46	C_2O_3 + NO	$\xrightarrow{O_2}$	FORM + NO_2 + HO_2 + XO2
R47	C_2O_3 + NO_2	$\longrightarrow$	PAN
R48	PAN	$\longrightarrow$	C_2O_3 + NO_2
R49	C_2O_3 + C_2O_3	$\longrightarrow$	2 FORM + 2 XO2 + 2 HO_2
R50	C_2O_3 + HO_2	$\longrightarrow$	0.79 FORM + 0.79 XO2 + 0.79 HO_2 + 0.79 OH
R51	OH	$\longrightarrow$	FORM + XO2 + HO_2
R52	PAR + OH	$\longrightarrow$	0.87 XO2 + 0.13 XO2N + 0.11 HO_2 + 0.11 ALD2 + 0.76 ROR - 0.11 PAR
R53	ROR	$\longrightarrow$	0.96 XO2 + 1.1 ALD2 + 0.94 HO_2 + 0.04 XO2N + 0.02 ROR - 2.1 PAR
R54	ROR	$\longrightarrow$	HO_2
R55	ROR + NO_2	$\longrightarrow$	
R56	O + OLE	$\longrightarrow$	0.63 ALD2 + 0.38 HO_2 + 0.28 XO2 + 0.3 CO + 0.2 FORM + 0.02 XO2N + 0.22 PAR + 0.2 OH
R57	OH + OLE	$\longrightarrow$	FORM + ALD2 - PAR + XO2 + HO_2
R58	O_3 + OLE	$\longrightarrow$	0.5 ALD2 + 0.74 FORM + 0.22 XO2 + 0.1 OH + 0.33 CO + 0.44 HO_2 - PAR

CB-IV reactions – continued on next page

CB-IV reactions – continued from previous page

Reaction No.		Reaction	
R59	NO_3 + OLE	$\longrightarrow$	0.91 XO2 + FORM + 0.09 XO2N + ALD2 + NO_2 - PAR
R60	O + ETH	$\longrightarrow$	FORM + 1.7 HO_2 + CO + 0.7 XO2 + 0.3 OH
R61	OH + ETH	$\longrightarrow$	XO2 + 1.56 FORM + 0.22 ALD2 + HO_2
R62	O_3 + ETH	$\longrightarrow$	FORM + 0.42 CO + 0.12 HO_2
R63	TOL + OH	$\longrightarrow$	0.44 HO_2 + 0.08 XO2 + 0.36 CRES + 0.56 TO2
R64	TO2 + NO	$\longrightarrow$	0.9 NO_2 + 0.9 HO_2 + 0.9 OPEN
R65	TO2	$\longrightarrow$	CRES + HO_2
R66	OH + CRES	$\longrightarrow$	0.4 CRO + 0.6 XO2 + 0.6 HO_2 + 0.3 OPEN
R67	CRES + NO_3	$\longrightarrow$	CRO + HNO_3
R68	CRO + NO_2	$\longrightarrow$	
R69	OH + XYL	$\longrightarrow$	0.7 HO_2 + XO2 + 0.2 CRES + 0.8 MGLY + 1.1 PAR + 0.3 TO2
R70	OPEN + OH	$\longrightarrow$	XO2 + 2 CO + 2 HO_2 + C_2O_3 + FORM
R71	OPEN + hν	$\longrightarrow$	C_2O_3 + HO_2 + CO
R72	OPEN + O_3	$\longrightarrow$	0.03 ALD2 + 0.62 C_2O_3 + 0.7 FORM + 0.03 XO2 + 0.69 CO + 0.08 OH + 0.76 HO_2 + 0.2 MGLY
R73	OH + MGLY	$\longrightarrow$	XO2 + C_2O_3
R74	MGLY + hν	$\longrightarrow$	C_2O_3 + HO_2 + CO
R75	O + ISOP	$\longrightarrow$	0.6 HO_2 + 0.8 ALD2 + 0.55 OLE + XO2 + 0.5 CO + 0.45 ETH + 0.9 PAR
R76	OH + ISOP	$\longrightarrow$	XO2 + FORM + 0.67 HO_2 + 0.13 XO2N + ETH + 0.4 MGLY + 0.2 C_2O_3 + 0.2 ALD2
R77	O_3 + ISOP	$\longrightarrow$	FORM + 0.4 ALD2 + 0.55 ETH + 0.2 MGLY + 0.1 PAR + 0.06 CO + 0.44 HO_2 + 0.1 OH
R78	NO_3 + ISOP	$\longrightarrow$	XO2N
R79	XO2 + NO	$\longrightarrow$	NO_2
R80	XO2 + XO2	$\longrightarrow$	
R81	XO2N + NO	$\longrightarrow$	
R82	XO2 + HO_2	$\longrightarrow$	

Appendix D
THE EPACMTP MODEL

D.1 Introduction

The EPA's Composite Model for leachate Migration and Transformation Products (EPACMTP) provides estimates of potential human exposure to hazardous chemicals leaching from land disposal facilities [64]. EPACMTP simulates the subsurface fate and transport of contaminants released from land disposal sites, and predicts the associated groundwater exposure in a domestic drinking water receptor well. This model is an improvement over the EPA's Composite Model for Landfills (EPACML) [62]. EPACML accounts for the first-order decay and sorption of chemicals, but disregards the formation and transport of transformation products. In addition, EPACML can describe only uniform, unidirectional groundwater flow. On the other hand, EPACMTP can take into consideration: (i) chain decay reactions and transport of daughter and grand-daughter products, (ii) effects of water-table mounding on groundwater flow and contaminant migration, (iii) finite source as well as continuous source scenarios, and (iv) metals transport.

EPACMTP consists of two modules: an unsaturated zone module called Finite Element and semi-analytical Contaminant Transport in the Unsaturated Zone (FECTUZ), and a saturated zone module called Combined Analytical-Numerical SAturated Zone in 3-Dimensions (CANSAZ-3D). FECTUZ is a one-dimensional model that simulates vertically downward steady-state flow and contaminant transport through the unsaturated zone above an unconfined aquifer. CANSAZ-3D simulates 3-D steady-state groundwater flow and transient or steady state contaminant transport. EPACMTP currently uses a simplified 2-D version of the CANSAZ-3D, and the modules are optimized for computational efficiency.

The following sections describe in detail the formulation and implementation of FECTUZ and CANSAZ-3D. The material presented here is summarized from the EPACMTP background document [64], which contains more detailed descriptions of the modules and the methods described here.

D.2 Description of the FECTUZ Module

The model for the Finite Element and semi-analytical Contaminant Transport in the Unsaturated Zone (FECTUZ) is a one-dimensional flow and transport code, that simulates vertical water flow and solute transport through the unsaturated zone above an unconfined aquifer. This model describes the migration of contaminants downward from a disposal unit, typically a landfill or surface impoundment, through the unsaturated zone underlying the disposal unit, to an unconfined aquifer. This model can simulate cases where the flow and transport are one-dimensional, in the downward vertical direction; the flow and transport are driven by the seepage from a landfill or a surface impoundment, which is assumed to occur at a constant rate. Flow is assumed to be always at steady state, while either transient or steady-state solute transport simulations can be performed. FECTUZ consists of two modules: a flow module and a solute transport module, and there are certain assumptions incorporated into both the unsaturated flow module and the unsaturated transport module.

The important assumptions in the flow module are as follows:

- the flow of the fluid phase is one-dimensional, isothermal, steady and governed by Darcy's law, and is not affected by the presence of dissolved chemicals, the fluid is slightly compressible and homogeneous,
- the soil profile consists of one or more, individually uniform, incompressible soil layers, and
- the effects of hysteresis in soil constitutive relations are negligible.

The important assumptions incorporated in the transport module are as follows:

- advection and dispersion are one-dimensional,
- transport in the porous medium is governed by Fick's law,

- sorption reactions can be described by Freundlich equilibrium isotherm,
- the effects of bio-chemical decay can be described by first-order degradation and zero-order production reactions, and
- fluid properties are independent of concentrations of contaminants.

A brief description of the mathematical formulation of the flow module and of the transport module is presented here. Additional details on the model formulation and the solution method can be found in the EPACMTP background document [64].

D.2.1 Steady State Flow Module Formulation

The governing equation for the steady state flow module of FECTUZ is given by Darcy's Law:

$$-K_s k_{rw}\left[\frac{d\psi}{dz}-1\right]=I \quad ; \quad \psi_l=0 \tag{D.1}$$

where ψ(L) is the pressure head, z(L) is the depth coordinate which is taken positive downward, K_s(L/T) is the saturated hydraulic conductivity, k_{rw} is the relative permeability, I(L/T) is the infiltration rate, and l(L) is the thickness of the unsaturated zone. The symbols inside the parentheses denote the dimensions of the physical quantities.

The permeability-water content relation is given assumed to follow the Mualem-van Genuchten model [205], and is given by

$$k_{rw}=S_e^{1/2}\left[1-\left(1-S_e^{1/\gamma}\right)^{\gamma}\right]^2 \tag{D.2}$$

where

$$S_e=\begin{cases}\left[1+(-\alpha\psi)^{\beta}\right]^{-\gamma} & , \quad \psi<0\\ 1 & , \quad \psi\geq 0\end{cases} \tag{D.3}$$

where the parameters α, β, and γ are soil specific shape parameters. Further, parameters β and γ are related through the equation $\gamma = 1 - 1/\beta$. Descriptive statistical values for α and β have been determined by Carsel and Parrish [30] for 12 soil classifications.

Solution Method for Flow Module

The solution of Equation D.1 involves the following steps:

- substitution of Equations D.2 and D.3 into Equation D.1,
- replacement of the derivative in Equation D.1 with a backward finite difference approximation, and
- solution of the resultant equation using a combined Newton-Raphson and bi-section method.

In this method, the unsaturated zone is discretized into a number of one-dimensional segments, with that a high resolution close to the water table and also close to layer interfaces for layered soils. The surface impoundment boundary condition is solved by using the Darcy's Law, and iteratively solving for the infiltration rate through the surface impoundment liner.

D.2.2 Solute Transport Module Formulation

The one-dimensional transport of solute species is modeled in FECTUZ using the following advection-dispersion equation:

$$\frac{\partial}{\partial z}\left[D\frac{\partial c_i}{\partial z}\right] - V\frac{\partial c_i}{\partial z} = \theta R_i \frac{\partial c_i}{\partial t} + \theta Q_i \lambda_i C_i - \sum_{m=1}^{M} \theta \xi_{im} Q_m \lambda_m c_m \qquad \text{(D.4)}$$

where c_i is the concentration of the ith species (M/L^3), D is the apparent dispersion coefficient (L^2/T), V is the Darcy velocity (L/T) obtained from the solution of the flow equation, R is the retardation factor, λ_i is the first order decay constant (1/T), Q is a coefficient to incorporate decay in the sorbed phase, and the summation on the right-hand side of Equation D.4 represents the production due to the decay of parent species, where M is the total number of parents. The coefficient ξ_{im} is a constant related to the decay reaction stoichiometry, representing the fraction of mth species that decays to the ith daughter species.

The dispersion coefficient D is defined as

$$D = \alpha_L V + \theta D^* \qquad \text{(D.5)}$$

where α_L is the longitudinal dispersivity (L) and D^* is the effective molecular diffusion coefficient (L^2/T). R signifies the effect of equilibrium sorption, whereas Q signifies the effect of decay. They are given by

$$\begin{aligned} R &= 1 + \frac{\rho_b}{\theta} k_1 \eta c^{\eta-1} \\ Q &= 1 + \frac{\rho_b}{\theta} k_1 c^{\eta-1} \end{aligned} \tag{D.6}$$

where k_1 and η are nonlinear Freundlich parameters.

The initial and boundary conditions of the one-dimensional transport problem can be given by one of the following equations:

$$\begin{aligned} c_i(z,0) &= c_i^{in} \\ -D\frac{\partial c_i}{\partial z}(0,t) &= v(c_i^0(t) - c_i) \\ c_I(0,t) &= c_i^0(t) \\ \frac{\partial c_i}{\partial t}(l,t) &= 0 \end{aligned} \tag{D.7}$$

where c_i^{in} is the initial concentration of the ith species, $c_i^0(t)$ is the leachate concentration emanating from the disposal facility, z is the downward vertical coordinate, and l is the depth of the water table.

Solution Method for Solute Transport Module

In the FECTUZ solute transport model, three solution methods are employed based on the complexity of the problem: (a) for single species, steady-state simulations involving a linear adsorption isotherm, an analytical solution is used, (b) for transient or steady-state decay chain simulations with linear sorption, a semi-analytical method is used, and (c) for cases involving nonlinear sorption, a numerical finite element solution is used.

D.3 Description of the CANSAZ-3D Module

The three-dimensional model, Combined Analytical-Numerical SAturated Zone 3-D (CANSAZ-3D), simulates the three-dimensional steady state groundwater flow and contaminant transport, and forms the saturated zone module of the EPACMTP. CANSAZ-3D describes advective-dispersive transport in an aquifer with a steady state flow field

and a patch source of contaminants at the water table. This model consists of two modules: (a) a module for the ground water flow in the saturated zone, and (b) a module for the contaminant migration through the saturated zone. The flow module estimates the hydraulic head and the flow velocities in an aquifer of constant thickness, whereas the transport module simulates the transport of dissolved contaminants, and estimates the contaminant concentrations at a receptor well. The flow module considers three-dimensional steady state flow, with an optional two-dimensional simulation. On the other hand, the transport module considers advection, hydrodynamic dispersion, equilibrium sorption, zero-order production, and first-order decay.

The important assumptions in the groundwater flow module are as follows:

- the aquifer is homogeneous
- the groundwater flow is steady-state, isothermal, and governed by Darcy's Law,
- the fluid is slightly compressible and homogeneous,
- the contribution of recharge from the unsaturated zone is small relative to the regional flow in the aquifer, and the saturated aquifer thickness is large relative to the head difference that establishes the regional gradient, and
- the principal directions of the hydraulic conductivity tensors are aligned with the Cartesian coordinate system.

The important assumptions in the solute transport module are as follows:

- the flow field is at a steady state,
- the aquifer is homogeneous and initially contaminant free,
- adsorption of contaminants onto solid phase is described by the Freundlich equilibrium isotherm,
- chemical/bio-chemical degradation of the contaminant can be described by a first-order process,
- the mass flux of contaminants through the source is constant,
- the chemical is dilute and is present in the solution or adsorbed onto the soil matrix,

- hydrodynamic dispersion can be described as a Fickian process, and
- the saturated zone is a chemically buffered system with constant and uniform geochemical characteristics.

A brief description of the groundwater flow module and of the transport module follows. Additional details can be found in the EPACMTP background document [64].

D.3.1 Formulation of the Groundwater Flow Module

The governing equation for steady state flow in three dimensions is

$$K_x \frac{\partial^2 H}{\partial^2 x} + K_y \frac{\partial^2 H}{\partial^2 y} + K_z \frac{\partial^2 H}{\partial^2 z} = 0 \tag{D.8}$$

where H is the hydraulic head (L), and K_x, K_y, and K_z are hydraulic conductivities (L/T). The boundary conditions are given by

$$\begin{aligned} H(0,y,z) &= H_1 \\ H(x_L,y,z) &= H_2 \\ \frac{\partial H}{\partial y}(x,0,z) &= 0 \\ \frac{\partial H}{\partial y}(x,\pm\frac{y_L}{2},z) &= 0 \\ \frac{\partial H}{\partial z}(x,y,0) &= 0, \qquad \text{and} \end{aligned} \tag{D.9}$$

$$-K_Z \frac{\partial H}{\partial z}(x,y,B) = \begin{cases} Ix_u \le x \le x_d, \quad -\frac{y_D}{2} \le y \le \frac{y_D}{2} \\ I_r \text{elsewhere} \end{cases}$$

where x_L, y_L and B are length, width and thickness of the aquifer system, x_d, x_u and y_D are the upstream and downstream coordinates and the width of the source, I is the infiltration rate through the rectangular surface patch source, and I_r is the recharge rate at the water table outside the patch area. The V_i's are the Darcy velocities obtained from the FECTUZ model, and are given by

$$V_l = -K_l \frac{dc}{dl}, \quad \text{where } l \text{ corresponds to } x, y \text{ or } z \tag{D.10}$$

These equations can be solved by either finite difference technique or finite element methods, by numerically discretizing the aquifer region of interest into three-dimensional elements. In addition, CANSAZ-3D can also simulate a 2-D groundwater flow in the $x-y$ plane. The present implementation of the EPACMTP model uses the finite difference technique for a 2-D solution. The details on the implementation of the finite element and finite difference techniques in CANSAZ-3D are presented in background document [64].

D.3.2 Formulation of the Saturated Zone Contaminant Migration Module

The three dimensional transport of contaminants in an aquifer can be described by

$$\begin{aligned} &\frac{\partial}{\partial x_i}\left(D_{ij}\frac{\partial c_p}{\partial x_j}\right) - V_i\frac{\partial c_p}{\partial x_i} \\ &= \phi Q_p \lambda_p c_p + \phi R_p \frac{\partial c_p}{\partial t} - \phi \sum_{m=1}^{M} \xi_{pm} Q_m \lambda_m c_m, \qquad \begin{aligned} p &= 1, n_c \\ i,j &= 1,2,3 \end{aligned} \end{aligned} \tag{D.11}$$

where c_p (M/L^3) is the concentration of the pth component species in the n_c member decay chain, λ_p (1/T) and R_p are the first order decay and retardation coefficients, Q_p and Q_m are correction factors to account for sorbed phase decay of species p and parent m, respectively, and ϕ is the aquifer effective porosity. The dispersion coefficients D_{ij} are given by

$$\begin{aligned} D_{ii} &= \alpha_{_L}\frac{V_i^2}{|V|} + \alpha_{_T}\frac{V_j^2}{|V|} + \alpha_{_V}\frac{V_k^2}{|V|} + \phi D^* \\ D_{ij} &= (\alpha_{_L} - \alpha_{_T})V_i V_j/|V| \end{aligned} \tag{D.12}$$

where $\alpha_{_L}, \alpha_{_T}$ and $\alpha_{_V}$ are the longitudinal, horizontal transverse and vertical dispersivity (L) respectively, and D^* is the effective molecular diffusion coefficient (L^2/T). The retardation factor R and the decay coefficient Q are given as in Equation D.6 in Appendix D.2.2. The factors ξ_{pm}, Q_m and λ_m are the same as described in Appendix D.2.2. Here i, j and k denote the x, y and z directions respectively, and the Einstein summation convention is used to simplify the notation.

The initial and boundary conditions for the problem are as follows:

upstream boundary: $c_p = 0$ $\quad x = 0,\ -\frac{y_L}{2} \le y \le \frac{y_L}{2},\ 0 \le z \le B$

downstream boundary: $\frac{\partial c_p}{\partial x} = 0$ $\quad x = x_L,\ -\frac{y_L}{2} \le y \le \frac{y_L}{2},\ 0 \le z \le B$

left boundary: $\frac{\partial c_p}{\partial y} = 0$ $\quad 0 \le x \le x_L,\ y = -\frac{y_L}{2},\ 0 \le z \le B$

right boundary: $\frac{\partial c_p}{\partial y} = 0$ $\quad 0 \le x \le x_L,\ y = \frac{y_L}{2},\ 0 \le z \le B$

bottom boundary: $\frac{\partial c_p}{\partial z} = 0$ $\quad x_u \le x \le x_d,\ -\frac{y_D}{2} \le y \le \frac{y_D}{2},\ z = 0$

top boundary: $c_p(x, y, z, 0) = c_p^0(t)$ $\quad x_u \le x \le x_d,\ -\frac{y_D}{2} \le y \le \frac{y_D}{2}$

and $\left[-D_{zz}\frac{\partial c_p}{\partial z} + V_z c_p\right]_{z=B} = 0$ $\quad x < x_u,\ x > x_d,\ y < \frac{y_D}{2},\ y > \frac{y_D}{2}$

Solution Method for the Transport Module

For steady state cases, the concentration in a receptor well is estimated by applying the de Hoog algorithm [47]. For time dependent simulations, the governing equation is solved by using the Laplace Transform Galerkin (GLT) technique, developed by Sudicky [196]. CANSAZ-3D has options for either a finite difference or a finite element solution. A detailed description of the solution method can be found in the background document [64].

Appendix E
PROGRAMS INCLUDED IN THE CDROM

A list of the programs that are required for the application of the SRSM, the SRSM-ADIFOR, the Web Based SRSM tool, and the RPM-3D, are presented here. Please note that the application of the SRSM requires some system dependent files and auxiliary software such as plotting utilities, whereas the application of the SRSM-ADIFOR requires the ADIFOR package. In addition, the SRSM and SRSM-ADIFOR require some standard numerical modules that are commercially available for a low cost such as Numerical Recipes [163]. These are not included here either because they cannot be redistributed (e.g., Numerical Recipes is a commercial package, whereas ADIFOR is free, but requires permission for use), or they are system dependent (e.g., the plotting utilities used here are system dependent). However, the code included here is easily configurable to work on any system that has plotting packages and a web server.

Table E.1: Programs required for the implementation of the SRSM

Stochastic Response Surface Method (SRSM) – Directory: srsm	
File	Description
`srsmmain.for`	Main program for the SRSM
`srsmsub.for`	Auxiliary subroutines required for the calculation of polynomial chaos expansions, and the calculation of unknown coefficients
`collsub.for`	Contains subroutines for collocation point generation
`recipes.for`	From Numerical Recipes [163] (not included here due to copyright considerations) subroutines required are: `GASDEV, LUBKSB, LUDCMP, PYTHAG, RAN2, SVBKSB, SVDCMP`)
`srsmcfg.h`	Configuration variables for the SRSM
`makefile`	Make file for the SRSM program
`chaos.map`	Contains Maple code for the generation of polynomial chaos expansions and auxiliary modules
requires a plotting program and a web server	

Table E.2: Programs required for the web based SRSM tool

Web Based SRSM – Directory: srsmweb	
File	Description
`srsm.html`	A web page for the SRSM
`srsm`	CGI-program for the SRSM web interface
`srsm-form.cgi`	Generates the collocation points based on input/output specification
`srsm-post-proc.cgi`	Calculates the coefficients of polynomial chaos expansion and plots the output *pdf*s
`srsm-form-errors.pl`	Error handler for input/output specification
`srsm-post-proc.pl`	Auxiliary Perl program for post processing

Table E.3: Programs required for the implementation of the SRSM-ADIFOR

SRSM-ADIFOR – Directory: srsmadi	
File	Description
`senscoll.for`	Collocation generation module
`polygen.for`	Polynomial chaos expansion routines
`polysens.for`	Modules for derivatives of polynomial chaos expansions
`sensmain.for`	Main program for the SRSM-ADIFOR
`recipes.for`	files from Numerical Recipes
`makefile`	Make file for the SRSM-ADIFOR program
requires the ADIFOR package	

Table E.4: Other Programs and Input Files

RPM – Directory: rpm4	
File	Description
`rpmgear.for`	RPM-IV code
`mercer.in`	RPM-IV input file for the Mercer County case study
`marathon.in`	RPM-IV input file for the Marathon Refinery case study
RPM-3D – Directory: rpm3d	
`makefile`	Makefile containing the list of files (45 files, all included in the folder)
`mercer.in`	RPM-3D input file for the Mercer County case study
`marathon.in`	RPM-3D input file for the Marathon Refinery case study

Vita

Sastry S. Isukapalli

1993	B.Tech. (Chemical Engineering), Indian Institute of Technology, Madras, INDIA.
1993-1994	Teaching Assistant, Department of Chemical Engineering, Rutgers, The State University of New Jersey.
1995	M.S. (Chemical Engineering), Rutgers, The State University of New Jersey.
1994-1997	Research Assistant, Environmental and Occupational Health Sciences Institute, a Joint Project of UMDNJ and Rutgers, The State University of New Jersey.
1997-1998	Research Teaching Specialist, Environmental and Occupational Health Sciences Institute, a Joint Project of UMDNJ and Rutgers, The State University of New Jersey.
1999	Ph.D. (Chemical Engineering), Rutgers, The State University of New Jersey.