EPA ###-#-##-###
Publication Number
Publication Number
www.epa.gov/crem
August 2008

Office of the Science Advisor

Guidance on the Development, Evaluation and Application of Environmental Models

Council for Regulatory Environmental Modeling
U.S. Environmental Protection Agency
Washington, DC 20460

Preface

This Guidance on the Development, Evaluation and Application of Environmental Models was prepared in response to the EPA Administrator's request that the EPA Council for Regulatory Environmental Modeling (CREM) helps to continue to strengthen EPA's development, evaluation and use of models (http://www.epa.gov/osp/crem/library/whitman.PDF).

An independent panel of experts established by EPA's Science Advisory Board reviewed the draft version of this document (http://cfpub.epa.gov/crem/crem_sab.cfm). In response to comments from the Board, the CREM made changes to the draft document.

This document is available in printed and electronic form. The electronic version provides direct links to the cited and other useful references identified in the document.

Disclaimer

This document provides guidance to those who develop, evaluate, and apply environmental models. It does not impose legally binding requirements and, depending on the circumstances, may not apply to a particular situation. The Environmental Protection Agency (EPA) retains the discretion to adopt approaches that differ from this guidance on a case-by-case basis.

Authors, Contributors and Reviewers

This document was developed under the leadership of EPA's Council for Regulatory Environmental Modeling. A number of individuals representing EPA's Core, Program and Regional offices have contributed to the authorship and review of this document.

PRINCIPAL AUTHORS:

Council for Regulatory Environmental Modeling:
Noha Gaber, Pasky Pascual, Neil Stiber, Elsie Sunderland

EPA Region 10:
Ben Cope

Office of Environmental Information:
Annett Nold

CONTRIBUTORS AND INTERNAL REVIEWERS:

EPA Core Offices:
Office of Research and Development:
Justin Babendreier, Thomas Barnwell (Retired), Ed Bender, Lawrence Burns (Retired), Gary Foley, Kathryn Gallagher, Kenneth Galluppi, Gerry Laniak, Haluk Ozkaynak, Kenneth Schere, Subhas Sikdar, Eric Weber, Joe Williams

Office of Environmental Information:
Ming Chang, Reggie Cheatham, Evangeline Cummings, Linda Kirkland, Nancy Wentworth

Office of General Counsel:
James Nelson (Retired), Barbara Pace, Manisha Patel, Carol Ann Sicilano

Science Advisory Board:
Jack Kooyoomjian

EPA Program Offices:
Office of Air and Radiation:
John Irwin (Retired), Joe Tikvart, Richard (Chet) Wayland, Jason West

Office of Prevention, Pesticides and Toxic Substances:
Lynn Delpire, Alan Dixon, Wen-Hsiung Lee, David Miller, Vince Nabholz, Steve Nako, Neil Patel, Randolph Perfetti (Retired), Scott Prothero, Donald Rodier

Office of Solid Waste and Emergency Response:
Peter Grevatt, Stephen Kroner, Larry Zaragoza

Office of Water:
Sharon E. Hayes, Marjorie Wellman, Denise Keehner, Lauren Wisniewski, Lisa McGuire, Mike Messner, James F. Pendergast

EPA Regional Offices:
Region 1:
Brian Hennessey

Region 2:
Kevin Bricke, Rosella O'Connor, Richard Winfield

Region 3:
Alan Cimorelli

Region 4:
Nancy Bethune, Brenda Johnson, Tim Wool

Region 5:
Bertram Frey, Arthur Lubin, Randy Robinson, Stephen Roy, Mary White

Region 6:
James Yarborough

Region 7:
Bret Anderson

Region 10:
David Frank, John Yearsley

Executive Summary

In pursuing its mission to protect human health and to safeguard the natural environment, the U.S. Environmental Protection Agency often relies on environmental models. In this Guidance, a model is defined as a "*simplification of reality that is constructed to gain insights into select attributes of a particular physical, biological, economic, or social system."* This Guidance provides recommendations for the effective use of models in environmental decision making. These recommendations are drawn from Agency white papers, EPA's Science Advisory Board reports, the National Research Council Report on Models in Environmental Regulatory Decision Making and peer-reviewed literature. For organizational simplicity, these recommendations are categorized into the following sections: *model development*, *model evaluation*, and *model application*.

Model Development can be viewed as a process that is achieved by following four main steps: (a) identify the environmental issue (or set of issues) that the model is intended to address; (b) develop the conceptual model; (c) construct the model framework (develop the mathematical model), and (d) parameterize the model to develop the application tool.

Model Evaluation is the process for generating information over the life cycle of the project that helps to determine whether a model and its analytical results are of a quality sufficient to serve as the basis for a decision. Model quality is an attribute that is meaningful only within the context of a specific model application. In simple terms, model evaluation provides information to help assess the following factors: (a) How have the principles of sound science been addressed during model development? (b) How is the choice of model supported by the quantity and quality of available data? (c) How closely does the model approximate the real system of interest? (d) How well does the model perform the specified task while meeting the objectives set by QA project planning?

Model Application, (i.e., model-based decision making), is strengthened when the science underlying the model is transparent. The elements of transparency emphasized in this Guidance are: (a) comprehensive documentation of all aspects of a modeling project (suggested as a list of elements relevant to any modeling project) and (b) effective communication between modelers, analysts, and decision makers. This approach ensures that there is a clear rationale for using a model for a specific regulatory application.

This Guidance recommends best practices to help determine when a model, despite its uncertainties, can be appropriately used to inform a decision. Specifically, it recommends that model developers and users: (a) subject their model to credible, objective peer review; (b) assess the quality of the data they use; (c) corroborate their model by evaluating the degree to which it corresponds to the system being modeled; and (d) perform sensitivity and uncertainty analyses. Sensitivity analysis evaluates the effect of changes in input values or assumptions on a model's results. Uncertainty analysis investigates the effects of lack of knowledge and other potential sources of error in the model (e.g., the "uncertainty" associated with model parameter values) and when conducted in combination with sensitivity analysis allows a model user to be more informed about the confidence that can be placed in model results. A model's quality to support a decision becomes better known when information is available to assess these factors.

Table of Contents

1. Introduction

1.1 Purpose and Scope of this Document

The U.S. Environmental Protection Agency (US EPA) uses a wide range of models to inform decisions that supports its mission of protecting human health and safeguarding the natural environment — air, water, and land — upon which life depends. These models includes atmospheric and indoor air models, chemical equilibrium models, economic models, exposure models, leaching and runoff models, multi-media models, risk assessment models, ground water and surface water models, and toxicokinetic models. These models range from simple to complex and may employ a combination of scientific, economic, socio-economic, or other types of data. As stated in the National Research Council (NRC) report on *Models in Environmental Regulatory Decision Making*, models are critical to the regulatory decision making process because the spatial and temporal scales linking environmental controls and environmental quality generally do not allow for an observational approach to understand the relationship between economic activity and environmental quality (NRC 2007). Models have a long history of helping to explain scientific phenomenon and of predicting outcomes and behavior in settings where empirical observations are limited or not available. For the purposes of this guidance, we use the NRC (NRC 2007) definition of a model as:

> *"A simplification of reality that is constructed to gain insights into select attributes of a particular physical, biological, economic, or social system."*

We further focus the content of this guidance on the subset of all models termed "computational models" by the NRC, which are models that use measurable variables, numerical inputs and mathematical relationships to produce quantitative outputs.

As models become increasingly significant in decision making, it is important that model development and application processes conform to appropriate protocols or standards that help ensure the utility, scientific soundness and defensibility of models and their outputs for decision making. It is also increasingly important to plan and manage the process of using models to inform decision making (Manno et al., 2008). This Guidance Document (hereon referred to as "guidance") aims to facilitate a widespread understanding of models and thereby promote their appropriate application to support informed decision making. The principles and practices described in the guidance are intended to be generally applicable to all models used to inform Agency decisions, regardless of domain, mode, conceptual basis, or form or level of rigor (i.e. varying from screening level applications to complex analyses) [USEPA 2001]. While the guidance includes discussion of the links between the modeling and decision processes, many of the recommendations apply equally to non-regulatory modeling projects.

This guidance presents recommendations that are drawn from Agency white papers on environmental modeling, EPA Science Advisory Board (SAB) reports, the National Research Council Report on Models in Environmental Regulatory Decision Making and peer-reviewed literature. It provides an overview of *best practices* for evaluating the quality of environmental models.

These practices, in turn, support the execution of mandatory Agency quality assurance (QA) processes for planning, implementing, and assessing of modeling projects that produce quality documentation (Box B1: Background on EPA Quality System). QA plans should contain performance criteria or ("specifications") for a model in the context of their intended use that are developed at the onset of each project. When documented using a series of associated tests of model quality ("checks") during model evaluation, these specifications provide a record of how well a model meets its intended use. These checks are the basis for a decision on model acceptability. The purpose of this guidance is to provide specific advice on how these "checks" are best performed throughout the development, evaluation and application of models. Following the best practices emphasized in this document in conjunction with a well-documented QA project plan helps to ensure that information EPA disseminates in support of model development and decisions that are informed by models heed the principles set by the Agency's Information Quality Guidelines (USEPA 2002a). Recognizing the diversity of modeling applications throughout the Agency, the following chapters are intended to outline some general principles that support the development process. These principles complement the systematic QA planning process for modeling projects that is outlined in existing guidance (USEPA 2002b). The principles presented in this guidance are also applicable to models not used for regulatory purposes as experience has shown that models developed for research and development have often found useful applications in environmental management purposes.

All underlined terms in this document are defined in the Glossary (Appendix A).

1.2 Intended Audience

This document is intended for a wide range of audiences, including model developers, computer programmers, model users, policy makers who work with models and affected stakeholders. Model users may include those who generate model output (i.e., those who set up, parameterize and run models) and those who are managers and are primarily users of model output. The main body of this document provides a general overview of principles of good modeling for all users (see Figure 1 for a depiction of the steps in modeling process). The appendices contain technical information and examples that may be more appropriate for specific user groups.

1.3 Organizational Framework

For organizational simplicity, the main body of this guidance is divided into three sections on model development, model evaluation and model application. However, evaluating a model and its input data to ensure their quality is in fact a process that should be undertaken and documented during all stages of model development and application.

Chapter 1 serves as a general introduction and outlines the scope of this guidance. Chapter 2 discusses the role of models in environmental decision making. Chapters 3 and 4 provide guidance on elements of model development and evaluation, respectively. Finally, Chapter 5 recommends practices for most effectively incorporating information from environmental models into policy decisions made by the Agency. The role of information from models in Agency decisions is illustrated in Figure 1. Several appendices referred to in the text contain more detailed technical information and examples that complement each of these chapters. As mentioned above, Appendix A is a glossary with definitions for all of the underlined terms in the guidance. Appendix A-2 summarizes the categories of models that are integral to environmental regulation. Appendix B presents additional background information on the QA program and other relevant topics. Appendix C presents an overview of best practices that may be used to evaluate models including more detailed information on the peer review process for models and specific technical guidance on tools for model evaluation.

1.4 Appropriate Implementation of this Document

Program and regional offices may rely on the details of this document, while modifying and clarifying recommendations, as appropriate and necessary. Each EPA office should be responsible for implementing the best practices described in this document in an appropriate manner to meet its needs.

As indicated by the use of non-mandatory language such as “may,” “should,” and “can,” this document provides recommendations and suggestions and does not create legal rights or impose legally binding requirements on EPA or the public.

The Models Knowledge Base is a companion product that was developed by the Council for Regulatory Environmental Modeling to complement this document. It is designed as a web-based inventory of information on models used in EPA. It provides convenient access to standardized documentation of the models’ development, scientific basis, user requirements, evaluation studies and application examples.

2. Modeling for Environmental Decision Support

2.1 Why are Models Important?

This guidance defines a model as "***a simplification of reality that is constructed to gain insights into select attributes of a particular physical, biological, economic, or social system***". However, a model developer sets boundary conditions, determines which aspects of the system are to be modeled, which processes are of importance, how these processes may be represented mathematically and what computational methods to use in implementing the mathematics. Models are thus based on simplifying assumptions and cannot be expected to completely replicate the complexity inherent in open environmental systems. Notwithstanding these limitations, models are essential for a variety of purposes in the environmental field. Fundamentally, models are used to investigate diagnostics (to assess *what happened*); process (to assess the causes and precursor conditions of *why did it happen*); or prognostics/ prediction (to forecast outcomes and future events or *what will happen*). Models, whether used to provide better understanding and characterization of complex systems under current conditions or under envisioned future circumstances or scenarios, play an important role in environmental management. Many environmental and human health questions cannot be addressed solely through empirical means. Models may be used to facilitate the analysis of systems that as a result of their complexity or spatial or temporal variability may not feasibly be isolated in experiments in the real world.

A variety of criteria may be used to construct classifications of models (See Appendix A-2). Models may be classified based on their conceptual basis and mathematical solution, the purpose for which they were developed and are applied, the domain or discipline that they apply to and the level of resolution and complexity at which they operate. For example, three categories of regulatory models have been identified, based on their purpose or application (CREM 2001):

(1) models used for policy analysis: the results of *policy analysis* models affect national policy decisions; they are used to set policy for large, multi-year programs or concepts, for example in the development of national policy on acid rain and phosphorus reduction in the Great Lakes;

(2) models used for national regulatory decision making: models used to inform *national regulatory* decision making where overall policy has been established include the use of a model to assist in determining the national decision regarding the regulation of a specific pesticide or using a model to aid in the establishment of national effluent limitations.

3) models used for implementation applications: models used for *implementation applications* apply to situations where policies and regulations have already been made. The development of these models and the use of their output may be driven by court-ordered schedules and the need for local actions.

Environmental models are one source of information for decision makers who need to consider many competing objectives. Within the Agency, models are used to simulate many different processes: natural (chemical, physical, and biological) systems, economic phenomena, and decision processes and many different types of models are employed, including economic, behavioral, physical, engineering design, health, ecological, and fate/transport models. A number of EPA programs make decisions based on information from environmental modeling applications, and the geographic scale of the problems addressed by a model can vary from the national scale to the individual site scale. Examples at different scales include:

- National air quality models used in decisions regarding emission requirements
- Watershed-scale water quality models used in decisions regarding permit limits for point sources
- Site-scale human health risk models used in decisions regarding hazardous waste cleanup measures

Box 1: Examples of EPA's Websites Containing Model Descriptions for Individual Programs

National Environmental Research Laboratory Models Web Site: http://www.epa.gov/nerl/topics/models.html
Atmospheric Sciences Modeling Division Web Site: http://www.epa.gov/asmdnerl/index.html
Office of Water's Water Quality Modeling Web Site: http://www.epa.gov/waterscience/wqm
Center for Subsurface Modeling Support Web Site: http://www.epa.gov/ada/csmos.html
National Center for Computational Toxicology Web Site: http://www.epa.gov/ncct
Support Center for Regulatory Atmospheric Modeling: http://www.epa.gov/scram001/aqmindex.htm

Models also have a number of other useful applications outside of the regulatory context. For example, because models include explicit mathematical statements about system mechanics, they serve as research tools for exploring new scientific issues and screening tools for simplifying and/or refining existing scientific paradigms or software (SAB 1993a, SAB 1989). Models can also help to study the behavior of ecological systems, design field studies, interpret data, and generalize results.

2.2 The Modeling Life-Cycle

The process of model development and application to address a specific decision making need can be viewed as generally following the iterative progression that is depicted in an idealized form in Figure 1 with further information provided in Box 2.

Box 2: Basic Steps in the Modeling Development Process (modified from Box 3-1 NRC Report on Models in Regulatory Environmental Decision Making)		
Model Development Step		**Modeling Issues**
Problem Identification: *to determine the right decision-relevant questions and establish modeling objectives*	Definition of Model Purpose	▪ Goal ▪ Decisions to be supported ▪ Predictions to be made
	Specification of Modeling Context	▪ Scale (spatial and temporal) ▪ Application domain ▪ User community ▪ Required inputs ▪ Desired output ▪ Evaluation criteria
Model Development: *to develop the conceptual model that reflects the underlying science of the processes being modeled; and develop the mathematical representation of that science and encoding these mathematical expressions in a computer program.*	Conceptual Model Formulation	▪ Assumptions (dynamic, static, stochastic, deterministic) ▪ State variables represented ▪ Level of process detail necessary ▪ Scientific foundations
	Computational Model Development	▪ Algorithms ▪ Mathematical/computational methods ▪ Inputs ▪ Hardware platforms and software infrastructure ▪ User interface ▪ Calibration/parameter determination ▪ Documentation
Model Evaluation: *to test that the model expressions have been encoded correctly into the computer program and test the model outputs by comparing it with empirical data*	Model Testing and Revision	▪ Theoretical corroboration ▪ Model components verification ▪ Corroboration (independent data) ▪ Sensitivity analysis ▪ Uncertainty analysis ▪ Robustness determination ▪ Comparison to evaluation criteria set during formulation
Model Application: *running the model and analyzing the model outputs to inform a decision.*	Model Use	▪ Analysis of Scenarios ▪ Predictions evaluation ▪ Regulations assessment ▪ Policy analysis and evaluation ▪ Model post-auditing

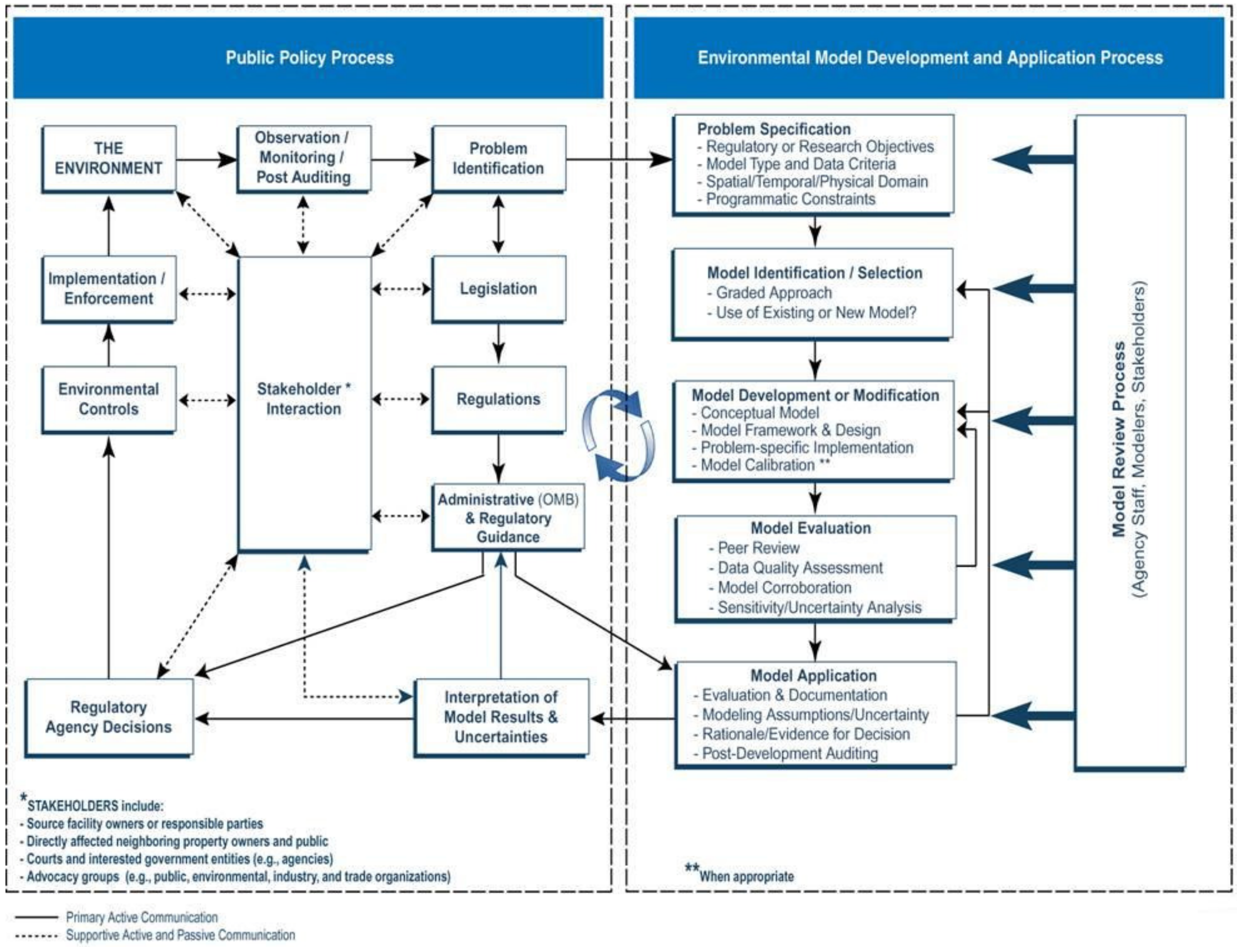


Figure 1: The Role of Modeling in the Public Policy Process

3. Model Development

Summary of Recommendations for Model Development

- Evaluation of a regulatory model should continue throughout the life of a model.
- Communication between model developers and model users is crucial at the problem identification stage
- Present a clear statement and description (in words, functional expressions, diagrams, and graphs, as necessary) of each element of the conceptual model. For each element, the modeler should document the science behind the conceptual model.
- When possible, test simple competing conceptual models/hypotheses.
- Use sensitivity analysis early and often.
- Determine the optimal level of model complexity by making appropriate tradeoffs among competing objectives.
- Where possible, model parameters should be characterized using direct measurements of sample populations.
- All input data should meet data quality acceptance criteria in the QA project plan for modeling.

3.1 Problem Identification

Models are developed in order to address real or perceived environmental problems. The first problem statement may be expressed in general terms that are easily understood by the general public. In the first stage of a modeling assessment, model developers and program staff build a more detailed specification of the problem to inform the model selection and development process. Communication between model developers and model users is crucial at this stage to clearly establish the objectives of the modeling process as ambiguity at the beginning of the process can undermine the chances for success (Manno et al 2008).

3.2 Problem Specification

Problem specification is best accomplished as a collaborative effort among model developers, intended users, and decision makers. In this initial stage of a project, the project team should work in an iterative process that identifies all of the aspects of the problem that will inform model selection and development. Problem specification should address the following components:

- **Identify the regulatory or research objectives**: statements of what questions a model has to answer. The statement of modeling objectives should include the state variables of concern, the stressors (model inputs) driving those state variables and their control options, appropriate temporal and spatial scales, user acceptance of the model, and the degree of accuracy and precision of the model.
- **Decide on the type of model or models to use:** Discuss alternatives and compare different types of models, e.g. empirical vs. mechanistic, static vs. dynamic, simulation vs. optimization, deterministic vs. stochastic, lumped vs. distributed
- **Specify the model domain characteristics:** identification of the environmental domain being modeled; specification of transport and transformation processes within that domain that are relevant to the policy/management/research objectives; specification of important time and space scales

inherent in transport and transformation processes within that domain in comparison with the time and space scales of the problem objectives; and any peculiar conditions of the domain that will affect model selection or new model construction.

- **Discuss the factors that could potentially constrain the modeling process:** time and budget, available data or resources to acquire more data, legal and institutional considerations, computer resource constraints, and experience and expertise of the modeling staff.

Once the need for a model has been identified, a conceptual model should be developed to represent the most important behaviors of the object or process relevant to the problem of interest. Literature, fieldwork, applicable anecdotal evidence, and relevant historical modeling projects may be considered in developing the conceptual model. The model developer should present a clear statement and description (in words, functional expressions, diagrams, and/or graphs) of each element of the conceptual model. For each element, the modeler should document the science behind the conceptual model (e.g., laboratory experiments, mechanistic evidence, empirical data supporting the hypothesis, peer reviewed literature) in mathematical form, when possible. To the extent feasible, the modeler should provide information on assumptions, scale, feedback mechanisms, and static/dynamic behaviors. When it is relevant, an appraisal of the strengths and weaknesses of each constituent hypothesis should be provided.

Using the conceptual model, the model developer can specify the relevant characteristics of the system to be modeled. This process includes: identification of the environmental domain being modeled; specification of transport and transformation processes within that domain that are relevant to the policy/management/research objectives; specification of important time and space scales inherent in transport and transformation processes within that domain in comparison with the time and space scales of the problem objectives; and any peculiar conditions of the domain that will affect model selection or new model construction.

Problem specification should also include a discussion of the potential programmatic considerations and constraints. These include regulatory requirements, time and budget, available data or resources to acquire more data, computer resource constraints, and experience and expertise of the modeling staff.

A "graded approach" should be applied in the specification of required model features. This approach involves an evaluation of the scope, rigor, and complexity of the modeling analysis in light of the intended use of results and degree of confidence needed in the results. A similar graded approach is used in subsequent stages in the model development and application process.

It is useful to qualitatively or quantitatively specify the acceptable range of uncertainty during the problem specification stage. Uncertainty is the term used in this guidance to describe *lack of knowledge* about models, parameters, constants, data, and beliefs. Defining the ranges of acceptable uncertainty helps project planners generate "specifications" for quality assurance planning and partially determines the appropriate boundary conditions and complexity for the model being developed.

Data Quality Objectives (DQOs)[1] (USEPA 2000a) enable the development of specifications for model quality and associated checks (Appendix B, Box B1: Background on EPA Quality System) during the problem specification stage. The DQOs provide guidance on how to state data needs when limiting decision errors (false positives or false negatives) relative to a given decision. False rejection decision errors (false positives) occur when the null-hypothesis (or baseline condition) is incorrectly rejected based on the sample data. The decision is made assuming the alternate condition or hypothesis to be true when in reality it is false. False acceptance decision errors (false negatives) occur when the null hypothesis (or baseline condition) cannot be rejected based on the available sample data. The decision is made assuming the baseline condition is true when in reality it is false. Well-defined DQOs are invaluable during later stages of model development (e.g., calibration and verification) and often attract the attention of peer reviewers. Included in these objectives should be a statement about the acceptable level of total uncertainty that will still enable model results to be used for the intended purpose (Appendix B, Box B2: Configuration Tests Specified in the QA program).

3.3 Model Identification and Selection

Once the problem specifications are understood, the project team will identify the type of model that is well-suited to the problem. Some of the choices among model types include: empirical vs. mechanistic, static vs. dynamic, simulation vs. optimization, deterministic vs. stochastic, and lumped vs. distributed models.

The scope (i.e., spatial, temporal and process detail) of models that can be used for a particular application can range from the simplest models to the very complex depending on the problem specification and data availability, among other factors. As different types of models are appropriate for solving different problems, a graded approach to model selection and development should be adopted. It is also important that a model's domain of applicability is well understood. A model's application niche is the set of conditions under which the use of a model is scientifically defensible.

For mechanistic modeling of common environmental problems, the project team may be able to use an existing "model framework" that is well-suited to the problem. A model framework is a formal mathematical specification of the concepts and procedures of the conceptual model, usually translated into computer software. There are many existing model frameworks in the public domain that can be used in environmental assessments. A number of institutions, including EPA, develop and maintain these model frameworks on an ongoing basis.

[1] The DQOs provide guidance on how to state data needs when limiting decision errors (false positives or false negatives) relative to a given decision. False rejection decision errors (false positives) occur when the null-hypothesis (or baseline condition) is incorrectly rejected based on the sample data. The decision is made assuming the alternate condition or hypothesis to be true when in reality it is false. False acceptance decision errors (false negatives) occur when the null hypothesis (or baseline condition) cannot be rejected based on the available sample data. The decision is made assuming the baseline condition is true when in reality it is false.

Box 3: Example of Model Selection Considerations: Arsenic in Drinking Water
(from Box 5-3 NRC Report on Models in Environmental Regulatory Decision Making)

A major challenge for regulatory model applications is which model to use to inform the decision making process. In this example, several models were available for estimating the incidence of cancer with different levels of arsenic in drinking water that differed according to how age and exposure were incorporated (Morales et al., 2000). All the models assumed that the number of cancers observed in a specific age group of a particular village followed a Poisson model with parameters, depending on the age and village exposure level. Linear, log, polynomial, and spline models for age and exposure were considered.

These various models differed substantially in their fitted values, especially in the critical low-dose area; which is so important for establishing the benchmark dose (BMD) used to set a reference dose (RfD). The fitted-dose response model was also strongly affected by whether Taiwanese population data were included as a baseline comparison group. The estimates of the BMD and associated lower limit (BMDL) varied by over an order of magnitude, depending on the particular modeling assumptions used.

Several strategies are available for choosing among multiple models. One strategy would be to pick the "best" model—for example, use one of the popular statistical goodness of fit, such as the Akieke (*sic*) information criterion (AIC) or the Bayesian information criterion (BIC). These approaches correspond to picking the model that maximizes log-likelihood, subject to a penalty function reflecting the number of model parameters, thus effectively forcing a trade-off between improving model fit by adding addition model parameters versus having a parsimonious description. In the case of the arsenic risk assessment, however, the noisiness of the data meant that many of the models explored by Morales et al. (2000) were relatively similar in terms of statistical goodness-of-fit criteria. In a follow-up paper, Morales et al. (2006) argued that it was important to address and account for the model uncertainty, because ignoring it will underestimate the true variability of the estimated model fit and, in turn, overestimate confidence in the resulting BMD and lead to "risky decisions" (Volinsky et al. 1997).

Morales et al. suggest the use of Bayesian model averaging (BMA) as a tool that avoids the need to pick one particular model by combining over a class of suitable models. In practice, estimates based on a BMA approach tend to approximate a weighted average of estimates based on individual models, the weights reflecting how well each individual model fits the observed data. More precisely, these weights can be interpreted as the probability that a particular model is the true model, given the observed data. The figures below show the results of applying a BMA procedure to the arsenic data. Figure (a) below plots individual fitted models, the width of each plotted line reflecting the weights. Figure (b) shows the estimated overall dose-response curve (solid line) fitted via BMA.

The shaded area shows the upper and lower limits (2.5% and 97.5% tiles) based on the BMA procedure. The dotted lines show upper and lower limits based on the best fitting models. Figure b (L30) effectively illustrates the inadequacy of standard statistical confidence intervals in characterizing uncertainty in settings where there is substantial model uncertainty. The BMA limits coincide closely with the individual curves at the upper level of the dose-response curve where all the individual models tend to give similar results.

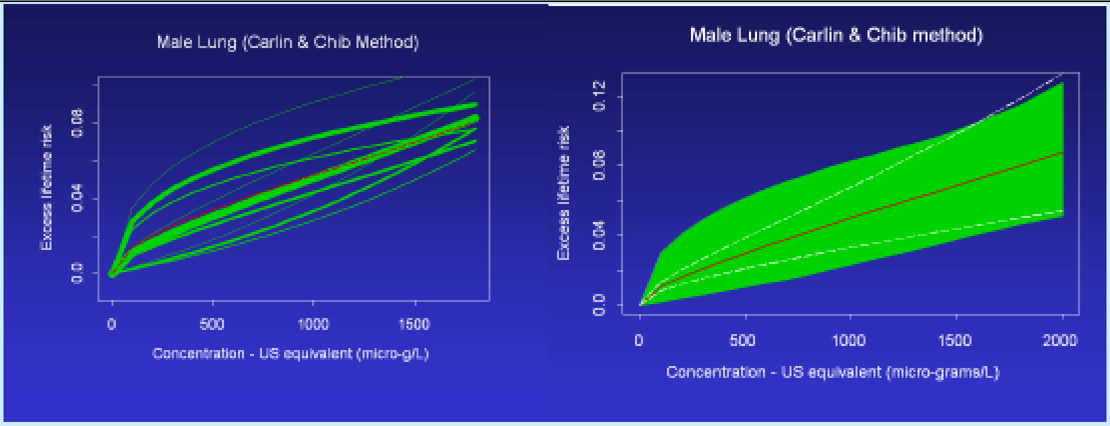


Figure 2 (a) Individual dose-response models, and (b) overall dose-response model fitted using the Bayesian model averaging approach. Source: Morales et al. 2000.

An existing model framework may or may not include all of the important elements of the conceptual model. Ideally, there are more than one model frameworks that will meet project needs, and the project team can select the best model for the specified problem. In some instances, there are no model frameworks appropriate to the task, and EPA will develop a new model framework for the project. Another potential option is to modify an existing framework to include additional capabilities necessary to address project needs.

Some assessments will require the linkage of multiple model frameworks. For example, air quality modeling often links meteorological, emissions, and air chemistry/transport models. They are linked in that the output from one model is used as input data to a second model. When employing linked models, the project team should evaluate each component model as well as the full system of integrated models at each stage of the model development and evaluation process.

In all cases, the model's documentation should provide a clear understanding of why and how the model will and can be used. The modeling literature should be reviewed before developing a new model to determine whether an existing model meets the identified needs and requirements.

3.3.1 Model Framework Evaluation

The model framework is a formal mathematical specification of the concepts and procedures of the conceptual model. The mathematical framework is usually translated into a form amenable to running on a computer (i.e., algorithm development and model coding). Several issues considered during model framework development include:

- Does sound science support the underlying hypothesis? (Sound science may be defined in part, but is not restricted to, peer reviewed theory and equations.)
- Is the complexity of the model appropriate for the problem at hand?
- Do the quality and quantity of data support the choice of model?
- Does the model structure reflect all the relevant inputs based on the conceptual model?
- Has the model code been developed? If so, has it been verified?

If multiple viable competing hypotheses about the system behavior exist, it may be useful to statistically compare the performance of these competing models with observational, field, or laboratory data (Chapter 4). The principles of scientific hypothesis testing (Platt 1964) should be applied to model development using an iterative approach to model evaluation (Hilborn and Mangel 1997).

3.3.2 Model Complexity

Due to the inherent trade-off between model framework uncertainty and data uncertainty, an optimal level of model complexity exists for every model (Fig. 3).

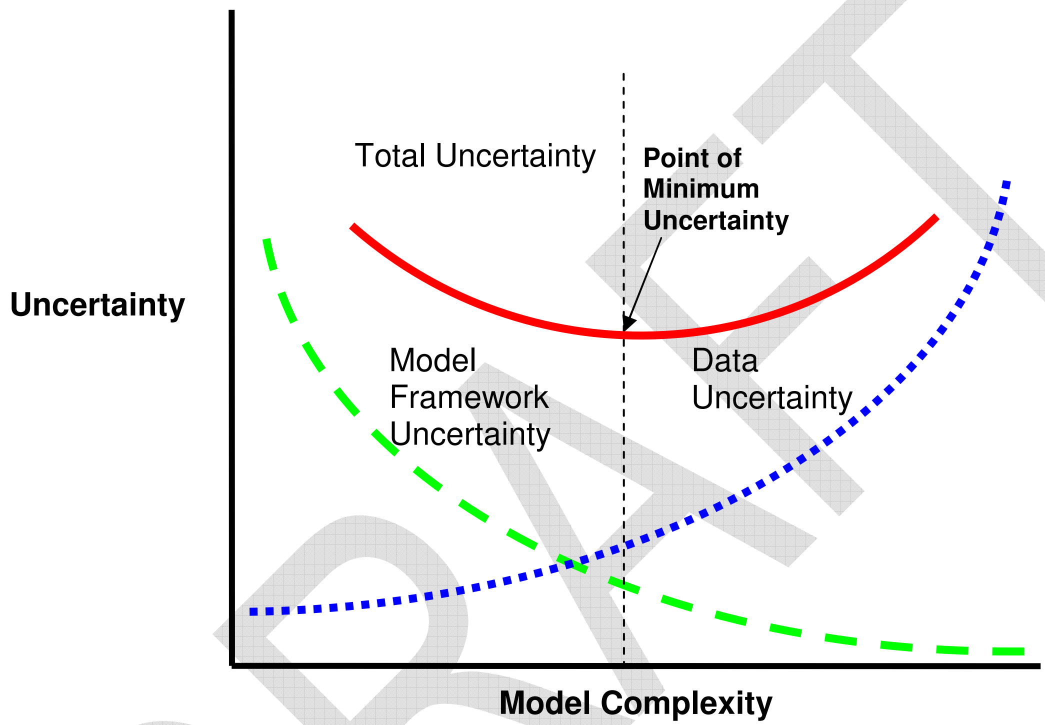


Figure 3 Illustration of relationship between model framework uncertainty and data uncertainty, and the combined effect on total model uncertainty. In this document, uncertainty is the term used to describe incomplete knowledge about specific factors, parameters (inputs) or models. Model framework uncertainty is a function of the soundness of the underlying scientific foundations of the model. Data uncertainty measurement errors, analytical imprecision and limited sample sizes during the collection and treatment of data that are used to characterize model parameters. An optimal level of complexity exists for every model, which is represented in the figure as the "point of minimum uncertainty." (Adapted from Hanna 1988).

For example, within the field of air-quality modeling it is sometimes necessary to compromise when choosing among the physical processes that will be treated explicitly in the model. If the objective is to estimate the pattern of concentration values in the near vicinity of one (or several) source(s), then typically chemistry is of little importance. For such situations, travel times from sources to receptors are often too short for chemical formation and destruction to greatly affect results. However, such situations demand that the air quality model properly treat near-field source emission effects such as: building wakes, initial characterization of source release conditions and size, rates of diffusion of pollutants released as they transport downwind, and land use effects on plume transport. If chemistry is to be treated explicitly, initial source release effects are typically unimportant because the pollutants are well-mixed over some volume of the atmosphere before the chemistry of interest has greatly affected the results. To date, attempts to treat both near-field dispersion effects and chemistry have been found to be inefficient and slow on desktop computers.

Because of these competing objectives, parsimony (economy or simplicity of assumptions) is a desirable modeling trait. As illustrated in Figure 3, the likelihood of degrading performance due to input uncertainty increases as a model's formulation increases in complexity (to explicitly treat more physical processes). This decrease in performance is a function of the increase in the number of input variables and thus an increase in data uncertainty. Because there may not be a "best" model for a given decision and different models contain different types and ranges of uncertainty, sensitivity analysis at early stages in the model development phase is useful for identifying the relative importance of model parameters. Sensitivity analysis is the computation of the effect of changes in input values or assumptions (including boundaries and model functional form) on the outputs (Morgan and Henrion 1990)

When sensitivity analyses (Chapter 4/Appendix C) show that a model parameter has an insignificant effect on outputs and there is no process-based rationale for inclusion in the model-framework, it may be appropriate to eliminate this parameter in order to constrain model complexity. However, because sensitivity surfaces are often extremely irregular, a variable of little significance during one application of a model may be more important in a different application. In past reviews of Agency models, the SAB has supported the general guiding principle of simplifying complex models, where possible, for the sake of transparency (SAB 1988) but has emphasized that care should be taken not to eliminate important parameters from process-based models simply because data are unavailable or difficult to obtain (SAB 1989). In any case, the quality and resolution of available data will ultimately constrain the type of model that can be applied. Hence, it is important to identify the existing data and and/or field collection efforts that are needed to adequately parameterize the model framework and support the application of a model. The NRC Committee on Models in Regulatory Environmental Decision making further recommended that models used in the regulatory process should be no more complicated that is necessary to inform regulatory decision. They further point out that it is often preferable to omit capabilities that do not improve model performance substantially (NRC 2007).

3.3.3 Model Coding and Verification

Model coding translates the mathematical equations that constitute the model framework into functioning computer code. Code verification ascertains that the computer code has no inherent numerical problems with obtaining a solution. Code verification tests whether the code performs according to its design specifications. It should include an examination of the numerical technique in the computer code for consistency with the conceptual model and governing equations (Beck et al 1994). Independent testing of the code once it is fully developed can be useful as an additional check of integrity and quality.

Steps taken early in the development of the computer code can help minimize programming errors later on and facilitate the code verification process. *Using "comment" lines to describe the purpose of each component within the code* during development makes future revisions and improvements by different modelers and programmers more efficient. *Using a flow chart* during the development of the conceptual model before beginning coding helps to show the overall structure of the model program. This provides a simplified description of the calculations that will be performed in each step of the model.

Breaking the program/model into component parts or modules is also useful for careful consideration of model behavior in an encapsulated way. This allows the modeler to test the behavior of each of the sub-components separately, expediting testing and increasing confidence in the program. A module is an independent piece of software that forms part of one or more larger programs. It is useful to break large models into discrete modules because testing and locating/correcting errors ("debugging") is more difficult with large programs. The approach also makes it easier to re-use relevant modules in future modeling projects, or to update/add/remove sections of the model without altering the overall structure of the program.

Time and resources can often be saved by *using generic algorithms for common tasks*, allowing efforts to be focused on developing and improving the original aspects of a new model. An algorithm is a precise rule (or set of rules) for solving some problem. Commonly used algorithms are often published as "recipes" with publicly available code (e.g., Press 1992). Existing Agency models and code should also be reviewed to minimize duplication in effort. The CREM models knowledge base, which will contain a web-accessible inventory of models, will provide a resource for model developers to support these efforts. Software engineering has evolved rapidly in the past ten years and continues to advance rapidly with changes in technology and user platforms. For example, some of the general recommendations for developing computer code given above do not apply to models that are developed using object-oriented platforms. Object-oriented platforms model systems use a collection of cooperating "objects." These objects are treated as instances of a class within a class hierarchy, where a class is a set of objects that share a common structure and behavior. The structure of a class is determined by the class variables, which represent the state of an object of that class and the behavior is given by the set of methods associated with the class (Booch 1994). When models are developed with object oriented platforms, the user should print out the actual mathematical relationships that the platform generates and this should be reviewed as part of the code validation process.

There are many available references on programming style and conventions that provide specific, technical suggestions for developing and testing computer code (e.g., *The Elements of Programming Style* (Kernigham and Plaugher 1988)). In addition to these recommendations, the *Guidance for Quality Assurance Project Plans for Modeling* USEPA 2002b suggests a number of practices during code verification to "check" how well it follows the "specifications" laid out during QA planning (Appendix B, Box B2: Configuration Tests Specified in the QA Program).

3.4 Development of Application Tools

Once a model framework is selected, the modeler populates the framework with the specific system characteristics that will comprise the model, including geographic boundaries of the model domain, boundary conditions, pollution source inputs, and model parameters. In this manner, the generic computational capabilities of the model framework are converted into an application tool to assess a specific problem occurring at a specific location. Model parameters are terms in the model that are fixed during a model run or simulation but can be changed in different runs either as a method for conducting sensitivity analysis or while performing an uncertainty analysis when probabilistic distributions are selected to model parameters or to achieve calibration (defined below) goals. Parameters can be quantities estimated from sample data that characterize statistical populations or constants such as the speed of light and gravitational force. Other activities at this stage of model development include creating a user guide for the model, assembling datasets for model input parameters, and determining hardware requirements.

3.4.1 Input Data

The accuracy and precision of input data used in the model is a major source of uncertainty. Accuracy refers to the closeness of a measured or computed value to its "true" value, where the "true" value is obtained with perfect information. Due to the natural heterogeneity and random variability (stochasticity) of many environmental systems, this "true" value exists as a distribution rather than a discrete value. Variability is the term used in this guidance to describe differences that are attributable to true heterogeneity or diversity in model parameters. Because of variability, the "true" value of model parameters is often a function of the degree of spatial and temporal aggregation. Precision refers to the quality of being reproducible in outcome or performance. With models and other forms of quantitative information, precision often refers to the number of decimal places to which a number is computed. This is a measure of the "preciseness" or "exactness" of the model.

The most appropriate data, as defined by QA protocols for field sampling, data collection, and analysis (USEPA 2002c, 2002d, 2000b) should always be selected for use in modeling analyses. Whenever possible, all parameters should be directly measured in the system of interest.

Box 4: Comprehensive Everglades Restoration Plan: An Example of the Interdependence of Models and Data from Measurements (*from NRC Report on Models in Environmental Regulatory Decision Making*)

The planned restoration of the Florida Everglades is the largest ecosystem restoration effort ever undertaken in terms of its geographical extent and number of individual components. The NRC Committee on Restoration of the Greater Everglades Ecosystem, which was charged with providing scientific advice on this effort, describes the role that modeling and measurements should play in implementing an adaptive approach to restoration (NRC 2003). Under the committee's vision, monitoring of hydrological and ecological performance measures should be integrated with mechanistic modeling and experimentation to better understand how the Everglades functions and how the system will respond to management practices and external stresses. Because the individual components of the restoration plan will be staggered in time, the early components can be used as experiments to provide scientific feedback to guide and refine implementation of later components of the plan.

The NRC panel on *Models in Environmental Regulatory Decision Making* further recommends that: "…using adapting strategies to coordinate data collection and modeling should be a priority for decision makers and those responsible for regulatory model development and application. The interdependence of measurements and modeling needs to be fully considered as early as the conceptual model development phase."

3.4.2 Model Calibration

Some models are "calibrated" to set parameters. Guidance on model calibration as a QA project plan element is given in Appendix B (Box B3: Quality Assurance Planning Suggestions for Model Calibration

Activities). In this guidance, calibration is defined as the process of adjusting model parameters within physically defensible ranges until the resulting predictions give the best possible fit to the observed data (USEPA 1994b). In some disciplines, calibration is also referred to as parameter estimation (Beck et al 1994).

Most process-oriented environmental models are underdetermined; that is, they contain more uncertain parameters than state variables that can be used to perform a calibration. Sensitivity analysis can be used to identify key processes influencing the state variables. In some instances, it is possible to directly measure the rate constant for a key process. An example of this practice would be to measure the rate of photosynthesis (process) in a lake in addition to the biomass of phytoplankton (state variable). Direct measurement of rate parameters can reduce model uncertainty.

In cases where a calibration database is developed and improved over time, it may be necessary to periodically reevaluate initial adjustments and estimates. When data for quantifying one or more parameter values are limited, calibration exercises can also be used to find solutions that result in the 'best fit' of the model. However, this type of calibration should be undertaken with caution, as these solutions will not provide meaningful information unless they are based on *measured* physically defensible ranges.

Because of these concerns, the use of calibration to improve model performance varies among EPA offices and regions. For a particular model, the appropriateness of calibration may be a function of the modeling activities undertaken. For example, it is standard practice for the Office of Water to calibrate well-established model frameworks such as CE-QUAL-W2 (a model for predicting temperature fluctuations in rivers), to a specific system (e.g., the Snake River). This calibration generates a site-specific tool (e.g., the "Snake River Temperature" model). In contrast, the Office of Air and Radiation (OAR) more commonly uses model frameworks and models that do not need site-specific adjustments. For example, certain types of air models (e.g., gaussian plume) are parameterized for a range of meteorological conditions, and thus, do not need to be "re-calibrated" for different geographic locations (assuming that the range of conditions is appropriate for the model). This, in combination with a desire to avoid artificial improvements in model performance by adjusting model inputs outside of the ranges supported by the empirical databases, prompted OAR to issue the following statement on model calibration in their *Guideline on Air Quality Models* (USEPA 2003b):

> *Calibration of models is not common practice and is subject to much error and misunderstanding. There have been attempts by some to compare model estimates and measurements on an event-by-event basis and then calibrate a model with results of that comparison. This approach is severely limited by uncertainties in both source and meteorological data and therefore it is difficult to precisely estimate the concentration at an exact location for a specific increment of time. Such uncertainties make calibration of models of questionable benefit. Therefore, model calibration is unacceptable.*

In general, however, models benefit from thoughtful adaptation to respond adequately to the specifics of each regulatory problem to which they are applied.

4. Model Evaluation

Summary of Recommendations for Model Evaluation

- Model evaluation provides information to determine when a model, despite its uncertainties, can be appropriately used to inform a decision.
- Model evaluation addresses the soundness of the science underlying a model, the quality and quantity of available data, the degree of correspondence with observed conditions and the appropriateness of a model for a given application.
- Recommended components of the evaluation process that help to determine when a model, despite its uncertainties, can be appropriately used to inform a decision that were reviewed in this section include: (a) credible, objective peer review; (b) QA project planning and data quality assessment; (c) qualitative and/or quantitative model corroboration; and (d) sensitivity and uncertainty analyses.
- Quality is an attribute of models that is meaningful only within the context of a specific model application. Deciding whether a model serves its intended purpose can be determined through in-depth discussions between the model developers and users responsible for its application for a particular problem.
- Information gathered during model evaluation allows the decision maker to be better positioned to formulate decisions and policies that take into account all relevant issues and concerns.

4.1 Introduction to Model Evaluation

> Models will always be constrained by computational limitations, assumptions and knowledge gaps. They can best be viewed as tools to help inform decisions rather than as machines to generate truth or make decisions. Scientific advances will never make it possible to build a perfect model that accounts for every aspect of reality or to prove that a given model is correct in all aspects for a particular regulatory application. These characteristics…suggest that model evaluation be viewed as an integral and ongoing part of the life cycle of a model, from problem formulation and model conceptualization to the development and application of a computational tool.
>
> *--NRC Panel on Models in Environmental Regulatory Decision Making (NRC 2007)*

The natural complexity of environmental systems means that it is difficult to develop complete mathematical descriptions of relevant processes, including all of the intrinsic mechanisms that govern their behavior. Thus, policy-makers often rely on models as tools to approximate reality when implementing decisions that affect environmental systems. The challenge facing model developers and users is determining when a model, despite its uncertainties, can be appropriately used to inform a decision.

Model evaluation provides a vehicle for dealing with this problem. In this guidance, model evaluation is defined as *the process used to generate information to determine whether a model and its analytical results are of a quality sufficient to serve as the basis for a decision.*

Different disciplines assign different meanings to the terms "model evaluation" and "model validation." For example, Suter (1993) found that among models used for risk assessments, misconception often arises in the form of the question: "Is the model valid?" and statements such as "no model should be used unless it has been validated.” The author further points out that “validated” in this context means: (a) proven to correspond exactly to reality, or (b) demonstrated through experimental tests to make consistently accurate predictions.

Because every model contains simplifications, predictions derived from the model can never be completely accurate and the model can never correspond exactly to reality. Additionally, “validated models” (e.g., those that have been shown to correspond to field data), do not necessarily generate accurate predictions of reality for multiple applications (Beck 2002a). Thus, some researchers assert that no model is ever truly “validated,” but can only be invalidated for a specific application (Oreskes et al 1994). Accordingly, this guidance focuses on the process and techniques that can be used for *model evaluation* rather than model validation or invalidation.

Verification is another commonly used term applied to the evaluation process. However, in this guidance and elsewhere, model verification typically refers to model code verification as defined in the model development section. For example, the NRC Committee on Models in Regulatory Environmental Decision Making provides the following definition:

> *Verification* refers to activities that are designed to confirm that the mathematical framework embodied in the module is correct and that the computer code for a module is operating according to its intended design so that the results obtained compare favorably with those obtained using known analytical solutions or numerical solutions from simulators based on similar or identical mathematical frameworks (NRC 2007).

In simple terms, model evaluation provides information to assess the following factors (Beck 2002b):
1. How have the principles of sound science been addressed during model development?
2. How is the choice of model supported by the quantity and quality of available data?
3. How closely does the model approximate the real system of interest?
4. How does the model perform the specified task while meeting the objectives set by QA project planning?

These four factors address two components of model quality. The first factor focuses on the intrinsic mechanisms and generic properties of a model, regardless of the particular task to which it is applied. In contrast, the latter three factors are evaluated in the context of the use of a model within a *specific set of conditions*. Hence, it follows that model quality is an attribute that is meaningful only within the context of a *specific model application*. A model's quality to support a decision becomes known when information is available to assess these factors.

The NRC Panel recommends that evaluation of a regulatory model should continue throughout the life of a model and that an evaluation plan could:
- Describe the model and its intended uses.
- Describe the relationship of the model to data, including the data for both inputs and corroboration.
- Describe how such data and other sources of information will be used to assess the ability of the model to meet its intended task.
- Describe all the elements of the evaluation plan by using an outline or diagram showing how the elements relate to the model’s life cycle.

- Describe the factors or events that might trigger the need for major model revisions or the circumstances that might prompt users to seek an alternative model. These could be fairly broad and qualitative.
- Identify responsibilities, accountabilities, and resources needed to ensure implementation of the evaluation plan.

As stated above, the goal of model evaluation is to ensure model quality. At EPA, quality is a concept defined by the Information Quality Guidelines (IQGs) (USEPA 2002a). The IQG applies to all information that is disseminated by EPA, including models, information from models and input data (Appendix B, Box B4: Definition of Quality). According to the IQG, quality has three major components: integrity, utility, and objectivity. Ensuring the objectivity of information from models by considering their accuracy, bias and reliability is emphasized in this guidance as part of the model evaluation process that addresses the questions listed above. While accuracy was defined in section 2.4, for the purposes of this guidance bias and reliability are defined as follows:

Bias describes any *systematic deviation* between a measured (i.e., observed) or computed value and its "true" value. Bias is affected by faulty instrument calibration and other measurement errors, systematic errors during data collection, and sampling errors such as incomplete spatial randomization during the design of sampling programs.

Reliability is the confidence that (potential) users have in a model and in the outputs of the model such that they are willing to use the model and accept its results (Sargent 2000). Specifically, reliability is a function of the performance record of a model and its conformance to best available, practicable science.

4.2 Best Practices for Model Evaluation

This guidance provides an overview of model evaluation that can help answer the questions posed in Chapter 4.1 to determine when a model, despite its uncertainties, can be appropriately used to inform a decision. In summary, these questions are intended to address the soundness of the science underlying a model, the quality and quantity of available data, the degree of correspondence with observed conditions and the appropriateness of a model for a given application.

Box 5: Examples of Life Cycle Model Evaluation
(from Box 4-5 in NRC Report on Models in Regulatory Environmental Decision Making)

An example that shows the value in evaluating a model from conceptual through use stages is the work of the Organization for Economic Cooperation and Development (OECD) to develop a screening model for assessing the persistence and long-range transport potential of chemicals. The goal of this effort was a consensus model that was evaluated against a broad set of available models and data. The evaluation process began at a workshop in 2001 where the model performance and evaluation goals were set before model selection and development began (OECD 2002). To act upon the recommendations, an OECD expert group was established in 2002. This group published a guidance document on the use of multimedia models for estimating environmental persistence and long-range transport. From 2003 to 2004, the expert group performed an extensive comparison of nine available multimedia fate and transport models to compare and assess their performance (Fenner et al. 2005; Klansmeier et al. 2006). Following this effort, the expert group developed a parsimonious consensus model representing the minimum set of key model components identified in the model comparison. The expert group then convened three international workshops to disseminate this consensus model and provide an on-going model evaluation forum (Scheringer et al. 2006). In this example, significant effort was invested (more than half of the total effort in the OECD case) in the conceptual and model formulation stages. Moreover, much of this effort focused on performance evaluation. The committee recognizes that each model's life cycle is different but notes that attention should be given to developing consensus-based approaches in the model concept and formulation stages. Conducting concurrent evaluations at these stages in this setting resulted in a high degree of buy-in from the various modeling groups.

The proposed "tools" or best practices emphasized in this guidance are: peer review of models, QA project planning including data quality assessment, model corroboration and sensitivity and uncertainty analysis. In this guidance, corroboration is defined as a qualitative and/or quantitative evaluation of the accuracy and predictive capabilities of a model.

As discussed in previous sections, the process of model evaluation is iterative in nature. Hence, the proposed qualitative and quantitative assessment techniques discussed below may be effectively applied throughout model development, testing and application and should not be interpreted as sequential steps for model evaluation.

The distinction between qualitative and quantitative assessments is given below:

Qualitative Assessments: Some of the uncertainty in model predictions may arise from sources whose uncertainty cannot be quantified. Examples are uncertainties about the theory underlying the model, the manner in which that theory is mathematically expressed to represent the environmental components, and theory being modeled. The subjective evaluations of experts may be needed to determine appropriate values for model parameters and inputs that cannot be directly observed or measured (e.g., air emissions estimates). Qualitative assessments are needed for these sources of uncertainty. Qualitative, corroboration activities may involve expert elicitation regarding the system's behavior and comparison with model forecasts.

Quantitative Assessments*:* The uncertainty in some sources—such as some model parameters and some input data—can be estimated through quantitative assessments involving statistical uncertainty and sensitivity analyses. In addition, comparisons can be made for the special purpose of quantitatively describing the differences to be expected between model estimates of current conditions and comparable field observations. However, model predictions are not what are directly observed, so special care needs to be taken in any exercise that attempts to make quantitative comparisons of model predictions with field data.

Model evaluation should always be conducted using a graded approach that is adequate and appropriate to the decision at hand (USEPA 2001, 2002b). This approach recognizes that model evaluation can be modified to the circumstances of the problem at hand and that programmatic requirements are varied. For example, a "screening" model used for risk management should undergo "rigorous" evaluation to avoid false negatives, while still not imposing unreasonable data-generation burdens (false positives) on the regulated community. A screening model is a type of model designed to provide a "conservative" or risk-averse answer. The appropriate degree of model evaluation is ideally identified by both decision-makers and modeling staff at the onset of new projects (§3.1).

External circumstances can affect the rigor required in model evaluation. For example, in cases where the likely result of the modeling will be costly control strategies and associated controversy, more detailed model evaluation may be necessary. In these cases, many aspects of the modeling may come under close scrutiny, and it is incumbent upon the modeler to answer questions that arise about the model and document the findings of the model evaluation process. A deeper level of model evaluation may also be appropriate when modeling unique or extreme situations not previously encountered.

Finally, as noted earlier, some assessments require the use of multiple, linked models. This linkage has implications for assessing uncertainty and applying the system of models. Each component model as well as the full system of integrated models needs to be evaluated.

The substance of the following discussion on peer review of models and quality assurance protocols for input data is drawn from existing guidance. In addition, this chapter (along with Appendix C) provides new guidance on model corroboration activities and the use of sensitivity and uncertainty analysis during model evaluation.

4.2.1 Scientific Peer-Review

Peer review provides the main mechanism for independent evaluation and review of environmental models used by the Agency. Peer review addresses questions (1) and (4) posed during model evaluation (§4.1). Thus, the purpose of peer review is twofold. First, peer review evaluates whether the assumptions, methods, and conclusions derived from environmental models are based on sound scientific principles. Secondly, peer review provides a useful check on the scientific appropriateness of a model for informing a specific regulatory decision. This is particularly important for secondary applications of existing models. As part of this second objective, peer reviews may focus on whether a model meets the objectives or specifications set as part of the quality assurance plan (see USEPA 2002b) at the onset of the modeling project (§3.1). Peer review is *not* a mechanism to comment on the *regulatory decisions* or policies that are informed by models (USEPA 2000c). Peer review charter questions and corresponding records for peer reviewers to answer those questions need to be incorporated into assessment planning as part of developing the QA Project Plan. All models that inform *significant*[2] regulatory decisions are candidates for peer review (USEPA 2000c, USEPA 1993). There are a number of reasons for initiating this review:

- Use of model results as a basis for major regulatory or policy/guidance decision-making
- Significant investment of Agency resources
- Inter-Agency or cross-Agency implications/applicability

[2] Executive Order 12866 (58 FR 51735) requires federal agencies to determine whether a regulatory action is "significant" and therefore, its *underlying scientific basis* is subject to review by the Office of Management and Budget (OMB) and the requirements of the Executive Order. The rule or regulation itself is *not* subject to peer review. The Order defines "significant regulatory action" as one that is likely to result in a rule that may: (1) Have an annual effect on the economy of $100 million or more or adversely affect in a material way the economy, a sector of the economy, productivity, competition, jobs, the environment, public health or safety, or State, local, or tribal governments or communities; (2) Create a serious inconsistency or otherwise interfere with an action taken or planned by another agency; (3) Materially alter the budgetary impacts of entitlements, grants, user fees, or loan programs or the rights and obligations of recipients thereof; or (4) Raise novel legal or policy issues arising out of legal mandates, the President's priorities, or the principles set forth in the Order .

Existing guidance recommends that the first application of a new model should undergo scientific peer review but for subsequent applications, the program manager should consider the scientific/technical complexity and/or the novelty of the particular circumstances (USEPA 1993). In the interests of conserving resources, peer-review of "similar" applications should be avoided. When a modeling project uses a well-established model framework (e.g., WASP), it is left to the discretion of project managers within the individual program offices and regions to determine when a modeling product is sufficiently similar to past applications and/or does not affect "significant" actions or policies, such that it should not be subject to peer review.

Models used for secondary applications (existing EPA models or proprietary models) will generally undergo a different type of evaluation than those developed with a specific regulatory information need in mind. By their nature, reviews of secondary applications models may deal more with uncertainty about the appropriate application of a model to a specific set of conditions than with the science underlying the model framework. Information from peer reviews is also helpful for choosing among multiple competing models for a specific regulatory application. Finally, peer review is a useful mechanism for identifying the limitations of existing models

Aspects of a model that should be reviewed in this process to establish scientific credibility are: (a) appropriateness of input data, (b) appropriateness of boundary condition specification, (c) documentation of inputs and assumptions, (d) the applicability and appropriateness of using default values, (e) documentation and justification for adjusting model inputs to improve model performance (calibration), and (f) model application with respect to the range of its validity, (g) supporting empirical data that strengthen or contradict the "conclusions" that are based on model results (SAB 1993a, USEPA 1993).

To be most effective and maximize its value, external peer review should begin as early in the model *development* phase as possible (USEPA 2000b). Because peer review involves significant time and resources, these allocations need to be components of the project planning and any related contracts. In the early stages, peer review can help to review the conceptual basis of models and potentially save time by redirecting misguided initiatives, identifying alternative approaches, or providing strong technical support for a potentially controversial position (SAB 1993a, USEPA 1993). Peer review at the development stage is also useful as an independent external review of model code (i.e., model verification). External peer review of the *applicability* of a model to a particular set of conditions should be considered well in advance of any decision-making as it helps to avoid inappropriate applications of a model for specific regulatory purposes (USEPA 1993).

The logistics of the peer review process are left to the discretion of the office managers responsible for use and application of the model for a given decision. Mechanisms for accomplishing external peer review include, but are not limited to, the following**:**

- Using an ad hoc panel of at least three scientists[3]
- Using an established external peer review mechanism such as the Science Advisory Board or the Science Advisory Panel
- Holding a technical workshop[4]

---

[3] The selection of an ad hoc panel of peer reviewers may create legal concerns under the Federal Advisory Committee Act (FACA). Compliance with this statute's requirements is best summarized in the Chapter two of the *Peer Review Handbook,* "Planning a Peer Review" (USEPA 2000c). Guidance may also be sought from the Cross-Cutting Issues Law Office of the Office of General Council.

Specific techniques for selecting the qualifications and number of reviewers needed for a given modeling project can be found in the guidelines for peer review (SAB 1993a, USEPA 2000c, USEPA 1993, USEPA 1994a) and are summarized in Appendix C of this guidance.

4.2.2 Quality Assurance Project Planning and Data Quality Assessment

Another aspect of model evaluation that addresses the issue of whether a model has been developed according to the principles of sound science is data quality. Some variability in data is unavoidable (see § 4.2.3.1), but adhering to the tenets of data quality assessment described in other Agency guidance[5] (Appendix C, Box C2: Quality Assurance Planning and Data Acceptance Criteria) helps to minimize data uncertainty.

Well-executed QA project planning also helps to ensure that a model performs the specified task, which was the fourth question related to model evaluation posed in Section 4.1. As discussed above, evaluating the degree to which a modeling project has met QA objectives is often a function of the external peer review process. The *Guidance for Quality Assurance Project Plans for Modeling* (USEPA 2002b) provides general information about how to document quality assurance planning for modeling (e.g., specifications or assessment criteria development, assessments of various stages of the modeling process, reports to management as feedback for corrective action, and finally the process for acceptance, rejection or qualification of the output for use) to conform with EPA policy and acquisition regulations. Data quality assessments are a key component of the QA plan for models.

The quality and quantity (representativeness) of supporting data used to parameterize and (when available) corroborate models should be evaluated during all relevant stages of a modeling project. Such assessments are needed to evaluate whether the available data are sufficient to support the choice of the model to be applied (question two, §4.1). In addition, model outputs cannot be meaningfully compared to observational data (question three, §4.1) unless these data are representative of the true system being modeled.

---

[4] Use of a 'one-shot' technical workshop does not implicate the same concerns under FACA, especially if EPA personnel see only the individual opinions of the workshop attendees. However, repeated meetings of the workshop group or an attempt to forge a group consensus at the end of a one-day meeting might implicate FACA requirements.

[5] Other guidance that can help to ensure the quality of data used in modeling projects includes:

- *Guidance for the Data Quality Objectives Process*, a systematic planning process for environmental data collection (USEPA 2000a).
- *Guidance on Choosing a Sampling Design for Environmental Data Collection*, on applying statistical sampling designs to environmental applications (USEPA 2002c).
- *Guidance for Data Quality Assessment: Practical Methods for Data Analysis*, to evaluate the extent to which data can be used for a specific purpose (USEPA 2000b).

4.2.3 Corroboration, Sensitivity Analysis, and Uncertainty Analysis

The question, "How closely does the model approximate the real system of interest?" is unlikely to have a simple answer. In most cases, it will not simply be a matter of comparing model results and empirical data. As noted in Section 3.1, in developing and using an environmental model it is important that modelers and decision-makers consider the acceptable degree of uncertainty within the context of a specific model application. An understanding of the uncertainties underlying a model is needed to meaningfully address this question. Where practical, the recommended analyses should be conducted and their results reported in the documentation supporting the model.

4.2.3.1 Types of Uncertainty

Uncertainties in the scientific sense are a component of all aspects of the modeling process. However, identifying the types of uncertainty that significantly influence model outcomes (qualitatively or quantitatively) and communicating their importance is key to successfully integrating information from models into the decision-making process. As defined in Chapter 3, uncertainty is the term used in this guidance to describe incomplete knowledge about specific factors, parameters (inputs) or models. For organizational simplicity, uncertainties that affect model quality are categorized in this guidance as:

a) **Model framework uncertainty**, resulting from incomplete knowledge about the factors that control the behavior of the system being modeled, limitations in spatial or temporal resolution, and simplifications of the system;

b) **Model input uncertainty**, resulting from data measurement errors, inconsistencies between measured values and those used by the model (e.g., in their level of aggregation/averaging), and parameter value uncertainty;

c) **Model niche uncertainty**, resulting from the use of a model outside of the system for which it was originally developed and/or developing a larger model from several existing models with different spatial or temporal scales.

Box 6: Example of Model Input Uncertainty

The NRC Panel on *Models in Environmental Regulatory Decision Making* provide a detailed example of the effect of model input uncertainty on policy decisions in their example "Ozone Modeling and the Irregular Swings Between Science and Policy." This example is summarized here.

The formation of ozone in the lower atmosphere (troposphere) is an exceedingly complex chemical process involving the interaction of oxides of nitrogen (NO_x), volatile organic compounds (VOCs), sunlight, and dynamic atmospheric processes. The basic chemistry of ozone formation was known in the early 1960s (Leighton 1961). Reduction of ozone concentrations in general requires control of either NO_x or VOC emissions of a combination of both. Due to the nonlinearity of atmospheric chemistry, the selection of the emission-control strategy has traditionally relied on air quality models.

One of the first attempts to include the complexity of atmospheric ozone chemistry in the decision-making process was a simple observations-based model, the so-called Appendix J curve (36 Fed. Reg. 8166 [1971]). The curve was used to indicate the percentage VOC emission reduction required to attain the ozone standard in an urban area based on peak concentration of photochemical oxidants observed in that area. Reliable NO_x data were virtually nonexistent at the time and Appendix J was based on data from measurements of ozone and VOC concentrations from 6 U.S. cities. The Appendix J curve was based on the hypothesis that reductions of VOC emissions were the most effective emission-control path, and this conceptual model helped define legislative mandates enacted by Congress that emphasized controlling these emissions.

The choice in the 1970s to concentrate on VOC controls was supported by early results from models. While new results regarding the higher than expected biogenic VOC emissions were being gathered in the 1980s, EPA continued on its path of emphasizing VOC controls, in part because the schedule set by Congress and EPA for attainment of ozone ambient air quality standards was not conducive to reflection on the basic elements of the science (Dennis, 2002).

Reductions in VOCs from the early 1970s to the early 1990s had little effect on ozone concentrations. Regional ozone models developed in the 1980s and 1990s suggested that control of NO_x emissions was necessary in addition to, or instead of, the control of VOCs (NRC, 1991, "Rethinking the Ozone Problem"). The shift in the 1990s toward regulatory activities focusing on NO_x controls was partly due to the realization that historical estimates of emissions and the effectiveness of various control strategies in reducing emissions were not accurate. Thus, part of the reason ozone concentrations have not been reduced as much as hoped over the past 3 decades has been because emissions of some pollutants were much higher than originally estimated and have not been reduced as much as originally predicted. Regulations go forward despite imperfect models and information. The potential harm from environmental hazards can cause regulatory activities to proceed before the science and models are perfected. In the case of ozone modeling, the inputs to the models (emissions inventories in this case) are often more important than the model science (description of atmospheric transport and chemistry in this case) and require as careful an evaluation as the evaluation of the model. These factors point to the potential synergistic role that measurements play in model development and application.

In reality, all three categories are interrelated. Uncertainty in the underlying model structure or model framework uncertainty is the result of incomplete scientific data or lack of knowledge about the factors that control the behavior of the system being modeled. Model framework uncertainty can also be the result of simplifications necessary to translate the conceptual model into mathematical terms as described in Section 3.3. In the scientific literature this type of uncertainty is also referred to as structural error (beck 1987), conceptual errors (Konikow 1992), uncertainties in the conceptual model (Usunoff et al 1992), or model error/uncertainty (USEPA 1997, Luis and McLaughlin 1992). Structural error relates to the mathematical construction of the algorithms that make up a model while the conceptual model refers to the science underlying a model's governing equations. Model error and model uncertainty are both generally synonymous with model framework uncertainty.

Many models are developed iteratively to update their underlying science and resolve existing model framework uncertainty as new information becomes available. Models with long lives may undergo important changes from version to version. The MOBILE model for estimating atmospheric vehicle emissions, the CMAQ (Community Multi-scale Air Quality) model, and the QUAL2 water quality models are examples of models that have had multiple versions and major scientific modifications and extensions in over two decades of their existence (Scheffe and Morris, 1993; Barnwell et al., 2004; EPA 1999c, as cited in NRC 2007).

When an appropriate model framework has been developed, the model itself may still be highly uncertain if the input data or database used to construct the application tool is not of sufficient quality. The quality of empirical data used for both model parameterization and corroboration tests is affected by both uncertainty and variability. This guidance uses the term data uncertainty to refer to the uncertainty that is caused by measurement errors, analytical imprecision and limited sample sizes during the collection and treatment of data.

In contrast to data uncertainty, variability results from the inherent randomness of certain parameters that is attributable to the heterogeneity and diversity in environmental processes. Variability includes: fluctuations in ecological conditions, differences in habitat, and genetic variances among populations (USEPA 1997). Variability in model parameters is largely dependent on the extent to which input data have been aggregated (both spatially and temporally). Data uncertainty is sometimes referred to as reducible uncertainty because it can be minimized with further study (USEPA 1997). Accordingly, variability is referred to as irreducible because it can be better characterized and represented but cannot be reduced with further study (USEPA 1997).

A model's application niche is the set of conditions under which the use of a model is scientifically defensible (USEPA 1994b). Application niche uncertainty is therefore a function of the appropriateness of a model for use under a specific set of conditions. Application niche uncertainty is particularly important when choosing among existing models for an application outside of the system for which it was originally developed and/or developing a larger model from several existing models with different spatial or temporal scales (Levins 1992).

A good example of application niche uncertainty is given in the SAB review of MMSOILS (Multimedia Contaminant Fate, Transport and Exposure Model) where they address the adequacy of using a screening-level model to characterize situations where there is substantial subsurface heterogeneity or where non-aqueous phase contaminants are present (conditions differ from default values) (SAB 1993b). In this example, the SAB considered the MMSOILS model acceptable within its original application niche, but unsuitable for more heterogeneous conditions.

4.2.3.2 Model Corroboration

> The interdependence of models and measurements is complex and iterative for several reasons. Measurements help to provide the conceptual basis of a model and inform model development, including parameter estimation. Measurements are also a critical tool for corroborating model results. Once developed, models can derive priorities for measurements that ultimately get used in modifying existing models or in developing new ones. Measurement and model activities are often conducted in isolation…Although environmental data systems serve a range of purposes, and basic research performance, the importance of models in the regulatory process requires measurements and models to be better integrated. Adaptive strategies that relay on iterations of measurements and modeling, such as those discussed in the 2003 NRC report titled *Adaptive Monitoring and Assessment for the Comprehensive Everglades Restoration Plan*, provides examples of how improved coordination might be achieved (NRC 2007).
>
> --*NRC Panel on Models in Environmental Regulatory Decision Making*

Model corroboration includes all quantitative and qualitative methods for evaluating the degree to which a model corresponds to reality. The rigor of these methods will vary depending on the type and purpose of the model application. Quantitative model corroboration uses statistics to estimate how closely the results of a model match measurements made in the real system. Qualitative corroboration activities may include expert elicitation to obtain beliefs about a system's behavior in a data-poor situation. These corroboration activities may potentially move model forecasts toward consensus.

For newly developed model frameworks or untested mathematical processes, formal corroboration procedures may be appropriate. Formal corroboration may involve formulation of hypothesis tests for model acceptance, tests on data sets independent of the calibration data set, and quantitative testing criteria. In many cases, collecting independent data sets for formal model corroboration is extremely costly or otherwise unfeasible. In such circumstances, model evaluation may be appropriately conducted using a combination of other tools for model evaluation that are discussed in this section.

The degree of similarity between calibration data and corroboration data provides a measure of robustness of model performance (Reckhow 1994, Borsuk et al 2002). Robustness is defined in this guidance as the capacity of a model to perform equally well across the full range of environmental conditions for which it was designed. The degree of similarity among data sets available for calibration and corroboration provides insight into the robustness of the model. For example, if the dataset used to calibrate a model is identical or statistically similar to the dataset used to corroborate a model, an independent measure of the model's performance has not been provided. In this case, the exercise has provided no insight into model robustness. Conversely, when model outputs are similar to corroboration data that are significantly different from the calibration data, the corroboration exercise provides a measure of both model performance and robustness.

Quantitative model corroboration methods are also recommended for choosing among multiple models that are available for the same application. In such cases, models may be ranked on the basis of their statistical performance in comparison to the observational data (e.g. USEPA 1992). The Office of Air and Radiation evaluates models in this manner and when a single model is found to perform better than others in a given category, it is recommended in the *Guidelines on Air Quality Models* for application in that category as a preferred model (USEPA 2003a). If no model is found to perform better through the evaluation exercise, then the preferred model is selected on the basis of other factors such as past use, public familiarity, cost or resource requirements, and availability.

Box 7: Example: Comparing Results from Models of Varying Complexity Modeling Mercury in Fish
(*From Box 5-4 in NRC Report on Models in Environmental Regulatory Decision Making*)

A potential benefit of the clean-air mercury rule, which requires reductions in mercury emissions from coal-fired power plants, is the reduction of human exposure and related health impacts from methylmercury by reducing concentrations of this toxin in fish. There are many challenges and uncertainties in understanding the impact of reductions in atmospheric mercury emissions on human health. In its assessment of the benefits and costs of this rule, EPA used multiple models to look at one particular issue—how changes in atmospheric deposition would affect mercury concentrations in fish—and applied the models to assess some of the uncertainties in this impact (EPA 2005).

EPA based its national-scale benefits assessment on results from the mercury maps (MMaps) model. This model assumes that there is a linear, steady-state relationship between atmospheric deposition of mercury and mercury concentrations in fish and thus assumes that a 50% reduction in mercury deposition rates results in a 50% decrease in fish mercury concentrations. In addition, MMaps assumes instantaneous adjustment of aquatic systems and their ecosystems to changes in deposition. Thus, there is no time lag in the conversion of mercury to methylmercury and its bioaccumulation in fish. MMaps also does not deal with sources of mercury other than those from atmospheric deposition. Despite those limitations, the agency concluded that no other available model was capable of performing a national-scale assessment.

To further investigate fish mercury concentrations and assess the effects of MMaps assumptions, EPA applied more detailed models, including the spreadsheet-based ecological risk assessment for the fate of mercury (SERAFM) model, to five well-characterized ecosystems. As opposed to the steady-state MMaps model, SERAFM is a dynamic model that calculates the temporal response of mercury concentrations in fish tissues to changes in mercury loading. It includes multiple land-use types for representing watershed loadings of mercury through soil erosion and runoff. SERAFM partitions mercury among multiple compartments and phases, including aqueous phase, abiotic participles (for example, silts), and biotic particles (for example, phytoplankton). Comparisons of SERAFM's predictions with observed fish mercury concentrations for a single fish species in four ecosystems showed that the model underpredicted mean concentrations for one water body, overpredicted mean concentrations for a second water body, and accurately predicted mean concentrations for the other two. The error bars for the observed fish mercury concentrations in these four ecosystems were large, making it difficult to assess the accuracy of the models. Modeling of the four ecosystems also showed how assumed physical and chemical characteristics of the specific ecosystem affected absolute fish mercury concentrations and the length of time before fish mercury concentrations reached steady state.

Although EPA concluded that the best available science supports the assumption of a linear relationship between atmospheric deposition and fish mercury concentrations for broad-scale use, the more detailed ecosystem modeling demonstrated that individual ecosystems were highly sensitive to uncertainties in model parameters. The agency also noted that there were many model uncertainties that could not be quantified. Finally, although the case studies cover the bulk of the key environmental characteristics, extrapolating the individual ecosystem case studies to account for the variability in ecosystems across the country indicated that they might not represent extreme conditions that could affect how atmospheric deposition of mercury would affect fish mercury concentrations in a water body.

This example illustrates the usefulness of investigating a variety of models at varying levels of complexity. A hierarchical modeling approach, such as that used in the mercury analysis, can provide justification for simplified model assumptions or can potentially provide evidence for a consistent bias that would negate the assumption that a simple model is appropriate for broad-scale application.

4.2.3.3 Sensitivity and Uncertainty Analysis

Sensitivity analysis is the study of how the response of a model can be apportioned to changes in a model's inputs (Saltelli et al 2000a). A model's sensitivity describes the degree to which the model result is affected by changes in a selected input parameter (Beck et al 1994). Sensitivity analysis is recommended as the principal evaluation tool for characterizing the most and least important sources of uncertainty in environmental models.

The NRC Panel points out that there are two critical but distinct issues in uncertainty analysis for regulatory environmental modeling: what kinds of analyses should be done to quantify uncertainty, and how these uncertainties should be communicated to policy makers. To communicate model uncertainty, the NRC Panel recommends that hybrid approaches where unknown quantities are treated probabilistically and others are explored in scenario-assessment mode by decision makers through a range of plausible values. The Panel further acknowledges that:

> *"Effective uncertainty communication requires a high level of interaction with the relevant decision makers to ensure that they have the necessary information about the nature and sources of uncertainty and their consequences. Thus, performing uncertainty for environmental regulatory activities requires extensive discussion between analysts and decision makers." (NRC 2007)*

Uncertainty analysis investigates the lack of knowledge about a certain population or the real value of model parameters. Uncertainty is sometimes reducible through further study and with the collection of additional data. Existing Agency guidance (e.g., USEPA 1997) distinguishes uncertainty analysis from methods that are used to account for variability in input data and model parameters. Variability in model parameters and input data can be better characterized through further study but is usually not reducible (USEPA 1997).

Although uncertainty and sensitivity analysis are closely related, uncertainty is parameter specific, and sensitivity is algorithm-specific with respect to model "variables." By investigating the "relative sensitivity" of model parameters, a user can become knowledgeable of the relative importance of parameters in the model. By knowing the "uncertainty" associated with parameter values and the "sensitivity" of the model to specific parameters, a user will be more informed regarding the confidence that can be placed in model results. Recommended techniques for conducting uncertainty and sensitivity analysis are discussed in Appendix C.

4.3 Evaluating Proprietary Models

This guidance defines proprietary models as those computer models for which the source code is not universally shared. To promote the transparency with which decisions are made, the Agency has a preference for using non-proprietary models when available. The Agency acknowledges there will be times when the use of proprietary models provides the most reliable and best accepted modeling alternatives.

When proprietary models are used, their use should be accompanied by comprehensive, publicly available documentation that the proprietary models adhere to the recommendations within this guidance. That is, the propriety models should be accompanied by documentation that describes:

- The conceptual model and theoretical basis (as described in §3.3.1) for the model;
- The techniques and procedures used to verify that the proprietary model is free from numerical problems or "bugs" and that it truly represents the conceptual model (as described in §3.3.3);

- The process used to evaluate the model (as described in §4.2) and the basis for concluding that the model and its analytical results are of a quality sufficient to serve as the basis for a decision (as described in §4.1); and
- To the extent practicable, access to input and output data such that third parties can replicate results derived from the use of the proprietary model.

4.4 Learning from Prior Experiences – Retrospective Analyses of Models

The NRC Committee on Models in Environmental Regulatory Decision Making emphasized that the final issue in managing the model evaluation process is the learning that is developed through the examination of prior modeling experiences. Retrospective analysis of models is important for developing improvements to individual models and regulatory policies as well as systematically enhancing the overall modeling field. One objective of such an analysis might be to investigate systematic strengths and weaknesses that are characteristic of various types of models. A second objective might be the study of processes (for example, approaches to model development and evaluation) that led to successful model applications.

Accordingly, the NRC committee pointed out that retrospective analyses could be considered from two perspectives. First, retrospective analyses could deal with broad classes of models – for example, models of groundwater flow, surface water, air pollution, and health risks assessment. The goal of such an approach would be to investigate whether there are systematic weaknesses that are characteristic of various types of models. For example, a researcher estimated that in 20-30% of groundwater modeling efforts, surprising occurrences indicated that the conceptual model underlying the computer model were invalid (Bredehoeft (2003, 2005) in NRC 2007).

The second perspective on retrospective model analysis concerns specific models that have been in use for a number of years. For such cases, the aim of a retrospective analysis would be to address the question, "how well does the model work?". The best chance for identifying and correcting conceptual errors is through an ongoing evaluation of the model against data, especially data taken under novel conditions. A key benefit of retrospective evaluations of models of individual models and of model classes is the identification of priorities for improving models.

Box 8: Example of a Retrospective Model Analysis at EPA
(*from Box 4-6 NRC Committee on Models in Environmental Regulatory Decision Making)*

EPA's Model Evaluation and Applications Research Branch is currently performing a retrospective analysis of the CMAQ model's ability to simulate the change in a pollutant associated with a known change in emissions (A. Gilliland, EPA, personal commun., May 19, 2006 and March 5, 2007). This study, which EPA terms a "dynamic evaluation" study, focuses on a rule issues (sic) by EPA in 1998 that required 22 states and the District of Columbia to submit State Implementation Plans providing NO_x emission reductions to mitigate ozone transport in the eastern United States. This rule, know (sic) as the NO_x SIP Call, requires emission reductions from the utility sector and large industrial boilers in the eastern and midwestern United States by 2004. Since theses sources are equipped with continuous emission monitor systems, the NOx SIP call represents a special opportunity to directly measure the emission changes and incorporate them into model simulations with reasonable confidence. Air quality model simulations were developed for summers 2002 and 2004 using the CMAQ model, and the resulting ozone predictions were compared to observed ozone concentrations. Two series of CMAQ simulations have been developed to test two different chemical mechanisms in CMAQ to consider model uncertainty that is associated with the representation of chemistry in the model. Given that regulatory applications use the model's prediction of the relative change in pollutant concentrations, dynamic evaluations such as these are particularly relevant to the way the model is used.

4.5 Summary of Model Evaluation

The individual elements of model evaluation may be summarized as follows (NRC 2007):

- **Scientific basis –** The scientific theories that form the basis for models.
- **Computational infrastructure –** The mathematical algorithms and approaches used in the execution of the model computations.
- **Assumptions and limitations** – The detailing of important assumptions used in the development or application of a computational model as well as the resulting limitations in the model that will affect the model's applicability.
- **Peer review** – The documented critical review of a model or its application conducted by qualified individuals who are independent of those who performed the work, but who are collectively at least equivalent in technical expertise
- (i.e., peers) to those who performed the original work. Peer review attempts to ensure that the model is technically adequate, competently performed, properly documented, and satisfies established quality requirements through the review of assumptions, calculations, extrapolations, alternate interpretations, methodology, acceptance criteria, and/or conclusions pertaining from a model or its application (modified from EPA 2006a)
- **Quality assurance and quality control (QA/QC) –** A system of management activities involving planning, implementation, documentation, assessment, reporting, and improvement to ensure that a model and its component parts are of the type needed and expected for its task and that they meet all required performance standards.
- **Data availability and quality** – The availability and quality of monitoring and laboratory data that can be used for both developing model input parameters and assessing model results.
- **Test cases –** Basic model runs where an analytical solution is available or an empirical solution is known with a high degree of confidence to ensure that algorithms and computational processes are implemented correctly.
- **Corroboration of model results with observations** – Comparison of model results with data collected in the field or laboratory to assess the accuracy and improve the performance of the model.
- **Benchmarking against other models** – Comparison of model results with other similar models.
- **Sensitivity and uncertainty analysis –** Investigation of what parameters or processes are driving model results as well as the effects of lack of knowledge and other potential sources of error in the model.
- **Model resolution capabilities** – The level of disaggregation of processes and results in the model compared to the resolution needs from the problem statement or model application. The resolution includes the level of spatial, temporal, demographic or other types of disaggregation.
- **Transparency –** The need for individuals and groups outside modeling activities to comprehend either the processes followed in evaluation or the essential workings of the model and its outputs.

5. Model Application

5.1 Introduction to Model Application for Environmental Decision Making

This chapter presents best practices and other recommendations for integrating the results of environmental models into Agency decisions. The model development and evaluation process culminates in a decision to accept (or not accept) the model for use in decision making. The decision to accept the model is not made by the model developer. Rather, this decision is made by the program manager charged with making regulatory decisions, in consultation with the model developers and project team. Model acceptance should be sought by the project team prior to application of the model to avoid confusion and potential re-work. The decision should be informed by good communication of the key findings of the model evaluation process discussed in previous chapters, including the critical issue of uncertainty.

Box 9: Examples of Major EPA Documents that Incorporate a Substantial Amount of Computational Modeling Activities

(From Table 2-2 in NRC Report on Models in Environmental Regulatory Decision Making)

Air Quality

Criteria Documents and Staff Paper for Establishing NAAQS

Summarize and assess exposures and health impacts for the criteria air pollutants (ozone, particulate matter, carbon monoxide, lead, nitrogen dioxide, and sulfur dioxide). Criteria documents include results from exposure and health modeling studies, focusing on describing exposure-response relationships. For example, the particulate matter criteria document placed emphasis on epidemiological models of morbidity and mortality (EPA 2004c). The Staff Paper takes this scientific foundation a step further by identifying the crucial health information and using exposure modeling to characterize risks that serve as the basis for the staff recommendation of the standards to the EPA Administrator. For example, models of the number of children exercising outdoors during those parts of the day when ozone is elevated had a major influence on decisions about the 8-hour ozone national ambient air quality standard (EPA 1996)

State Implementation Plan (SIP) Amendments

A detailed description of the scientific methods and emissions reduction programs a state will use to carry out its responsibilities under the CAA for complying with NAAQS. A SIP typically relies on results from activity, emissions, and air quality modeling. Model generated emissions inventories serve as input to regional air quality models and are used to test alternative emission-reduction schemes to see whether they will result in air quality standards being met (e.g., ADEC 2001; TCEQ 2004). Regional scale modeling has become part of developing state implementation plans for new 8-hour ozone and fine particulate matter standards. States, local governments, and their consultants do this analysis.

Regulatory Impact Assessments for Air Quality Rules

RIAs for air quality regulations document the costs and benefits of major emission control regulations. Recent RIAs have included emissions, air quality, exposure, and health and economic impacts modeling results (e.g., EPA 2004b)

Water Regulations

Total Maximum Daily Load (TMDL) Determinations

For each impaired water body, a TMDL documents a state-designated water quality standard need to meet a designated use for that water body and the amount by which pollutant loads need to be reduced to meet the standard. TMDLs utilize water quality and/or nutrient loading models. States and their consultants do the majority of this modeling, with EPA occasionally doing the modeling for particularly contentious TMDLs (EPA 2002b; George 2004; Shoemaker 2004; Wool 2004).

Leaking Underground Storage Tank Program

Assesses the potential risks associated with leaking underground gasoline storage tanks. At an initial screening level, it may assess one-dimensional transport of a conservative contaminant using an analytical model (Weaver 2004).

Development of Maximum Contaminant Level for Drinking Water

Assess drinking water standards for public water supply systems. Such assessments can include exposure, epidemiology, and dose-response modeling (EPA 2002c; NRC 2001b; 2005b).

Pesticides and Toxic Substances Program

Pre-manufacturing Notice Decisions

Assess risks associated with new manufactured chemicals entering the market. Most chemicals are screened initially as to their environmental and human health risks using structure-activity relationship models.

Pesticide Reassessments

Requires that all existing pesticides undergo a reassessment based on cumulative (from multiple pesticides) and aggregate (exposure from multiple pathways) health risk. This includes the use of pesticide exposure models.

Solid and Hazardous Wastes Regulations

Superfund Site Decision Documents

Includes the remedial investigation, proposed plan, and record of decision documents that detail the characteristics and cleanup of Superfund sites. For many hazardous waste sites, a primary modeling task is using groundwater modeling to assess the movement of toxic substances through the substrate (Burden 2004). The remedial investigation for a mining megasite might include water quality, environmental chemistry, human health risk, and ecological risk assessment modeling (NRC 2005a).

Human Health Risk Assessment

Benchmark Dose (BMD) Technical Guidance Document

EPA relies on both laboratory animal and epidemiological studies in assessing the noncancer effects of chronic exposure to pollutants (that is, the reference dose [RfD] and the inhalation reference concentration, [RfC]). These data are modeled to estimate the human dose-response. EPA recommends the use of BMD modeling, which essentially fits the experimental data to use as much of the available data as possible (EPA 2000).

Guidelines for Carcinogenic Risk Assessment

The cancer guidelines set forth a revised set of recommended principles and procedures to guide EPA scientists and others in assessing the cancer risks resulting from exposure to chemicals or other agents in the environment. One of the principal advancements was to describe approaches that consider mode-of-action data, if available, in the quantitative assessment. The guidelines are also used to inform agency decision makers and the public about risk assessment procedures (EPA 2005a).

Ecological Risk Assessment

Guidelines for Ecological Risk Assessment

The ecological risk assessment guidelines provide general principles and give examples to show how ecological risk assessment can be applied to a wide range of systems, stressors, and biological, spatial, and temporal scales. They describe the strengths and limitations of alternative approaches and emphasize processes and approaches for analyzing data rather than specifying data collection techniques, methods or models (EPA 1998).

5.2 Assessing the Problem

Once a model is accepted for use by decision makers, it is applied to the problem identified in the first stages of the modeling process. Model application commonly involves a shift from hindcasting (testing the model against past observed conditions) in the model development and evaluation phase, to forecasting (predicting a future change) in the application phase. This may involve a collaborative effort between modelers and program staff to devise management scenarios that represent different regulatory alternatives. Some model applications may entail trial-and-error model simulations, where model inputs are changed iteratively until a desired environmental condition is achieved.

The use of the model in a proposed decision requires the incorporation of the model application into the public process. The public process is best served when there is transparency, which is accomplished with good written documentation of relevant characteristics of the model for general readership, as well as sharing of specific model files and data upon the request of external parties such as technical consultants and university scientists.

The following sections describe how the transparency of a modeling process can be achieved and documented (§5.1) and how reasoned, evidence-based decision making (§5.2) can be implemented.

5.3 Transparency

The objective of transparency is to enable communication between modelers, decision makers, and the public. Model transparency is achieved when modeling processes are documented with clarity and completeness at an appropriate level of detail. When models are transparent, they can be used reasonably and effectively in a regulatory decision.

5.3.1 Documentation

Documentation enables decision makers and other users of models to understand the process by which a model was developed and used. In the course of modeling, many choices must be made and options selected which may lead to biases in the model results. Documentation of this process and its limitations and uncertainties is essential to increasing the utility and acceptability of model outcomes. Modelers and project teams should document all relevant information about the model to the extent practicable, particularly when a controversial decision is involved. In legal proceedings, the quality and thoroughness of the written documentation of the model and responses to comments on the model during the public process could affect the outcome of the legal challenge. Models should include a clear explanation of their relationship to the scenario of the particular application. This explanation should describe the limitations of the available information when applied to other scenarios. Disclosure about the state of science used in a model and future plans to update the model can to establish the record of reasoned, evidence based application to inform decisions. For example, EPA successfully defended a challenge to a model used in its TMDL program when it explained that it was basing its decision on the best, available scientific information and that it intended to refine its model as better information surfaced[6].

When a court reviews EPA modeling decisions, they generally give some deference to EPA's technical expertise, unless it is without substantial basis in fact. As discussed in Section 4.2.3 regarding

[6] NRDC v. Muszynski, 268 F.3s 91 (2d Cir., 2001).

corroboration, deviations from empirical observations are to be expected. In substantive legal disputes, the courts generally look to the record supporting EPA's decisions. The record will be examined for justification as to why the model was reasonable[7]. The record should contain not only documentation of model development, evaluation, and application but also the Agency's responses to comments on the model raised during peer review and the public process for the action. The organization of this guidance document offers a general outline for model documentation. A more detailed outline is shown below. These elements are adapted from EPA Region 10's standard practices for modeling projects.

Box 10: Recommended Elements for Model Documentation

1. Management Objectives

- Scope of problem
- Technical objectives that result from management objectives
- Level of analysis needed
- Level of confidence needed

2. Conceptual Model

- System boundaries (spatial and temporal domain)
- Important time and length scales
- Key processes
- System characteristics
- Source description
- Available data sources (quality and quantity)
- Data gaps
- Data collection programs (quality and quantity)
- Mathematical model
- Important assumptions

3. Choice of Technical Approach

- Rationale for approach in context of management objectives and conceptual model
- Reliability and acceptability of approach
- Important assumptions

4. Parameter Estimation

- Data used for parameter estimation
- Rationale for estimates in the absence of data
- Reliability of parameter estimates

5. Uncertainty/Error

- Error/uncertainty in inputs, initial conditions, and boundary conditions
- Error/uncertainty in pollutant loadings
- Error/uncertainty in specification of environment
- Structural errors in methodology (e.g., effects of aggregation or simplification)

6. Results

- Tables of all parameter values used for analysis
- Tables or graphs of all results used in support of management objectives or conclusions
- Accuracy of results

7. Conclusions of analysis in relationship to management objectives

8. Recommendations for additional analysis, if necessary

Note: The QA project plan for models (USEPA 2002b) *includes a documentation and records component that also describes the types of records and level of detailed documentation to be kept depending on the scope and magnitude of the project.*

[7] American Iron and Steel Inst. v. EPA, 115 F.3d 979 (D.C. Cir. 1997).

5.3.2 Effective Communication

The modeling process should include effective communication of uncertainty such that information about uncertainty is provided to anyone who is interested in the model results. All technical information should be documented in a manner that decision makers and stakeholders can readily interpret and understand. Recommendations for improving clarity that have been adapted from the Risk Characterization Handbook (USEPA 2000d) include the following:

- Be brief as possible while still providing all necessary details.
- Use plain language that is understood by modelers, policy makers, and the informed lay person.
- Avoid jargon and excessively technical language. Define specialized terms upon first use.
- Provide the model equations.
- Use clear and appropriate methods to efficiently display mathematical relationships.
- Describe quantitative outputs clearly.
- Use understandable tables and graphics to present technical data (see Morgan and Henrion, 1990 for suggestions).

It is important to clearly identify the conclusions of the modeling project and other relevant points of the modeling project. The challenge is to characterize these essentials for decision makers, while also providing them with more detailed information about the modeling process and its limitations. Decision makers should have sufficient insight into the model framework and its underlying assumptions that they can apply model results appropriately. This is consistent with QA planning practices that assert that the quality of data and any limitations on their use should be discussed with respect to their intended use in all technical reports (USEPA 2000e).

5.4 Application of Multiple Models

The situation sometimes arises where multiple models exist that may be applicable to a certain decision making need, e.g. several air quality models, each with its own strengths and weaknesses, might be applied for regulatory purposes. In other situations stakeholders use alternative models (developed by industry and academic researchers) to produce alternative risk assessments (e.g., CARES pesticide exposure model developed by industry). One approach to address this issue is to use multiple models of varying complexities to simulate the same phenomena (NRC 2007). Using multiple models in such a manner might allow insights into how sensitive results are to different modeling choices and how much trust to put in results from any one model. Some experience has shown that running multiple models increases the confidence in the model results (Manno et al., 2008) Box 11 shows an example of this approach. However, resource limitations or regulatory time constraints may limit the capacity to fully evaluate all possible models.

Box 11: Use of Multiple Models of Varying Complexity for Estimating Mercury in Fish
(From Box 5-4 in NRC Report on Models in Environmental Regulatory Decision Making)

A potential benefit of the clean-air mercury rule, which requires reductions in mercury emissions from coal-fired power plants, is the reduction of human exposure and related health impacts from methylmercury by reducing concentrations of this toxin in fish. There are many challenges and uncertainties in understanding the impact of reductions in atmospheric mercury emissions on human health. In its assessment of the benefits and costs of this rule, EPA used multiple models to look at one particular issue—how changes in atmospheric deposition would affect mercury concentrations in fish—and applied the models to assess some of the uncertainties in this impact (EPA 2005).

EPA based its national-scale benefits assessment on results from the mercury maps (MMaps) model. This model assumes that there is a linear, steady-state relationship between atmospheric deposition of mercury and mercury concentrations in fish and thus assumes that a 50% reduction in mercury deposition rates results in a 50% decrease in fish mercury concentrations. In addition, MMaps assumes instantaneous adjustment of aquatic systems and their ecosystems to changes in deposition. Thus, there is no time lag in the conversion of mercury to methylmercury and its bioaccumulation in fish. MMaps also does not deal with sources of mercury other than those from atmospheric deposition. Despite those limitations, the agency concluded that no other available model was capable of performing a national-scale assessment.

To further investigate fish mercury concentrations and assess the effects of MMaps assumptions, EPA applied more detailed models, including the spreadsheet-based ecological risk assessment for the fate of mercury (SERAFM) model, to five well-characterized ecosystems. As opposed to the steady-state MMaps model, SERAFM is a dynamic model that calculates the temporal response of mercury concentrations in fish tissues to changes in mercury loading. It includes multiple land-use types for representing watershed loadings of mercury through soil erosion and runoff. SERAFM partitions mercury among multiple compartments and phases, including aqueous phase, abiotic participles (for example, silts), and biotic particles (for example, phytoplankton). Comparisons of SERAFM's predictions with observed fish mercury concentrations for a single fish species in four ecosystems showed that the model underpredicted mean concentrations for one water body, overpredicted mean concentrations for a second water body, and accurately predicted mean concentrations for the other two. The error bars for the observed fish mercury concentrations in these four ecosystems were large, making it difficult to assess the accuracy of the models. Modeling of the four ecosystems also showed how assumed physical and chemical characteristics of the specific ecosystem affected absolute fish mercury concentrations and the length of time before fish mercury concentrations reached steady state.

Although EPA concluded that the best available science supports the assumption of a linear relationship between atmospheric deposition and fish mercury concentrations for broad-scale use, the more detailed ecosystem modeling demonstrated that individual ecosystems were highly sensitive to uncertainties in model parameters. The agency also noted that there were many model uncertainties that could not be quantified. Finally, although the case studies cover the bulk of the key environmental characteristics, extrapolating the individual ecosystem case studies to account for the variability in ecosystems across the country indicated that they might not represent extreme conditions that could affect how atmospheric deposition of mercury would affect fish mercury concentrations in a water body.

This example illustrates the usefulness of investigating a variety of models at varying levels of complexity. A hierarchical modeling approach, such as that used in the mercury analysis, can provide justification for simplified model assumptions or can potentially provide evidence for a consistent bias that would negate the assumption that a simple model is appropriate for broad-scale application.

5.5 Model Post-Audit

Due to time complexity, constraints, scarcity of resources, and/or lack of scientific understanding, technical decisions are often based on incomplete information and imperfect models. Furthermore, even if model developers strive to use the best science available, advances in knowledge and understanding are ongoing. Given this reality, decision makers should use model results in the context of an iterative, ever-improving process of continuous model refinement. Given the continuously evolving state of science and modeling knowledge and the need to demonstrate the accountability of model-based decisions, the practice of model post-audit is important to assess and improve models and their ability to provide valuable predictions for management decisions. Whereas corroboration demonstrates the degree to which a model corresponds to past system behavior, a model post-audit assesses its ability to model future conditions (Anderson and Woessner 1992). A model post-audit involves monitoring the modeled system, after implementation of a remedial or management action, to determine whether the actual system response concurs with that predicted by the model. Due to resource constraints, post-auditing of all models is not considered feasible. Targeted audits on commonly used models may provide valuable information for improvement of model frameworks and/or model parameter estimates. In its review of the TMDL program, the National Research Council recommended that EPA implement this approach by selectively targeting "some post implementation TMDL compliance monitoring for verification data collection to assess model prediction error" (NRC 2001). It is also important to evaluate the process of model use in decision making to evaluate the effectiveness of the methods used to engage decision makers and other stakeholders in the modeling process (Manno et al., 2008).

Appendix A: Glossary of Frequently Used Terms

Accuracy: Closeness of a measured or computed value to its "true" value, where the "true" value is obtained with perfect information. Due to the natural heterogeneity and stochasticity of many environmental systems, this "true" value exists as a distribution rather than a discrete value. In these cases, the "true" value will be a function of spatial and temporal aggregation.

Algorithm: A precise rule (or set of rules) for solving some problem.

Analytical Models: Models that can be solved mathematically in terms of analytical functions. For example, some models that are based on relatively simple differential equations can be solved analytically by combinations of polynomials, exponential, trigonometric, or other familiar functions.

Applicability and Utility: One of EPA's five Assessment Factors (see definition) that describes the extent to which the information is relevant for the Agency's intended use (USEPA 2003b).

Application Niche: The set of conditions under which the use of a model is scientifically defensible. The identification of application niche is a key step during model development. Peer review should include an evaluation of application niche. An explicit statement of application niche helps decision makers to understand the limitations of the scientific basis of the model (USEPA 1993).

Application Niche Uncertainty: Uncertainty as to the appropriateness of a model for use under a specific set of conditions (see application niche).

Assessment Factors: Considerations recommended by EPA for evaluating the quality and relevance of scientific and technical information. These include: (1) soundness, (2) applicability and utility, (3) clarity and completeness, (4) uncertainty and variability, (5) evaluation and review (USEPA 2003b).

Bias: Systematic deviation between a measured (i.e., observed) or computed value and its "true" value. Bias is affected by faulty instrument calibration and other measurement errors, systematic errors during data collection, and sampling errors such as incomplete spatial randomization during the design of sampling programs.

Boundaries: The spatial and temporal conditions and practical constraints under which environmental data are collected. Boundaries specify the area or volume (spatial boundary) and the time period (temporal boundary) to which a model application will apply (USEPA 2000a).

Boundary Conditions: Sets of values for state variables and their rates along problem domain boundaries, sufficient to determine the state of the system within the problem domain.

Calibration: The process of adjusting model parameters within physically defensible ranges until the resulting predictions give the best possible fit to the observed data (USEPA 1994b). In some disciplines, calibration is also referred to as "parameter estimation" (Beck et al 1994).

Checks: Specific tests in a quality assurance plan that are used to evaluate whether the specifications (performance criteria) for the project developed at its onset have been met.

Clarity and Completeness: One of EPA's five Assessment Factors (see definition) that describes the degree of clarity and completeness with which the data, assumptions, methods, quality assurance, sponsoring organizations and analyses employed to generate the information are documented (USEPA 2003b).

Class (see object oriented platform): A set of objects that share a common structure and behavior. The structure of a class is determined by the class variables, which represent the state of an object of that class and the behavior is given by the set of methods associated with the class (Booch 1994).

Code: Instructions, written in the syntax of a computer language, which provide the computer with a logical process. Code may also refer to a computer program or subset. The term code describes the fact that computer languages use a different vocabulary and syntax than algorithms that may be written in standard language.

Complexity: The opposite of simplicity. Complex systems tend to have a large number of variables, multiple parts, mathematical equations of a higher order, and are more difficult to solve. In relation to computer models, complexity generally refers to the level in difficulty in solving mathematically posed problems as measured by the time, number of steps or arithmetic operations, or memory space required (called time complexity, computational complexity, and space complexity, respectively).

Conceptual Basis: This is the underlying scientific foundation of model algorithms or governing equations. The conceptual basis for a model is either empirical (based on statistical relationships between observations) or mechanistic (process-based) or a combination. See definitions for: empirical model and mechanistic model.

Conceptual Model: A hypothesis regarding the important factors that govern the behavior of an object or process of interest. This can be an interpretation or working description of the characteristics and dynamics of a physical system (USEPA 1994b).

Confounding Errors: Errors induced by unrecognized effects from variables that are not included in the model. The unrecognized, uncharacterized nature of these errors makes them more difficult to describe and account for in statistical analysis of uncertainty (Small and Fishbeck 1999).

Constants: Quantities with have fixed values (e.g., the speed of light and the gravitational force) representing known physical, biological, or ecological activities.

Corroboration (model): Quantitative and qualitative methods for evaluating the degree to which a model corresponds to reality. In some disciplines, this process has been referred to as validation. In general, the term "corroboration" is preferred because it implies a claim of usefulness and not truth.

Data Uncertainty: Uncertainty (see definition) that is caused by measurement errors, analytical imprecision and limited sample sizes during the collection and treatment of data. Data uncertainty, in contrast to variability (see definition) is the component of total uncertainty that is "reducible" through further study.

Debug: The identification and removal of bugs from computer code. Bugs are errors in computer code that range from typos to misuse of concepts and equations.

Deterministic Model: A model that provides a solution for the state variables rather than a set of probabilistic outcomes. Because this type of model does not explicitly simulate the effects of data uncertainty or variability, changes in model outputs are solely due to changes in model components or in the boundary conditions or initial conditions.

Domain (spatial and temporal): The spatial and temporal domain of a model cover the extent and resolution with respect to time and space for which the model has been developed and over which it should be evaluated.

Domain Boundaries (spatial and temporal): The limits of space and time that bound a model's domain and are specified within the boundary conditions (see boundary conditions).

Dynamic Model: A model providing the time-varying behavior of the state variables.

Empirical Model: An empirical model is one where the structure is determined by the observed relationship among experimental data (Suter 1993). These models can be used to develop relationships that are useful for forecasting and describing trends in behavior but they are not necessarily mechanistically relevant.

Environmental Data: Information collected directly from measurements, produced from models, and compiled from other sources such as databases and literature (USEPA 2002a).

Evaluation (model): The process used to generate information to determine whether a model and its results are of a quality sufficient to serve as the basis for a regulatory decision.

Evaluation and Review: One of EPA's five Assessment Factors (see definition) that describes the extent of independent verification, validation and peer review of the information or of the procedures, measures, methods or models (USEPA 2003b).

Extrapolation: Extrapolation is a process that uses assumptions about fundamental causes underlying the observed phenomena in order to project beyond the range of the data. In general, extrapolation is not considered a reliable process for prediction; however, there are situations where it may be necessary and useful.

Expert Elicitation: a systematic process for quantifying, typically in probabilistic terms, expert judgments about uncertain quantities. Expert elicitation may be used to characterize uncertainty and fill data gaps where traditional scientific research is not feasible or data are not yet available. Typically, the necessary quantities are obtained through structured interviews and/or questionnaires. Procedural steps can be used to minimize the effects of heuristics and bias in expert judgments.

False Positives: Also known as false rejection decision errors. False positives occur when the null-hypothesis or baseline condition is incorrectly rejected based on the sample data. The decision is made assuming the alternate condition or hypothesis to be true when in reality it is false (USEPA 2000a).

False Negatives: Also known as false acceptance decision errors. False negatives occur when the null hypothesis or baseline condition cannot be rejected based on the available sample data. The decision is made assuming the baseline condition is true when in reality it is false (USEPA 2000a).

Forcing/Driving Variables: External or exogenous (from outside the model framework) factors that influence the state variables calculated within the model. These may include, for example, climatic or environmental conditions (temperature, wind flow, oceanic circulation, etc.).

Function: A mathematical relationship between variables.

Forms (models): Models can be represented and solved in different forms, including: analytic, stochastic, and simulation.

Graded approach: process of basing the level of application of managerial controls applied to an item or work according to the intended use of results and degree of confidence needed in the results (USEPA 2002b).

Integrity: One of three main components of quality in EPA's Information Quality Guidelines. Integrity refers to the protection of information from unauthorized access or revision to ensure that the information is not compromised through corruption or falsification (USEPA 2002a).

Intrinsic Variation: The variability (see definition) or inherent randomness in the real-world processes.

Loading: The rate of release of a constituent of interest to a particular receiving medium.

Measurement Errors: Errors in the observed data that are a function of human or instrumental error during collection. Such errors may be independent or random. When a persistent bias or mis-calibration is present in the measurement device, measurement errors may be correlated among observations (Small and Fishbeck 1999). In some disciplines, measurement error may be referred to as observation error.

Mechanistic Model: A model that has a structure that explicitly represents an understanding of physical, chemical, and/or biological processes. Mechanistic models quantitatively describe the relationship between some phenomenon and underlying first principles of cause. Hence, in theory, they are useful for inferring solutions outside of the domain that the initial data was collected and used to parameterize the mechanisms.

Model: A simplification of reality that is constructed to gain insights into select attributes of a physical, biological, economic, or social system. A formal representation of the behavior of system processes, often in mathematical or statistical terms. The basis can also be physical or conceptual (NRC 2007).

Model Coding: The process of translating the mathematical equations that constitute the model framework into a functioning computer program.

Model Framework: The model framework is the system of governing equations, parameterization and data structures that make up the mathematical model. It is a formal mathematical specification of the concepts and procedures of the conceptual model consisting of generalized algorithms (computer code/software) for different site or problem-specific simulations (USEPA 1994b).

Model Framework Uncertainty: The uncertainty in the underlying science and algorithms of a model. Model framework uncertainty is the result of incomplete scientific data or lack of knowledge about the factors that control the behavior of the system being modeled. Model framework uncertainty can also be the result of simplifications necessary to translate the conceptual model into mathematical terms.

Model Pedigree: A qualitative or quantitative determination of the rigor with which a model has been developed and evaluated. In some cases, a model's pedigree may be represented by a track record that reflects the quality of a model's development and evaluation. Model pedigree is concerned with the source of data used in model development, the origin of the model framework, and the extent of evaluation performed on the model.

Modes (of models): Manner in which a model operates. Models can be designed to represent phenomena in different modes. Prognostic (or predictive) models are designed to forecast outcomes and future events, while diagnostic models work "backwards" to assess causes and precursor conditions.

Module: An independent or self contained component of a model, which is used in combination with other components, and forms part of one or more larger programs.

Noise: Inherent variability that the model does not characterize (see definition for variability).

Objectivity: One of three main components of quality in EPA's Information Quality Guidelines. Objectivity includes whether disseminated information is being presented in an accurate, clear, complete and unbiased manner. In addition, objectivity involves a focus on ascertaining accurate, reliable and unbiased information (USEPA 2002a).

Object-Oriented Platforms: Type of user interface that models systems using a collection of cooperating "objects." These objects are treated as instances of a class within a class hierarchy, where a class is a set of objects that share a common structure and behavior. The structure of a class is determined by the class variables, which represent the state of an object of that class and the behavior is given by the set of methods associated with the class (Booch 1994).

Parameters: Terms in the model that are fixed during a model run or simulation but can be changed in different runs as a method for conducting sensitivity analysis or to achieve calibration goals.

Parametric Variation: When the value of a parameter itself is not a constant and includes natural variability. Consequently, the parameter should be described as a distribution (Shelly et al 2000).

Parameter Uncertainty: Uncertainties (see definition) related to parameter values.

Perfect Information: The state of information where there is no uncertainty. The current and future values for all parameters are known with certainty. The state of perfect information includes knowledge about the values of parameters with natural variability.

Precision: The quality of being reproducible in amount or performance. With models and other forms of quantitative information, precision refers specifically to the number of decimal places to which a number is computed as a measure of the "preciseness" or "exactness" with which a number is computed.

Probability Density Function: Mathematical, graphical, or tabular expression of the relative likelihoods with which an unknown or variable quantity may take various values. The sum (or integral) of all likelihoods equals one for discrete (continous) random variables (Cullen and Frey 1999). These distributions arise from the fundamental properties of the quantities we are attempting to represent. For example, quantities formed from adding many uncertain parameters tend to be normally distributed, and quantities formed from multiplying uncertain quantities tend to be lognormal (Morgan and Henrion 1990).

Programs (computer): Instructions, written in the syntax of a computer language, that provide the computer with a step-by-step logical process. Computer programs are also referred to as code.

Qualitative Assessments: Some of the uncertainty in model predictions may arise from sources whose uncertainty cannot be quantified. Examples are uncertainties about the theory underlying the model, the manner in which that theory is mathematically expressed to represent the environmental components, and theory being modeled. The subjective evaluations of experts may be needed to determine appropriate values for model parameters and inputs that cannot be directly observed or measured (e.g., air emissions estimates). Qualitative, corroboration activities may involve the elicitation of expert judgment on the true behavior of the system and agreement with model-forecasted behavior.

Quantitative Assessments: The uncertainty in some sources—such as some model parameters and some input data—can be estimated through quantitative assessments involving statistical uncertainty and sensitivity analyses. In addition, comparisons can be made for the special purpose of quantitatively describing the differences to be expected between model estimates of current conditions and comparable field observations.

Quality: A broad term that includes notions of integrity, utility, and objectivity (USEPA 2002a).
Reducible Uncertainty: Uncertainty in models that can be minimized or even eliminated with further study and additional data (USEPA 1997). See data uncertainty.

Reliability: The confidence that (potential) users have in a model and in the information derived from the model such that they are willing to use the model and the derived information (Sargent 2000). Specifically, reliability is a function of the performance record of a model and its conformance to best available, practicable science.

Robustness: The capacity of a model to perform well across the full range of environmental conditions for which it was designed.

Screening Model: A type of model designed to provide a "conservative" or risk-averse answer. Screening models can be used with limited information and are conservative, and in some cases they can be used in lieu of refined models, even when time or resources are not limited.

Sensitivity: The degree to which the model outputs are affected by changes in a selected input parameters (Beck et al 1994).

Sensitivity Analysis: The computation of the effect of changes in input values or assumptions (including boundaries and model functional form) on the outputs (Morgan and Henrion 1990). The study of how uncertainty in a model output can be systematically apportioned to different sources of uncertainty in the model input (Saltelli et al 2000a). By investigating the "relative sensitivity" of model parameters, a user can become knowledgeable of the relative importance of parameters in the model.

Sensitivity Surface: A theoretical multi-dimensional "surface" that describes the response of a model to changes in its parameter values. A sensitivity surface is also known as a response surface.

Simulation Models: Represent the development of a solution by incremental steps through the model domain. Simulations are often used to obtain solutions for models that are too complex to be solved analytically. For most situations, where a differential equation is being approximated, the simulation model will use finite time step (or spatial step) to "simulate" changes in state variables over time (or space).

Soundness: One of EPA's five Assessment Factors (see definition) that describes the extent to which the scientific and technical procedures, measures, methods or models employed to generate the information are reasonable for and consistent with, the intended application (USEPA 2003b).

Specifications: Acceptance criteria set at the onset of a quality assurance plan that help to determine if the intended objectives of the project have been met. Specifications are evaluated using a series of associated checks (see definition).

State variables: The dependent variables calculated within the model, which are also often the performance indicators of the models that change over the simulation.

Statistical Models: Models built using observations within a probabilistic framework. Include simple linear or multivariate regression models obtained by fitting observational data to a mathematical function.

Steady State Model: A model providing the long-term or time-averaged behavior of the state variables.

Stochasticity: Fluctuations in ecological processes that are due to natural variability and inherent randomness.

Stochastic Model: A model that includes variability (see definition) in model parameters. This variability is a function of: 1) changing environmental conditions, 2) spatial and temporal aggregation within the model framework, 3) random variability. The solutions obtained by the model or output is therefore a function of model components and random variability.

Transparency: The clarity and completeness with which data, assumptions and methods of analysis are documented. Experimental replication is possible when information about modeling processes is properly and adequately communicated (USEPA 2002a).

Uncertainty: The term used in this document to describe lack of knowledge about models, parameters, constants, data, and beliefs. There are many sources of uncertainty, including: the science underlying a model, uncertainty in model parameters and input data, observation error, and code uncertainty. Additional study and collecting more information allows error that stems from uncertainty to be minimized/reduced (or eliminated). In contrast, variability (see definition) is irreducible but can be better characterized or represented with further study (USEPA 2002b, Shelly et al 2000).

Uncertainty Analysis: Investigates the effects of lack of knowledge or potential errors on the model (e.g, the "uncertainty" associated with parameter values) and when conducted in combination with sensitivity analysis (see definition) allows a model user to be more informed about the confidence that can be placed in model results.

Uncertainty and Variability: One of EPA's five Assessment Factors (see definition) that describes the extent to which the variability and uncertainty (quantitative and qualitative) in the information or in the procedures, measures, methods or models are evaluated and characterized (USEPA 2003b).

Utility: One of three main components of quality in EPA's Information Quality Guidelines. Utility refers to the usefulness of the information to the intended users (USEPA 2002a).

Variable: A measured or estimated quantity which describes an object or can be observed in a system and which is subject to change.

Variability: Variability refers to observed differences attributable to true heterogeneity or diversity. Variability is the result of natural random processes and is usually not reducible by further measurement or study (although it can be better characterized) (USEPA 1997).

Verification (code): Examination of the algorithms and numerical technique in the computer code to ascertain that they truly represent the conceptual model and that there are no inherent numerical problems with obtaining a solution (Beck et al 1994).

Appendix A-2: Categories of Environmental Regulatory Models

This section is taken from Appendix C of the NRC Report on Models in Environmental Regulatory Decision Making.

Models can be categorized according to their fit into a continuum of processes that translate human activities and natural systems interactions into human health and environmental impacts. The categories of models that are integral to environmental regulation include human activity models, natural systems models, emissions models, fate and transport models, exposure models, human health and environmental response models, economic impact models, and noneconomic impact models. Examples of models in each of these categories are discussed below.

HUMAN ACTIVITY MODELS

Anthropogenic emissions to the environment are inherently linked to human activities. Activity models simulate the human activities and behaviors that result in pollutants. In the environmental regulatory modeling arena, examples of modeled activities are the following:

- Demographic information, such as the magnitude, distribution, and dynamics of human populations, ranging from national growth projections to local travel activity patterns on the order of hours.
- Economic activity, such as the macroeconomic estimates of national economic production and income, final demands for aggregate industrial sectors, prices, international trade, interest rates, and financial flows.
- Human consumption of resources, such as gasoline or feed, may be translated into pollutant releases, such as nitrogen oxides or nutrients. Human food consumption is also used to estimate exposure to pollutants such as pesticides. Resource consumption in dollar terms may be used to assess economic impacts.
- Distribution and characteristics of land use are used to assess habitat, impacts on the hydrogeologic cycle and runoff, and biogenic pollutant releases.

Human Activity Models

Model	Type	Use	Additional Information
TRANSCAD, TRANPLAN, MinUTP	Travel demand forecasting models	Develops estimations of motor vehicle miles traveled for use in estimating vehicle emissions. Can be combined with geographic information systems (GIS) for providing spatial and temporal distribution of motor vehicle activity.	http://www.caliper.com/tcovu.htm
DRI	Forecasts national economic indicators	Model can forecast over 1,200 economic concepts including aggregate supply, demand, prices, incomes, international trade, interest rates, etc. The eight sectors of the model are: domestic spending, domestic income, tax sector, prices, financial, international trade, expectations, and aggregate supply.	EIA 1993
E-GAS	National and regional economic activity model	Emissions growth factors for various sector for estimating volatile organic compounds, nitrogen oxides, and carbon monoxide emissions.	Young et al. 1994
YIELD	Crop-growth yield model	Predicts temporal and spatial crop yield.	Hayes et al. 1982

NATURAL SYSTEMS PROCESS AND EMISSIONS MODELS

Natural systems process and emissions models simulate the dynamics of ecosystems that directly or indirectly give rise to fluxes of nutrients and other environmental emissions.

Natural Systems Process and Emissions Models

Model	Type	Use	Additional Information
Marine Biological Laboratory General Ecosystem Model (MBL-GEM)	Plot-scale nutrient cycling of carbon and nitrogen	Simulates plot-level photosynthesis and nitrogen uptake by plants, allocation of carbon and nitrogen to foliage, stems, and fine roots, respiration in these tissues, turnover of biomass through litter fall, and decomposition of litter and soil organic matter.	http://ecosystems.mbl.edu/Research/Models/gem/welcome.html
BEIS	Natural emissions of volatile organic compounds	Simulates nitric oxide emissions from soils and volatile organic compound emissions from vegetation. Input to grid models for NAAQS attainment (CAA)	http://www.epa.gov/asmdnerl/biogen.html Vukovich and Pierce 2002
Natural Emissions Model	Natural emissions of methane and nitrous oxide	Models methane and nitrous oxide emissions from the terrestrial biosphere to atmosphere.	http://web.mit.edu/globalchange/www/tem.html#nem

EMISSIONS MODELS

These models estimate the rate or the amount of pollutant emissions to water bodies and the atmosphere. The outputs of emission models are used to generate inventories of pollutant releases that can then serve as an input to fate and transport models.

Emissions Models

Model	Type	Use	Additional Information
PLOAD	Releases to water bodies	GIS bulk loading model providing annual pollutant loads to waterbodies. Conducts simplified analyses of sediment issues, including a bank erosion hazard index.	http://www.epa.gov/ost/basins EPA 2001
SPARROW	Releases to water bodies	Relates nutrient sources and watershed characteristics to total nitrogen. Predicts contaminant flux, concentration, and yield in streams. Provides empirical estimates (including uncertainties) of the fate of contaminants in streams.	http://water.usgs.gov/nawqa/sparrow/ Schwarz et al. 2006
MOBILE MOVES NONROAD	Releases to air	Factors and activities for anthropogenic emissions from mobile sources. Estimates current and future emissions (hydrocarbons, carbon monoxide, nitrogen oxides, particulate matter, hazardous air pollutants, and carbon dioxide) from highway motor vehicles. Model used to evaluate mobile source control strategies, control strategies for state implementation plans, and for developing environmental impact statements, in addition to other research.	http://www.epa.gov/otaq/m6.htm http://www.epa.gov/otaq/nonrdmdl.htm EPA 2004, EPA 2005a, Glover and Cumberworth 2003

FATE AND TRANSPORT MODELS

Fate and transport models calculate the movement of pollutants in the environment. A large number of EPA models fall into this category. They are further categorized into the transport media they represent: subsurface, air, and surface water. In each medium, there are a range of models with respect to their complexity, where the level of complexity is a function of the following:

- The number of physical and chemical processes considered.

- The mathematical representation of those processes and their numerical solution.
- The spatial and temporal scales over which the processes are modeled.

Even though some fate and transport models can be statistical models, the majority is mechanistic (also referred to as process-based models). Such models simulate individual components in the system and the mathematical relationships among the components. Fate and transport model output has traditionally been deterministic, although recent focus on uncertainty and variability has led to some probabilistic models.

Subsurface Models

Subsurface transport is governed by the heterogeneous nature of the ground, the degree of saturation of the subsurface, as well as the chemical and physical properties of the pollutants of interest. Such models are used to assess the extent of toxic substance spills. They can also assess the fate of contaminants in sediments. The array of subsurface models is tailored to particular application objectives, for example, assessing the fate of contaminants leaking from underground gasoline storage tanks or leaching from landfills. Models are used extensively for site-specific risk assessments; for example, to determine pollutant concentrations in drinking-water sources. The majority of models simulate liquid pollutants; however, some simulate gas transport in the subsurface.

Subsurface Models

Model	Type	Use	Additional Information
MODFLOW	3D finite difference for ground water transport	Risk Assessments (RBCA) Superfund Remediation (CERCLA). Modular three-dimensional model that simulates ground water flow. Model can be used to support groundwater management activities.	http://water.usgs.gov/nrp/gwsoftware/modflow2000/modflow2000.html Prudic et al. 2004, Wilson and Naff 2004
PRZM	Hydrogeological	Pesticide leaching into the soil and root zone of plants (FIFRA). Estimates pesticide and nitrogen fate in the crop zone root and can simulate soil temperature, volatilization and vapor phase transport in soil, irrigation, and microbial transformation.	http://www.epa.gov/ceampubl/products.htm EPA 2005b
BIOPLUME	Two-dimensional finite difference and Method of Characteristics (MOC) model	Simulates organic contaminants in groundwater due to natural processes of dispersion, advection, sorption, and biodegradation. Simulates aerobic and anaerobic biodegradation reactions.	http://www.epa.gov/ada/csmos/models.html EPA 1998

Surface Water Quality Models

Surface water quality models are often related to, or are variations of, hydrological models. The latter are designed to predict flows in water bodies and runoff from precipitation, both of which govern the transport of aqueous contaminants. Of particular interest in some water quality models is the mixing of contaminants as a function of time and space, for example, following a point-source discharge into a river. Other features of these models are the biological, chemical, and physical removal mechanisms of contaminants, such as degradation, oxidation, and deposition, as well as the distribution of the contaminants between the aqueous phase and organisms.

Surface Water Quality Models

Model	Type	Use	Additional Information
HSPF	Combined watershed hydrology and water quality	Total maximum daily load determinations TMDL (CWA). Watershed model simulating nonpoint pollutant load and runoff, fate and transport processes in streams.	http://www.epa.gov/ceampubl/swater/hspf/ Donigan 2002
WASP	Compartment modeling for aquatic systems	Supports management decisions by predicting water quality responses to pollutants in aquatic systems. Multicompartment model that examines both the water column and underlying benthos.	http://www.epa.gov/athens/wwqtsc/html/wasp.html Brown 1986, Brown and Barnwell 1987
QUAL2E	Steady-state and quasi-dynamic water quality model	Stream water quality model used as a planning tool for developing TMDLs. The model can simulate nutrient cycles, benthic and carbonaceous demand, algal production, among other parameters.	http://www3.bae.ncsu.edu/ Regional-Bulletins/Modeling-Bulletin/qual2e.html Brown 1986, Brown and Barnwell 1987

Air Quality Models

The fate of gaseous and solid particle pollutants in the atmosphere is a function of meteorology, temperature, relative humidity, other pollutants, and sunlight intensity, among other things. Models that simulate concentrations in air have one of three general designs: plume models, grid models, and receptor models. Plume models are used widely for permitting under requirements to assess the impacts of large new or modified emissions sources on air quality or to assess air toxics (HAPs) concentrations close to sources. Plume models focus on atmosphere dynamics. Grid models are used primarily to assess concentrations of secondary criteria pollutants (e.g., ozone) in regional airsheds to develop plans (SIPs) and rules with the objective of attaining ambient air quality standards (NAAQS). Both atmospheric dynamics and chemistry are important components of 3-D grid models. In contrast to mechanistic plume and grid models, receptor models are statistical; they determine the statistical contribution of various sources to pollutant concentrations at a given location based on the relative amounts of pollutants at source and receptor. Most air quality models are deterministic.

Air Quality Models

Model	Type	Use	Additional Information
CMAQ	3-D Grid	SIP development, NAAQS setting (CAA). The model provides estimates of ozone, particulates, toxics, and acid deposition and simulates chemical and physical properties related to atmospheric trace gas transformations and distributions. Model has three components including, meteorological system, an emissions model for estimating anthropogenic and natural emissions, and a chemistry-transport modeling system.	http://www.epa.gov/asmdnerl/CMAQ/index.html Byun and Ching 1999
UAM	3-D Grid	Model calculates concentrations of inert and chemically reactive pollutants and is used to evaluate air quality, particularly related to ambient ozone concentrations.	Systems Applications International, Inc., 1999
REMSAD	3-D Grid	Using simulation of physical and chemical processes in the atmosphere that impact pollutant concentrations, model calculates concentration of inert and chemically reactive pollutants.	http://www.remsad.com ICF Consulting 2005
ICSC CALPUFF	Plume	PSD permitting; toxics exposure (CAA, TSCA). Non-steady-state air quality dispersion model that simulates long range transport of pollutants.	
CMB	Receptor	Relative contributions of sources. Receptor model used for air resource management purposes.	http://www.epa.gov/scram001/receptor_cmb.htm Coulter 2004

EXPOSURE MODELS

The primary objective of exposure models is to estimate the dose of pollutant which humans or animals are exposed to via inhalation, ingestion and/or dermal uptake. These models bridge the gap between concentrations of pollutants in the environment and the doses humans receive based on their activity. Pharmacokinetic models take this one step further and estimate dose to tissues in the body. Since exposure is inherently tied to behavior, exposure models may also simulate activity, for example a model that estimates dietary consumption of pollutants. In addition to the Lifeline model described below, other examples of models that estimate dietary exposure to pesticides include Calendex and CARES. These models can be either deterministic or probabilistic, but are well-suited for probabilistic methods due to the variability of activity within a population.

Exposure Models

Model	Type	Use	Additional Information
Lifeline	Diet, water and dermal of single chemical	Aggregate dose of pesticide via multiple pathways.	http://www.thelifeline group.org Lifeline Group, Inc. 2006
IEUBK	Multipathway, single chemical	Dose of lead to children's blood via multiple pathways. Estimates exposure from lead in media (air, water, soil, dust, diet, and paint and other sources) using pharmacokinetic models to predict blood lead levels in children 6 months to 7 years old. The model can be used as a tool for the determination of site-specific cleanup levels.	http://www.epa.gov/ superfund/programs/ lead/products.htm EPA 1994
Air Pollutants Exposure Model (APEX)	Inhalation exposure model	Simulates an individual's exposure to an air pollutant and their movement through space and time in indoor or outdoor environments. Provides dose estimates and summary exposure information for each individual.	http://www.epa.gov/ttn/ fera/human_apex.html Richmond et al. 2001

HUMAN HEALTH AND ENVIRONMENT RESPONSE MODELS

Human Health Effects Models

Health effects models provide a statistical relationship between a dose of a chemical and an adverse human health effect. Health effects models are statistical methods, hence models in this category are almost exclusively empirical. They can be further classified as toxicological and epidemiological. The former refer to models derived from observations in controlled experiments, usually with nonhuman subjects. The latter refer to models derived from observations over large populations. Health models use statistical methods and assumptions that ultimately assume cause and effect. Included in this category are models that extrapolate information from non-human subject experiments. Also, physiologically based pharmacokinetic models can help predict human toxicity to contaminants through mathematical modeling of absorption, distribution, storage, metabolism, and excretion of toxicants. The output from health models is almost always a dose, such as a safe level (for example, reference dose [RfD]), a cancer potency index (CPI), or an expected health end point (for example, lethal dose for 50% of the population (LD50) or number of asthma cases). There also exist model *applications* that facilitate the use of the statistical methods.

Human Health Effects Models

Model	Type	Use	Additional Information
Benchmark dose model	Software tool for applying a variety of statistical models to analyze dose-response data	To estimate risk of pollutant exposure. Models fit to dose-response data to determine a benchmark dose that is associated with a particular benchmark response.	http://cfpub.epa.gov/ ncea/cfm/recordisplay. cfm?deid=20167 EPA 2000
Linear Cancer model	Statistical analysis method	To estimate the risk posed by carcinogenic pollutants	

Ecological Effects Models

Ecological effects models, like human health effects models, define relationships between a level of pollutant exposure and a particular ecological indicator. Many ecological effects models simulate aquatic environments, and ecological indicators are related directly to environmental concentrations. Examples of ecological effects indicators that have been modeled are: algae blooms, BOD, fish populations, crop yields, coast line erosion, lake acidity, and soil salinity.

Ecological Effects Models

Model	Type	Use	Additional Information
AQUATOX	Integrated fate and effects of pollutants in aquatic environment	Ecosystem model that predicts the environmental fate of chemicals in aquatic ecosystems, as well as direct and indirect effects on the resident organisms. Potential applications to management decisions include water quality criteria and standards, TMDLs, and ecological risk assessments of aquatic systems.	http://www.epa.gov/waterscience/models/aquatox/ Hawkins 2005, Rashleigh 2007
BASS	Simulates fish populations exposed to pollutants (mechanistic)	Models dynamic chemical bioconcentration of organic pollutants and metals in fish. Estimates are being used for ecological risks to fish in addition to realistic dietary exposures to humans and wildlife.	http://www.epa.gov/athens/research/modeling/bass.html
SERAFM	Steady-state modeling system used to predict mercury concentration in wildlife	Predicts total mercury concentrations in fish and speciated mercury concentrations in water and sediments.	http://www.epa.gov/ceampubl/swater/serafm/index.htm Knightes 2005
PATCH	Movement of invertebrates in their habitat	Provides population estimates of territorial terrestrial vertebrate species over time, in addition to survival and fecundity rates, and orientation of breeding sites. Determine ecological effects of regulation.	http://www.epa.gov/wed/pages/models/patch/patchmain.htm Lawler et al. 2006

ECONOMIC IMPACT MODELS

This category includes a broad group of models that are used in many different aspects of EPA's activities including: rulemaking (regulatory impact assessments), priority setting, enforcement, and retrospective analyses. Models that produce a dollar value as output belong in this category. Models can be divided into cost models, which may include or exclude behavior responses, and benefit models. The former incorporate economic theory on how markets (supply, demand, and pricing) will respond as a result of an action. Economic models are traditionally deterministic, though there is a trend toward greater use of uncertainty methods in cost-benefit analysis.

Economic Impact Models

Model	Type	Use	Additional Information
ABEL	Micro Economic	Assess a single firm's ability to pay compliance costs or fees. Estimates claims from defendants that they cannot afford to pay for compliance, clean-up or civil penalties using information from tax return data and cash-flow analysis. Used for settlement negotiations.	http://iaspub.epa.gov/ edr/edr_proc_qry. navigate?P_LIST_ OPTION_CD=CSDIS& P_REG_AUTH_ IDENTIFIER=1&P_ DATA_IDENTIFIER= 90389&P_VERSION=1
Nonroad Diesel Economic [illegible]	Macro economic for impact of the nonroad diesel [illegible]	Multimarket model to analyze how producers and consumers are expected to respond to compliance costs associated with the rule. [illegible]	http://www.epa.gov/ttn/ atw/nsps/cinsps/ ci_nsps_eia_reportfinal [illegible]

Noneconomic Impact Models

Model	Type	Use	Additional Information
TDM (Travel Demand Management)	Model used to evaluate travel demand management strategies	Evaluates travel demand management strategies to determine vehicle-trip reduction effects. Model used to support transit policies including HOV lanes, carpooling, telecommuting, and pricing and travel subsidies.	http://www.fhwa.dot.go v/environment/cmaqeat/ descriptions_tdm_evalua tion_model.htm
CERES-Wheat	Crop-growth yield model	Simulates effects of planting density, weather, water, soil, and nitrogen on crop growth, development, and yield. Predicts management strategies that impact crop yield.	http://nowlin.css.msu.ed u/wheat_book/
PHREEQE-A	Models effects of acidification on stone	Simulates the effects of acidic solutions on carbonate stone.	Parkhurst et al. 1990

Appendix B: Supplementary Material on Quality Assurance Planning and Protocols

This section consists of a series of text boxes meant to supplement concepts and references made in the main body of the document. They are not meant to provide a comprehensive discussion on QA practices, and each box should be considered as a discrete unit. Individually, each of the text boxes provides additional background material for specific sections of the main document. The complete QA manuals for each subject area discussed in this guidance and referred to below should be consulted for more complete information on QA planning and protocols.

Box B1: Background on EPA Quality System

The EPA Quality System defined in EPA Order 5360.1 A2, Policy and Program Requirements for the Mandatory Agency-wide Quality System (USEPA 2000e), covers environmental data produced from models as well as "any measurement or information that environmental processes, location, or conditions; ecological or health effects and consequences; or the performance of environmental technology." For EPA, environmental data includes information collected directly from measurements, produced from models, and compiled from other sources such as databases and literature.

The EPA Quality System is based on an American National Standard (ANSI 1994). Consistent with minimum specifications of this standard, §6.a.(7) of EPA Order 5360.1 A2 states that EPA organizations will develop a Quality System that includes "approved Quality Assurance (QA) Project Plans, or equivalent documents defined by the Quality Management Plan, for all applicable projects and tasks involving environmental data with review and approval having been made by the EPA QA Manager (or authorized representative defined in the Quality Management Plan). The approval of the QA Project Plan containing the specifications for the product(s) and the checks against those specifications (assessments) for implementation is an important management control assuring records to avoid fiduciary "waste and abuse" (Federal Managers' Financial Integrity Act of 1982[8] with annual declarations including conformance to the EPA Quality System). The assessments (including peer review) support the product acceptance for models and their outputs and approval for use such as supporting environmental management decisions by answering questions, characterizing environmental processes or conditions and direct decision support such as economic analyses (process planned in Group D in the Guidance for QA Project Plans for Modeling). EPA's policies for QA Project Plans are provided in Chapter 5 of the EPA Manual 5360 A1 (USEPA 2000e), EPA Quality Manual for Environmental Programs (USEPA 2000f) for in-house modeling and Requirements for Quality Assurance Project Plans (QA/R-5) (USEPA 2002b) for modeling done through extramural agreements (e.g., contracts 48 CFR 46, grants and cooperative agreements 40 CFR 30, 31, and 35). For Interagency Agreements QA requirements need to be negotiated and written into the agreement if the project is funded by EPA but if funds are received by EPA the EPA Manual 5360 A1 (USEPA 2000e) applies.

EPA Order 5360.1 A2 also states that EPA organization's Quality Systems include "use of a systematic planning approach to develop acceptance or performance criteria for all work covered" and "assessment of existing data, when used to support Agency decisions or other secondary purposes, to verify that they are of sufficient quantity and adequate quality for their intended use."

[8] Federal Managers Financial Integrity Act of 1982, P.L. 97-255 -- (H.R. 1526), September 8, 1982.

Box B2: Configuration Tests Specified in the QA program

During code verification the final set of computer code is scrutinized to assure that the equations are programmed correctly and that sources of error, such as rounding, are minimal. This process is likely to be more extensive for new computer code. For existing code, the criteria used for previous verification, if known, can be described or cited. Any additional criteria specific to the modeling project can be specified, along with how the criteria were established. Possible departures from the criteria are discussed, along with how the departures can affect the modeling process.

Software code development inspections: Software requirements, software design, or code are examined by an independent person or groups other than the author(s) to detect faults, programming errors, violations of development standards, or other problems. All errors found are recorded at the time of inspection, with later verification that all errors found have been successfully corrected.

Software code performance testing: Software used to compute model predictions is tested to assess its performance relative to specific response times, computer processing usage, run time, convergence to solutions, stability of the solution algorithms, the absence of terminal failures, and other quantitative aspects of computer operation.

Tests for individual model module: Checks ensure that the computer code for each module is computing module outputs accurately and within any specific time constraints. (Modules are different segments or portions of the model linked together to obtain the final model prediction.)

Model framework testing: The full model framework is tested as the ultimate level of integration testing to verify that all project-specific requirements have been implemented as intended.

Integration tests: The computational and transfer interfaces between modules need to allow an accurate transfer of information from one module to the next, and ensure that uncertainties in one module are not lost or changed when that information is transferred to the next module. These tests detect unanticipated interactions between modules and track down cause(s) of those interactions. (Integration tests should be designed and applied in a hierarchical way by increasing, as testing proceeds, the number of modules tested and the subsystem complexity.)

Regression tests: All testing performed on the original version of the module or linked modules is repeated to detect new "bugs" introduced by changes made in the code to correct a model.

Stress testing (of complex models): This ensures that the maximum load (e.g., real-time data acquisition and control systems) does not exceed limits. The stress test should attempt to simulate the maximum input, output, and computational load expected during peak usage. The load can be defined quantitatively using criteria such as the frequency of inputs and outputs or the number of computations or disk accesses per unit of time.

Acceptance testing: Certain contractually required testing may be needed before the new model or the client accepts model application. Specific procedures and the criteria for passing the acceptance test are listed before the testing is conducted. A stress test and a thorough evaluation of the user interface is a recommended part of the acceptance test.

Beta testing of the pre-release hardware/software: Persons outside the project group use the software as they would in normal operation and record any anomalies encountered or answer questions provided in a testing protocol by the regulatory program. The users report these observations to the regulatory program or specified developers, who address the problems before release of the final version.

Reasonableness checks: These checks involve items like order-of-magnitude, unit, and other checks to ensure that the numbers are in the range of what is expected.
Note: This section is adapted from (USEPA 2002b).

Box B3: Quality Assurance Planning Suggestions for Model Calibration Activities
Information related to objectives and acceptance criteria for calibration activities that generally appear at the beginning of this QA Project Plan element includes the following:

Objectives of model calibration: This includes expected accomplishments of the calibration and how the predictive quality of the model might be improved as a result of implementing the calibration procedures.

Acceptance criteria: The specific limits, standards, goodness-of-fit, or other criteria on which a model will be judged as being properly calibrated (e.g., the percentage difference between reference data values from the field or laboratory and predicted results from the model). This includes a mention of the types of data and other information that will be necessary to acquire in order to determine that the model is properly calibrated (e.g., field data, laboratory data, predictions from other accepted models). In addition to addressing these questions when establishing acceptance criteria, the QA Project Plan can document the likely consequences (e.g., incorrect decision-making) for selecting data that do not satisfy one or more of these areas (e.g., are non-representative, are inaccurate), as well as procedures in place to minimize the likelihood of selecting such data.

Justifying the calibration approach and acceptance criteria: Each time a model is calibrated, it is potentially altered. Therefore, it is important that the different calibrations, the approaches taken (e.g., qualitative versus quantitative), and their acceptance criteria are properly justified. This justification can refer to the overall quality of the standards being used as a reference or of the quality of the input data (e.g., whether data are sufficient for statistical tests to achieve desired levels of accuracy).

Box B4: Definition of Quality
As defined by EPA's Information Quality Guidelines (USEPA 2002a), quality is a broad-term that includes notions of integrity, utility, and objectivity. Integrity refers to the protection of information from unauthorized access or revision to ensure that it is not compromised through corruption or falsification. In the context of environmental models, often integrity is most relevant to protection of code from unauthorized or inappropriate manipulation (see Box 2). Utility refers to the usefulness of the information to the intended users. The utility of modeling projects is aided by the implementation of a systematic planning approach that includes the development of acceptance or performance criteria (see Box 1). Objectivity involves two distinct elements, presentation and substance. Objectivity includes whether disseminated information is being presented in an accurate, clear, complete and unbiased manner. In addition, objectivity involves a focus on ascertaining accurate, reliable and unbiased information.

EPA's five general assessment factors (USEPA 2003b) for evaluating the quality and relevance of scientific and technical information supporting Agency actions are: (a) soundness, (b) applicability and utility, (c) clarity and completeness, (d) uncertainty and variability, (e) evaluation and review. Soundness refers to the extent to which a model is appropriate for its intended application and is a reasonable representation of reality. Applicability and utility describe the extent to which the information is relevant and appropriate for the Agency's intended use. Clarity and completeness refer to documentation of the data, assumptions, methods, quality controls, and analysis employed to generate the model outputs. Uncertainty and variability highlight the extent to which limitations in knowledge and information and natural randomness in input data and models are evaluated and characterized. Evaluation and review evaluate the extent of independent application, replication, evaluation, validation and peer review of the information or of the procedures, measures, methods or models employed to generate the information.

Appendix C: Best Practices for Model Evaluation

C.1 Introduction

This appendix presents a practical guide to the best practices for model evaluation (please see §4.1 for descriptions of these practices). These best practices are:

- Scientific peer review (§4.1.1)
- Quality assurance project planning (§4.1.2)
- Corroboration (§4.1.3)
- Sensitivity Analysis (§4.1.3)
- Uncertainty Analysis (§4.1.3)

The objective of model evaluation is to determine whether a model is of sufficient quality to inform a regulatory decision. For each of these best practices, this appendix provides a conceptual overview for model evaluation and introduces a suite of "tools" that may be used in partial fulfillment of the best practice. The appropriate use of these tools is discussed and citations to primary references are provided. Users are encouraged to obtain more complete information about tools of interest, including their theoretical basis, details of their computational methods, and the availability of software.
Figure C.1.1 provides an overview of the steps in the modeling process that are discussed in this guidance. Items in bold in the figure, including peer review, model corroboration, uncertainty analysis and sensitivity analysis, are discussed in this section on model evaluation.

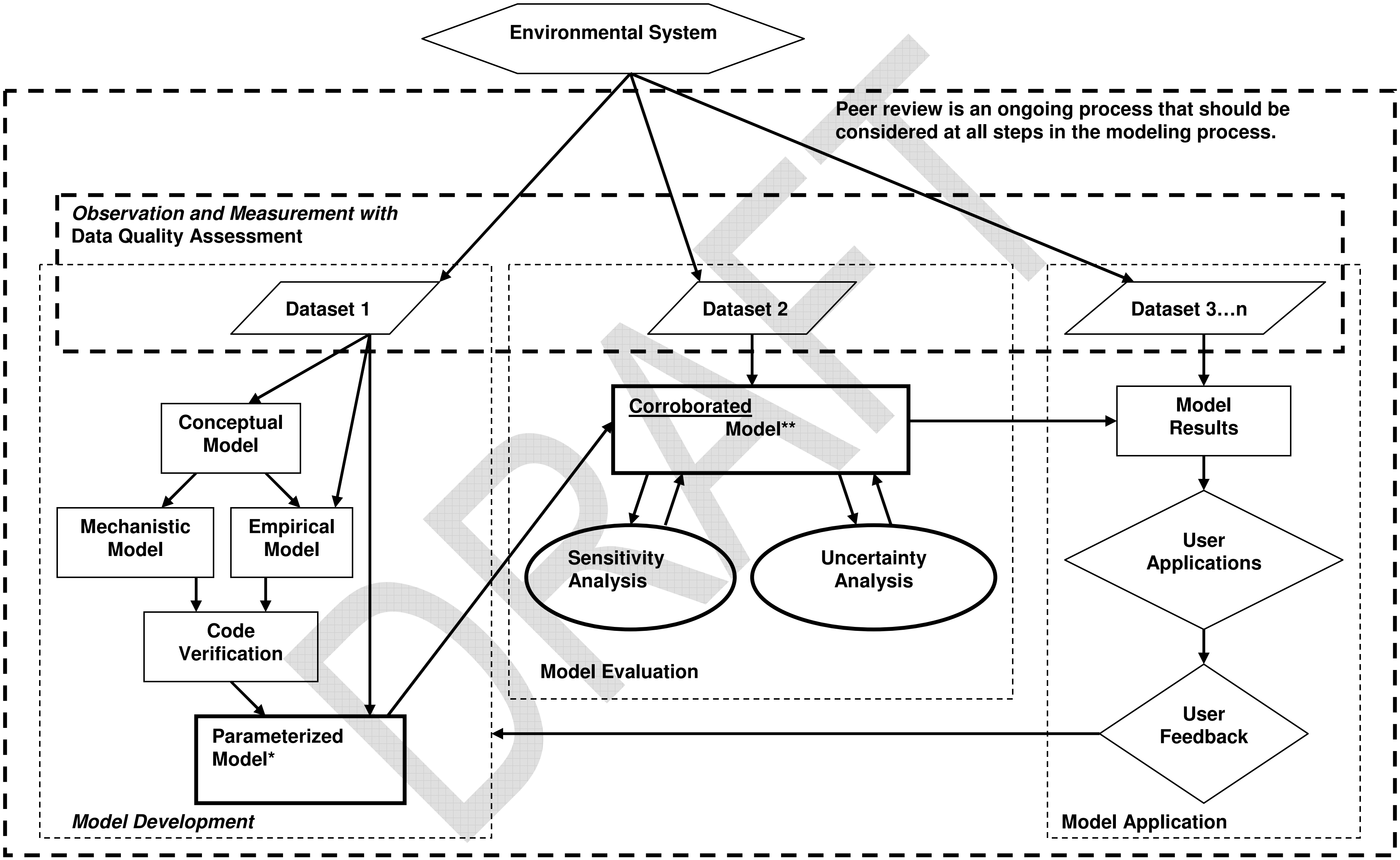


Figure C.1.1. The Modeling Process. * In some disciplines parameterization may include, or be referred to as, calibration.
** Qualitative and/or quantitative corroboration should be performed when necessary.

C.2 Scientific Peer Review

EPA policy states that major scientifically and technically based products related to Agency decisions normally should be peer-reviewed. Agency managers determine and are accountable for the decision whether to employ peer review in particular instances and, if so, its character, scope, and timing. EPA has published guidance to provide a resource for program managers responsible for implementing the peer review process for models (Beck et al 1994). More specifically, this guidance discussed peer review mechanisms, the relationship of external peer review to the process of environmental regulatory model development and application, documentation of the peer review process, and specific elements of what could be covered in an external peer review of model development and application.

The general process for external peer review of models is as follows (Beck et al 1994, Press 1992):

- Step 0: The program manager within the originating office (AA-ship or Region) identifies elements of the regulatory process that would benefit from the use of environmental models. A review/solicitation of currently available models and related research should be conducted. If it is concluded that the development of a new model is necessary, a research/development work plan would be prepared.
- Step 0b: (optional) The program manager may consider internal and/or external peer review of the research/development concepts to determine whether they are of sufficient merit and whether the model is likely to achieve the stated purpose.
- Step 1: The originating office develops a new or revised model or evaluates the possible novel application of model developed for a different purpose.
- Step 1b: (optional) The program manager may consider internal and/or external peer review of the technical or theoretical basis prior to final development, revision or application at this stage. For model development, this review should evaluate the stated application niche.
- Step 2: Initial Agency-wide (internal) peer review/consultation of model development and/or proposed application may be undertaken by the developing originating office. Model design, default parameters, etc. and/or intended application are revised (if necessary) based on consideration of internal peer review comments.
- Step 3: External peer review is considered by the originating office. Model design, default parameters, etc. and/or intended application are revised (if necessary) based on consideration of internal peer review comments.
- Step 4: Final Agency-wide evaluation/consultation may be implemented by the originating office. This step should consist of consideration of external peer review comments and documentation of the Agency's response to scientific/technical issues.

(Note: Steps 2 and 4 are relevant when there is either an internal Agency standing or ad hoc peer review committee or process).

Box C1: Elements of External Peer Review for Environmental Regulatory Models (Box 2-4 from NRC Report on Models in Environmental Regulatory Decision Making)

Model Purpose/Objectives

- What is the regulatory context in which the model will be used and what broad scientific question is the model intended to answer?
- What is the model's application niche?
- What are the model's strengths and weaknesses?

Major Defining and Limiting Considerations

- Which processes are characterized by the model?
- What are the important temporal and spatial scales?
- What is the level of aggregation?

Theoretical Basis for the Model – formulating the basis for problem solution

- What algorithms are used within the model and how were they derived?
- What is the method of solution?
- What are the shortcomings of the modeling approach?

Parameter Estimation

- What methods and data were used for parameter estimation?
- What methods were used to estimate parameters for which there were no data?
- What are the boundary conditions and are they appropriate?

Data Quality/Quantity

Questions related to model design include:

- What data were utilized in the design of the model?
- How can the adequacy of the data be defined taking into account the regulatory objectives of the model?

Questions related to model application include:

- To what extent are these data available and what are the key data gaps?
- Do additional data need to be collected and for what purpose?

Key Assumptions

- What are the key assumptions?
- What is the basis for each key assumption and what is the range of possible alternatives?
- How sensitive is the model toward modifying key assumptions?

Model Performance Measures

- What criteria have been used to assess model performance?
- Did the data bases used in the performance evaluation provide an adequate test of the model?
- How does the model perform relative to other models in this application niche?

Model Documentation and Users Guide

- Does the documentation cover model applicability and limitations, data input, and interpretation of results?

Retrospective

- Does the model satisfy its intended scientific and regulatory objectives?
- How robust are the model predictions?
- How well does the model output quantify the overall uncertainty?

Source: USEPA 1994b.

C.3 Quality Assurance Project Planning

Box C2: Quality Assurance Planning and Data Acceptance Criteria

The QA Project Plan needs to address the following four issues regarding information on how non-direct measurements are acquired and used on the project (USEPA 2002d):

- The need and intended use of each type of data or information to be acquired
- How the data will be identified or acquired, and expected sources of these data
- The method of determining the underlying quality of the data
- The criteria established for determining whether the level of quality for a given set of data is acceptable for use on the project

Acceptance criteria for individual data values generally address issues such as the following:

Representativeness: Were the data collected from a population sufficiently similar to the population of interest and the model-specified population boundaries? Were the sampling and analytical methods used to generate the collected data acceptable to this project? How will potentially confounding effects in the data (e.g., season, time of day, location, and scale incompatibilities) be addressed so that these effects do not unduly impact the model output?

Bias: Would any characteristics of the data set directly impact the model output (e.g., unduly high or low process rates)? For example, has bias in analysis results been documented? Is there sufficient information to estimate and correct bias? If using data to develop probabilistic distributions, are there adequate data in the upper and lower extremes of the tails to allow for unbiased probabilistic estimates?

Precision: How is the spread in the results estimated? Is the estimate of variability sufficiently small to meet the uncertainty objectives of the modeling project as stated in Element A7 (Quality Objectives and Criteria for Model Inputs/Outputs) (e.g., adequate to provide a frequency of distribution)?

Qualifiers: Have the data been evaluated in a manner that permits logical decisions on the data's applicability to the current project? Is the system of qualifying or flagging data adequately documented to allow data from different sources to be used on the same project (e.g., distinguish actual measurements from estimated values, note differences in detection limits)?

Summarization: Is the data summarization process clear and sufficiently consistent with the goals of this project (e.g., distinguish averages or statistically transformed values from unaltered measurement values)? Ideally, processing and transformation equations will be made available so that their underlying assumptions can be evaluated against the objectives of the current project.

C.4 Corroboration

In this guidance, corroboration is defined as all quantitative and qualitative methods for evaluating the degree to which a model corresponds to reality. In practical terms, it is the process of "confronting models with data" (Hilborn and Mangel 1997). In some disciplines, this process has been referred to as validation. In general, the term "corroboration" is preferred because it implies a claim of usefulness and not truth.

Corroboration is used to understand how consist the model is with data. However, uncertainty and variability affect how accurately both models and data represent reality because both models and data (observations) are approximations of some system. Thus, to conduct corroboration meaningfully (i.e., as a tool to assess how well a model represents the system being modeled), this process should begin by characterizing the uncertainty and variability in the corroboration data. As discussed in Section 4.1.3.1, variability stems from the natural randomness or stochasticity of natural systems and can be better captured or characterized in a model but not reduced. In contrast, uncertainty can be minimized with improvements in model structure (framework), improved measurement and analytical techniques, and more comprehensive data for the system being studied. Hence, even a "perfect" model (that contains no measurement error, predicts the correct ensemble average) may deviate from observed field measurements at a given time.

Depending on the type (qualitative and/or quantitative) and availability of data, corroboration can involve hypothesis testing and/or estimates of the likelihood of different model outcomes.

C.4.1 Qualitative Corroboration

Qualitative model corroboration involves expert judgment and tests of intuitive behavior. This type of corroboration uses "knowledge" of the behavior of the system in question, but does not treat model corroboration in a formalized statistical manner. Expert knowledge can establish model reliability through: (a) *consensus* and (b) *consistency*. For example, an expert panel consisting of model developers, other parties and stakeholders could be convened to determine whether there is agreement that the methods and outputs of a model are consistent with processes, standards and results used in other models. Expert judgment can also establish model credibility by determining if model-predicted behavior of a system agrees with best-available understanding of internal processes and functions.

C.4.2 Quantitative Methods

When data are available, model corroboration may involve comparing model predictions to independent empirical observations to investigate how well a model's description of the world fits the observational data. This involves the use of both statistical measures for goodness of fit and numerical procedures to facilitate these calculations. The can be done graphically or by calculating various statistical measures of fit of a model's results to data.

Recall that a model's *application niche* is the set of conditions under which the use of a model is scientifically defensible (§5.2.3); it is the domain of a model's intended applicability. If the model being evaluated purports to estimate an average value across the entire system, then one method to deal with corroboration data is to stratify model results and observed data into "regimes," subsets of data within which system processes operate similarly. Corroboration is then performed by comparing the average of model estimates and observed data within each regime (ASTM 2000).

<u>C.4.2.1 Graphical Methods</u>

Graphical methods can be used to compare the *distribution* of model outputs to independent observations. The degree to which these two distributions overlap, and their respective shapes provide an indication of model performance with respect to the data. Alternately, the differences between observed and predicted data pairs can be plotted and the resulting probability density function (PDF) used to indicate precisions and bias. Graphical methods for model corroboration can be used to indicate bias, precision and kurtosis of model results. Skewness indicates the relative precision of model results, while bias is a reflection of accuracy. Kurtosis refers to the amplitude of the PDF.

<u>C.4.2.2 Deviance Measures</u>

Methods for calculating model bias:

Mean error calculates the average deviation between models and data (e = model-data) by dividing the sum of errors (Σe) by total number of data points compared (m). $MeanError = \frac{\Sigma e}{m}$ (in original measurement units)

Similarly, **mean % error** provides a unit-less measure of model bias: $MeanError(\%) = \frac{\Sigma e / s}{m} * 100$,

Where: "s" is the sample or observational data in original units.

Methods for calculating bias and precision:

Mean square error (MSE): $MSE = \frac{\Sigma e^2}{m}$, large deviations in any single data pair (model-data) can dominate this metric.

Mean absolute error: $MeanAbsError = \frac{\Sigma |e|}{m}$

<u>C.4.2.3 Statistical Tests</u>

A more formal hypothesis testing procedure can also be used for model corroboration. In such cases, a test is performed to determine if the model outputs are statistically significantly different from the empirical data. Important considerations in these tests are the probability of making type I and type II errors and the shape of the data distributions as most of these metrics assume the data are distributed normally. The test-statistic used should also be based on the number of data-pairs (observed and predicted) available.

There are a number of comprehensive texts that may help analysts determine the appropriate statistical and numerical procedures for conducting model corroboration. These include:

- Efron, B. and R. Tibshirani, *An Introduction to the Bootstrap*. 1993. Chapman and Hall, New York.
- Gelman, A.J.B., H.S. Carlin and D.B. Rubin, *Bayesian Data Analysis*. 1995. Chapman and Hall, New York.
- McCullagh, P. and J.A. Nelder, *Generalized Linear Models*. 1989. Chapman and Hall, New York.
- Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, *Numerical Recipes*. 1986. Cambridge University Press, Cambridge, UK.
- Snedecor, G.W. and W.G. Cochran, *Statistical Methods*, 1989. Eighth Ed., Iowa State University Press.

C.4.3 Evaluating Multiple Models

> *Models are metaphorical (albeit sometimes accurate) descriptions of nature, and there can never be a "correct" model. There may be a "best" model, which is more consistent with the data than any of its competitors, or several models may be contenders because each is consistent in some way with the data and none clearly dominates the others. It is the job of the ecological detective to determine the support that the data offer for each competing model or hypothesis.*
> - Hillborn and Mangel 1997, *Ecological Detective*

In the simplest sense, a first cut of model performance is obtained by examining which model minimizes the sum of squares between observed and model-predicted data.

Sum of Squares (SSq):

$$SSq = \sum (pred - obs)^2$$

Sum of squares is equal to the squared differences between model predicted values and observational values. If data are used to fit models and estimate parameters, the fit will automatically improve with each higher order model, e.g., simple linear model: $y = a+bX$ vs. a polynomial model: $y = a+bX+cX^2$.

It is therefore useful to apply a penalty for additional parameters to determine if the improvement in model performance (minimizing SSq deviation) justifies an increase in model complexity. The question is essential whether the decrease in the sum of squares is statistically significant.

The SSq is best applied when comparing several models using a single data set. However, if several data sets are available the Normalized Mean Square Error (NMSE) is typically a better statistic, as it is normalized to the product of the means of the observed and predicted values (see discussion and references §C.4.4.4).

C.4.4 An Example Protocol for Selecting a Set of Best Performing Models

During the development phase of an air quality dispersion model and in subsequent upgrades, model performance is constantly evaluated. These evaluations generally compare simulation results using simple methods that do not account for the fact that models only predict a portion of the variability seen in the observations. To fill a part of this void, the U.S. Environmental Protection Agency (EPA) developed a standard that has been adopted by the ASTM International, designation D6589 – 00 for Statistical Evaluation of Atmospheric Dispersion Model Performance (ASTM 2000). The following discussion summarizes some of the issues discussed in D6589.

<u>C.4.4.1 Define Evaluation Objectives</u>

Performing a statistical model evaluation involves defining those evaluation objectives (features or characteristics) within the pattern of observed and modeled concentration values that are of interest to compare. As yet, no one feature or characteristic has been found that can be defined within a concentration pattern that will fully test a model's performance. For instance, the maximum surface concentration may appear unbiased through a compensation of errors in estimating the lateral extent of the dispersing material and in estimating the vertical extent of the dispersing material. Adding into

consideration that other biases may exist (e.g., in treatment of the chemical and removal processes during transport, in estimating buoyant plume rise, in accounting for wind direction changes with height, in accounting for penetration of material into layers above the current mixing depth, in systematic variation in all of these biases as a function of atmospheric stability), one can appreciate that there are many ways that a model can falsely give the appearance of good performance.
In principle, modeling diffusion involves characterizing the size and shape of the volume into which the material is dispersing as well as the distribution of the material within this volume. Volumes have three dimensions, so a model evaluation will be more complete if it tests the model's ability to characterize diffusion along more than one of these dimensions.

C.4.4.2 Define Evaluation Procedures
Having selected evaluation objectives for comparison, the next step would be to define a evaluation procedure (or series of procedures), which define how each evaluation objective will be derived from the available information. Development of statistical model evaluation procedures begins by providing technical definitions of the terminology used in the goal statement. In the following discussion, we use a plume dispersion model example, but the thought process is valid as well for regional photochemical grid models.

Suppose the evaluation goal is to test the ability of models to replicate the average centerline concentration as a function of transport downwind and as a function of atmospheric stability. Several items are defined to achieve the stated goal, namely: 1) what is an 'average centerline concentration', 2) what is 'transport downwind', and 3) how will 'stability' be defined?

What questions arise in defining the average centerline concentration? Given a sampling arc of concentration values, a decision is needed of whether the centerline concentration is the maximum value seen anywhere along the arc, or whether the centerline concentration is that seen near the center of mass of the observed lateral concentration distribution. If one chooses the latter concept, then a definition is needed of how 'near' the center of mass one has to be, to be representative of a centerline concentration value. One might decide to select all values within a specific range (nearness to the center of mass). In such a case, either a definition or a procedure will be needed to define how this specific range will be determined. A decision will have to be made on the treatment of observed zero (and near measurement threshold) concentrations. To discard such values is to say that low concentrations cannot occur near a plume's center of mass, which is a dubious assumption. One might test to see if conclusions reached regarding 'best performing model' are sensitive to the decision made on the treatment of near-zero concentrations.

What questions arise in defining 'transport downwind'? During near-calm wind conditions, when transport may have favored more than one direction over the sampling period, 'downwind' is not well described by one direction. If plume models are being tested, one might exclude near-calm conditions, since plume models are not meant to provide meaningful results during such conditions. If puff models or grid models are being tested, one might sort the near-calm cases into a special regime for analysis.

What questions arise in defining the 'stability'? For surface releases, surface-layer Monin-Obukhov length, L, has been found to adequately define stability effects, whereas, for elevated releases, Z_i/L, where Z_i is the mixing depth, has been found to be a useful parameter for describing stability effects. Each model likely has its own meteorological processor. It is likely that different processors will have different values for L and Z_i for each of the evaluation cases. There is no one best way to deal with this problem. One solution might be to sort the data into regimes using each of the model's input values, and see if the conclusions reached as to best performing model are affected.

What questions arise if one is grouping data together? If one is grouping data together for which the emission rates are different, one might choose to resolve this by normalizing the concentration values by dividing by the respective emission rates. To divide by the emission rate, one has either a constant emission rate over the entire release, or the downwind transport is sufficiently obvious that one can compute an emission rate based on travel time, that is appropriate for each downwind distance.

Characterizing the plume transport direction is highly uncertain, even with meteorological data collected specific for the purpose. Thus, we expect that the simulated position of the plume will not overlap the observed position of the plume. A decision will have to be made as to how one will compare a feature (or characteristic) in a concentration pattern, when uncertainties in transport direction are large. Will the observed and modeled patterns be shifted, and if so, in what manner?

This discussion is not meant to be exhaustive, but to be illustrative of how the thought process might evolve. It is seen that in defining terms, other questions arise that when resolved will eventually develop an analysis that will compute the evaluation objective from the available data. There likely is more than one answer to the questions that develop. This may cause different people to develop different objectives and procedures for the same goal. If the same set of models is chosen as the best performing, regardless of which path is chosen, one can likely be assured that the conclusions reached are robust.

C.4.4.3 Define Trends in Modeling Bias

In this discussion, references to observed and modeled values refer to the observed and model evaluation objectives (e.g., regime averages). A plot of the observed and modeled values as a function of one of the model input parameters is a direct means for detecting model bias. Such comparison have been recommended and employed in a variety of investigations, e.g., Fox [55], Weil et al. [56], Hanna [57]. In some cases the comparison is the ratio, formed by dividing the modeled value by the observed value, plotted as a function of one or more of the model input parameters. If the data have been stratified into regimes, one can also display the standard error estimates on the respective modeled and observed regime averages. If the respective averages are encompassed by the error bars (typically plus and minus two times the standard error estimates), one can assume the differences are not significant. As described by Hanna [11], this a 'seductive' inference. Procedures to provide a robust assessment of the significance of the differences are defined in ASTM D6589 (ASTM 2000).

C.4.4.4 Summary of Performance

As an example of overall summary of performance, we will discuss a procedure constructed using the scheme introduced by Cox and Tikvart [58] as a template. The design for statistically summarizing model performance over several regimes is envisioned as a five-step procedure.

1. Form a replicate sample using concurrent sampling of the observed and modeled values for each regime. Concurrent sampling associates results from all models with each observed value, so that selection of an observed value automatically selects the corresponding estimates by all models.
2. Compute the average of observed and modeled values for each regime.
3. Compute the Normalize Mean Square Error, *NMSE,* using the computed regime averages, and store the value of the *NMSE* computed for this pass of the bootstrap sampling.
4. Repeat steps 1 through 3 for all Bootstrap sampling passes (typically of order 500).
5. Implement the procedure described in ASTM D 6589 (ASTM 2000) to detect: a) which model has the lowest computed *NMSE* value (call this the 'base' model); b) which models have *NMSE* values that are significantly different from the 'base' model.

In the Cox and Tikvart (1990) analysis, the data were sorted into regimes (defined in terms of Pasquill stability category and low/high wind speed classes), and bootstrap sampling was used to develop standard error estimates on the comparisons. The performance measure was the Robust Highest Concentration (computed from the raw observed cumulative frequency distribution), which is a comparison of the highest concentration values (maxima), which most models do not contain the physics to simulate. This procedure can be improved if intensive field data are used and the performance measure is the **NMSE** computed from the modeled and observed regime averages of centerline concentration values as a function of stability along each downwind arc, where each regimes is a particular distance downwind for a defined stability range.

The data demands are much greater for using regime averages, than for using individual concentrations. Procedures that analyze groups (regimes) of data include intensive tracer field studies, with a dense receptor network, and many experiments. Whereas, Cox and Tikvart (1990) devised their analysis to make use of very sparse receptor networks having one or more years of sampling results. With dense receptor networks, attempts can be made to compare average modeled and 'observed' centerline concentration values, but there are only a few of these experiments that have sufficient data to allow stratification of the data into regimes for analysis. With sparse receptor networks, there are more data for analysis, but there is insufficient information to define the observed maxima relative to the dispersing plume's center of mass. Thus, there is uncertainty as to whether or not the observed maxima are representative of centerline concentration values. It is not obvious that the average of the N (say 25) observed maximum hourly concentration values (for a particular distance downwind and narrowly defined stability range) is the ensemble average centerline concentration the model is predicting. In fact, one might anticipate that the average of the N maximum concentration values is likely to be higher than the ensemble average of the centerline concentration. Thus the testing procedure outlined by Cox and Tikvart (1990) may favor selection of poorly formed models that routinely underestimate the lateral diffusion (and thereby overestimate the plume centerline concentration). This in turn, may bias the performance of such models in their ability to characterize concentration patterns for longer averaging times.

It is therefore concluded that once a set of "best performing models" has been selected from an evaluation using intensive field data that tests a model's ability to predict the average characteristics to be seen in the observed concentration patterns, then evaluations using sparse networks are seen as useful extensions to further explore the performance of well-formulated models for other environs and purposes.

C.5 Sensitivity Analysis

This section provides a broad overview of uncertainty and sensitivity analyses and introduces various methods used to conduct the latter. A table at the end of this section summarizes these methods' primary features and citations to additional resources for computational detail.

C.5.1 Introducing Sensitivity Analyses and Uncertainty Analysis

A model approximates reality in the face of scientific uncertainties. Section 4.1.3.1 identifies and defines various sources of model uncertainty. External peer reviewers of EPA models have consistently recommended that EPA communicate this uncertainty by conducting uncertainty and sensitivity analyses, two related disciplines. <u>Uncertainty analysis</u> investigates the effects of lack of knowledge or potential errors of model inputs (e.g, the "uncertainty" associated with parameter values) and when conducted in combination with sensitivity analysis (see definition) allows a model user to be more informed about the confidence that can be placed in model results. <u>Sensitivity analysis</u> measures the effect of changes in input values or assumptions (including boundaries and model functional form) on the outputs (Morgan and Henrion 1990); it is the study of how uncertainty in a model output can be systematically apportioned to different sources of uncertainty in the model input (Beck et al 1994). By investigating the "relative sensitivity" of model parameters, a user can become knowledgeable of the relative importance of parameters in the model.

Consider a model represented as a function *f*, with inputs x_1 and x_2, and with output *y*, such that $y = f(x_1,x_2)$. Figure C.1 schematically depicts how uncertainty analysis and sensitivity analysis would be conducted for this model. Uncertainty analysis would be conducted by determining how the model output y responds to variation in inputs x_1 and x_2, the graphic depiction of which is referred to as the model's *response surface.* Sensitivity analysis would be conducted by apportioning the respective contributions of x_1 and x_2 to changes in y. The schematic should *not* be construed to imply that uncertainty analysis and sensitivity analysis are sequential events. Rather, uncertainty analysis and sensitivity analysis are generally conducted by trial and error, with each type of analysis informing the other. Indeed, in practice, the distinction between these two related disciplines may be irrelevant. For purposes of clarity, the remainder of this appendix will refer exclusively to sensitivity analysis.

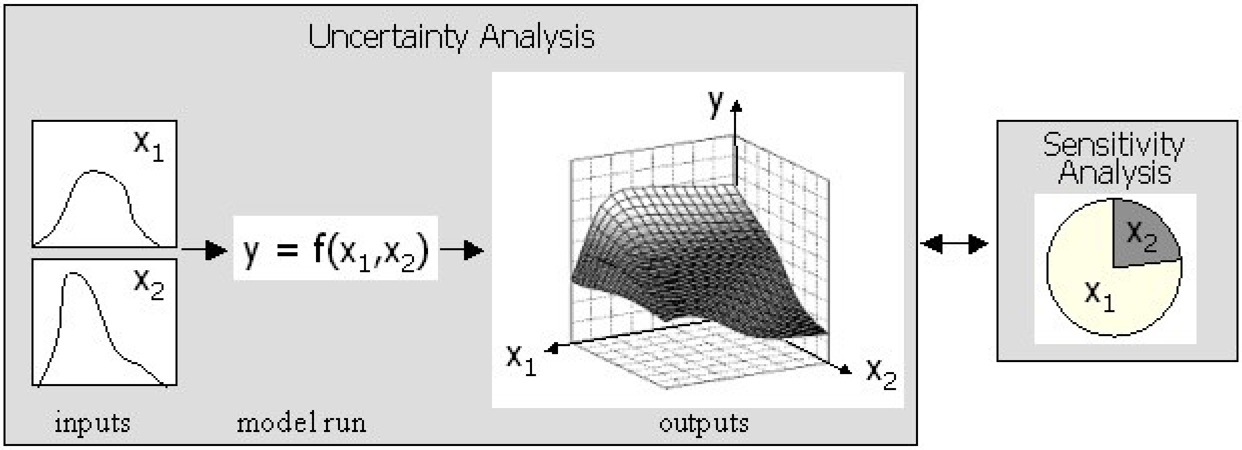


Figure C.5.1 Uncertainty and sensitivity analyses. Uncertainty analysis investigates the effects of lack of knowledge or potential errors of model inputs. Sensitivity analysis evaluates the respective contributions of inputs x_1 and x_2 to output y.

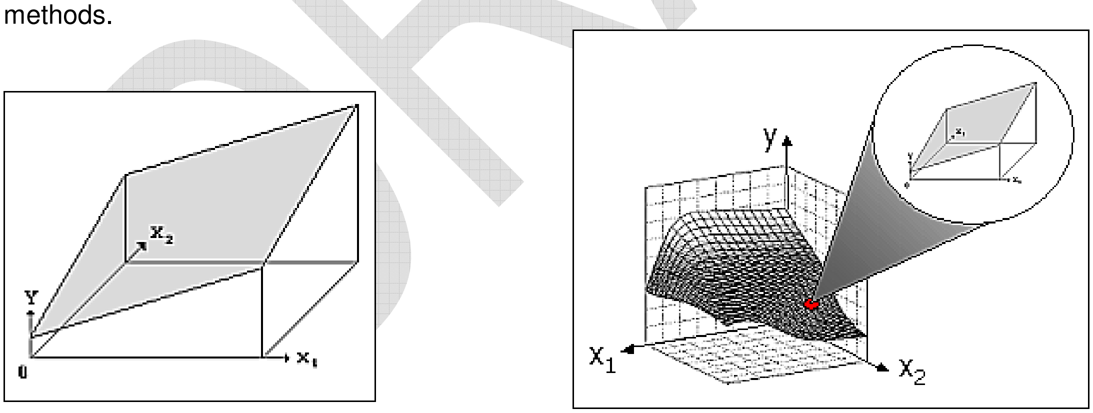
C.5.2 Sensitivity Analysis and Computational Complexity

Choosing the appropriate uncertainty analysis/sensitivity analysis method is often a matter of trading off between the amount of information one wants from the analyses and the computational difficulties of the analyses. These computational difficulties are often inversely related to the number of assumptions one is willing or able to make about the shape of a model's response surface.

Consider once again a model represented as a function *f*, with inputs x_1 and x_2, and with output *y*, such that $y = f(x_1,x_2)$. *Sensitivity* measures how output changes with respect to an input. This is a straightforward enough procedure with differential analysis if the analyst:

(a) is able to assume that the model's response surface is a hyperplane, as in Figure C.5.2 (1);
(b) accepts that the results apply only to specific points on the response surface and that these points are monotonic first order, as in Figure C.5.2 (2);[9] or
(c) is unconcerned about interactions among the input variables.

Otherwise, sensitivity analysis may be more appropriately conducted using more intensive computational methods.



(1) (2)

Figure C.5.2. It's hyperplane and simple. (1) A model response surface that is a hyperplane can simplify sensitivity analysis computations. (2) Alternatively, these same computations can be used for other response surfaces but only as approximations around a single locus.

[9] Related to this issue are the terms *Local* and *Global Sensitivity Analysis*. The former refers to SA conducted around a nominal point of the response surface, while the latter refers to sensitivity analysis across the entire surface.

This guidance suggests that, depending on assumptions underlying the model, the analyst should use non-intensive sensitivity analysis techniques to initially identify those inputs that generate the most sensitivity, then apply more intensive methods to this smaller subset of inputs. It may therefore be useful to categorize the various sensitivity analysis techniques into methods that (a) can be quickly used to screen for the more important input factors; (b) are based on differential analyses; (c) are based on sampling; and (d) based on variance methods.

C.5.3 Screening Tools

C.5.3.1 Tools That Require No Model Runs

Cullen and Frey [50] suggest that summary statistics measuring input uncertainty can serve as preliminary screening tools without additional model runs (and if the models are simple and linear), indicating proportionate contributions to output uncertainty:

(a) *Coefficient of Variation*. The coefficient of variation is the standard deviation normalized to the mean (σ/μ) in order to reduce the possibility that inputs that take on large values are given undue importance.

(b) *Gaussian Approximation.* Another approach to apportioning input variance is Gaussian approximation. Using this method, the variance of a model's output is estimated as the sum of the variances of the inputs (for additive models) or the sum of the variances of the log-transformed inputs (for multiplicative models), weighted by the squares on any constants which may be multiplied by the inputs as they occur in the model (Cullen and Frey 1999).

C.5.3.2 Scatterplots

Cullen and Frey (1999) suggest that a high correlation between an input and an output variable may indicate substantial dependence of the variation in output and the variation of the input. A simple, visual assessment of the influence of an input on the output is therefore possible using scatterplots, with each plot posing a selected input against the output, as in Figure C.5.3.

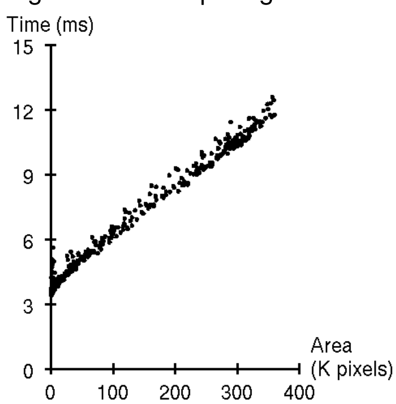


Figure C.5.3. Correlation as indication of input effect. The high correlation between the input variable area and the output variable time (holding all other variables fixed), is an indication of the possible effect of area's variation on the output.

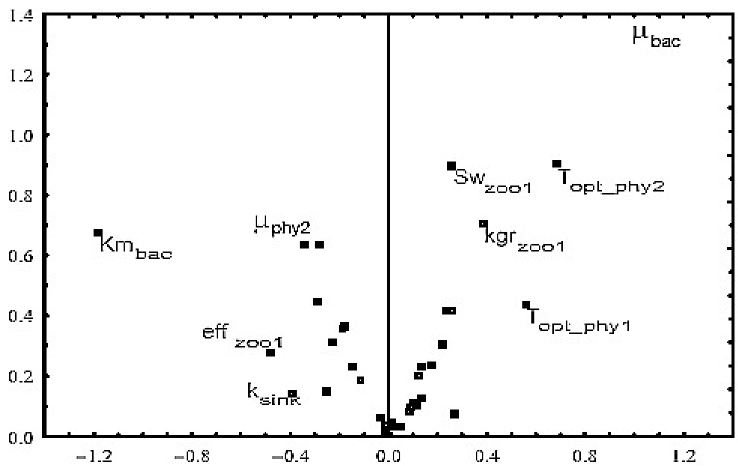
C.5.3.3 Morris's OAT

The key concept underlying one-at-a-time (OAT) sensitivity analyses is to choose a base case of input values and to perturb each input variable by a given percentage away from the base value while holding all other input variables constant. Most OAT sensitivity analysis methods yield *local* measures of sensitivity (see footnote 1) that depend on the choice of base case values. To avoid this bias, Saltelli et al. (2000b) recommend using Morris's OAT for screening purposes because it is a *global* sensitivity analysis method in that the technique entails computing a number of local measures (randomly extracted across the input space) and then taking their average.

Morris's OAT provides a measure of the importance of an input factor in generating output variation, and while it does not quantify interaction effects, it does provide an indication of the presence of interaction.

Figure C.5.4 presents the results that one would expect to obtain from applying Morris's OAT (Cossarini et al 2002). Computational methods for this technique are described in (Saltelli et al 2000b).

Interaction



Importance of input in output variation

Figure C.5.4. **An application of Morris's OAT.** Cossarini et al. (2002) investigated the influence of various ecological factors on energy flow through a food web. Their sensitivity analysis indicated that maximum bacteria growth and bacteria mortality (μ_{bac} and Km_{bac}, respectively) have the largest (and opposite) effects on energy flow, as indicated by their values on the horizontal axis. These effects, as indicated by their values on the vertical axis, resulted from interactions with other factors.

C.5.4 Methods Based on Differential Analysis

As noted previously, differential analyses may be used to analyze sensitivity if the analyst is willing either to assume that the model response surface is hyperplanar or to accept that the sensitivity analysis results are local and that they are based on hyperplanar approximations tangent to the response surface at the nominal scenario (Morgan and Henrion 1990, Saltelli et al 2000b).

Differential analyses entail four steps. First, select base values and ranges for input factors. Second, using these input base values, develop a Taylor series approximation to the output. Third, estimate uncertainty in output in terms of its expected value and variance using variance propagation techniques. Finally, use the Taylor series approximations to estimate the importance of individual input factors (Saltelli et al 2000b). Computational methods for this technique are described in (Morgan and Henrion 1990).

C.5.5 Methods Based on Sampling

One approach to estimating the impact of input uncertainties is to repeatedly run a model using randomly sampled values from the input space. The most well-known method using this approach is Monte Carlo analysis. In a Monte Carlo simulation, a model is run repeatedly. With each run, different input values are drawn randomly from the probability distribution functions of each input, thereby generating multiple output values (Morgan and Henrion 1990, Cullen and Frey 1999). One can view a Monte Carlo simulation as a process through which multiple scenarios generate multiple output values; although each execution of the model run is deterministic, the set of output values may be represented as a cumulative distribution function and summarized using statistical measures (Cullen and Frey 1999).

EPA proposes several best principles of good practice for the conduct of Monte Carlo simulations (USEPA 1997). They include the following:

- Conduct preliminary sensitivity analyses to identify significant model components and input variables that make important contributions to model uncertainty;
- When deciding upon a probability distribution function (PDF) for input variables,

- Consider the following questions: (a) is there any mechanistic basis for choosing a distributional family; (b) is the PDF likely to be dictated by physical, biological or other properties and mechanisms; (c) is the variable discrete or continuous; (d) what are the bounds of the variable; (e) is the PDF symmetric or skewed, and if skewed, in which direction.
- Base the PDF on empirical, representative data;
- If expert judgment is used as the basis for the PDF, document explicitly the reasoning underlying this opinion.
- Discuss the presence or absence of covariance among the input variables, which can significantly affect the output.

The preceding points merely summarize some of the main points raised in EPA's Guidance on Monte Carlo Analysis. That document should be consulted for more detailed guidance. Conducting Monte Carlo analysis may be problematic for models containing a large number of input variables. Fortunately, there are several approaches to dealing with this problem:

- *Brute Force Approach*. One approach is to increase sheer computing power. For example, EPA's ORD is developing a Java-based tool that facilitates Monte Carlo analyses across a cluster of PCs by harnessing the computing power of multiple workstations to conduct multiple runs for a complex model (Babendreier and Castleton 2002).
- *Smaller, structured trials*. The value of Monte Carlo lies not in the randomness of sampling, but rather in achieving representative properties of sets of points in the input space. Therefore, rather than sampling data from entire input space, computations may be through *stratified sampling* by dividing the input sample space into strata and sampling from within each stratum. A widely used method for stratified sampling is *Latin hypercube sampling*, comprehensively described in (Cullen and Frey 1999).
- *Response surface model surrogate*. The analyst may also choose to conduct Monte Carlo not on the complex model directly, but rather on a response surface representation of it. The latter is a simplified representation of the relationship between a selected number of model outputs and a selected number of model inputs, with all other model inputs held at fixed values (Morgan and Henrion 1990, Saltelli et al 2000b).

C.5.6 Methods based on Variance

Consider once again a model represented as a function *f*, with inputs x_1 and x_2, and with output y, such that $y = f(x_1,x_2)$. The input variables are affected by uncertainties and may take on any number of possible values. Let X denote an input vector randomly chosen from among all possible values for x_1 and x_2. The output y for a given X can also be seen as a realization of a random variable Y. Let E(Y│X) denote the expectation of Y conditional on a fixed value of X. If the total variation in y is matched by the variability in E[Y│X] as x_1 is allowed to vary, then this is an indication that variation in x_1 significantly affects y.
The variance-based approaches to sensitivity analysis are based on the estimation of what fraction of total variation of y is attributable to variability in E[Y│X] as a subset of input factors are allowed to vary. Three methods for computing this estimation (correlation ratio, Sobol, and Fourier amplitude sensitivity test) are featured in (Saltelli et al 2000b).

C.5.7 Which Method to Use?

A panel of sensitivity analysis experts was recently assembled to conduct a review of various sensitivity analysis methods. The panel refrained from explicitly recommending a "best" method and instead developed a list of attributes for preferred sensitivity analysis methods. The panel recommended that methods should preferably be able to (a) deal with a model regardless of assumptions about a model's linearity and additivity; (b) consider interaction effects among input uncertainties; and (c) cope with differences in the scale and shape of input PDFs; (d) cope with differences in input spatial and temporal dimensions; and (e) evaluate the effect of an input while all other inputs are allowed to vary as well (Frey 2002, Saltelli 2002). Of the various methods discussed above, only those based on variance (§ C.5.6) are characterized by these attributes. When one or more of the criteria are not important, the other tools discussed in this section will provide a reasonable sensitivity assessment.

As mentioned earlier, choosing the most appropriate sensitivity analysis method will often entail a trade-off between computational complexity, model assumptions, and the amount of information needed from the sensitivity analysis. As an aid to sensitivity analysis method selection, Table 1 below summarizes the features and caveats of the methods discussed above.

Method	Features	Caveats	Reference
Screening Methods	May be conducted independent of model run.	Potential for significant error if model is non-linear.	Cullen and Frey 1999 at pp. 247-8.
Morris's One-at-a-Time	Global sensitivity analysis	Indicates, but does not quantify interactions.	Saltelli et al 2000b at p. 68.
Differential Analyses	Global sensitivity analysis for linear model; local sensitivity analysis for nonlinear model.	No treatment of interactions among inputs. Assumes linearity, monotonicity, and continuity.	Cullen and Frey 1999 at pp. 186-94. Saltelli et al 2000b at 183-91
Monte Carlo Analyses	Intuitive No assumptions regarding response surface.	Depending on number of input variables, may be time-consuming to run, but methods to simplify are available. May rely on assumptions regarding input PDFs.	Cullen and Frey 1999 at pp. 196-237 Morgan and Henrion 1990 at 198-216.
Variance-Based	Robust and independent of model assumptions. Addresses interactions.	May be computationally difficult.	Saltelli et al 2000b at 167-97

C.6 Uncertainty Analysis

C.6.1 Model Suitability

An evaluation of model suitability to resolve application niche uncertainty (§4.1.3.1) should precede any evaluation of data uncertainty and model performance. The extent to which a model is suitable for a proposed application depends on:

- Mapping of model attributes to the problem statement
- The degree of certainty needed in model outputs
- The amount of reliable data available or resources available to collect additional data
- Quality of the state of knowledge on which the model is based
- Technical competence of those undertaking simulation modeling

Appropriate data should be available before any attempt is made to apply a model. A model that needs detailed, precise input data should not be used when such data are unavailable.

C.6.2 Data Uncertainty

There are two statistical paradigms that can be adopted to summarize data. The first employs classical statistics and is useful for capturing the most likely or "average" conditions observed in a given system. This is known as the "frequentist" approach to summarizing model input data. Frequentist statistics rely on measures of central tendency (median, mode, mean values) and represent uncertainty as the deviation from these metrics. A frequentist or "deterministic" model produces a single set of solutions for each model run. In contrast, the alternate statistical paradigm employs a probabilistic framework, which summarizes data according to their "likelihood" of occurrence. Input data are represented as distributions rather than a single numerical value and models outputs capture a range of possible values.

The classical view of probability defines the probability of an event occurring by the value to which the long run frequency of an event or quantity converges as the number of trials increases (Morgan and Henrion 1990). Classical statistics relies on measures of central tendency (mean, median, mode) to define model parameters and their associated uncertainty (standard deviation, standard error, confidence intervals).

In contrast to the classical view, a subjectivist or Bayesian view is that the probability of an event is the degree currently of belief that a person has that it will occur, given all of the relevant information currently known to that person. This framework involves the use of probability distributions based on likelihoods functions to represent model input values and employs techniques like Bayesian updating and Monte Carlo methods as statistical evaluation tools (Morgan and Henrion 1990).

Literature Cited

Literature Cited in Main Text and Appendices A, B, C:

Anderson, M. and Woessner, W., 1992. The role of the postaudit in model validation, *Advances in Water Resources*, 15, p. 167-173

ANSI, 1994. *Specifications and Guidelines for Quality Systems for Environmental Data Collection and Technology Programs*, ANSI/ANSQ, E4-1994.

ASTM, 2000. *Standard Guide for Statistical Evaluation of Atmospheric Dispersion Model Performance (D 6589)*, 2000. (Available at http://222.astm.org), 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428.

Babendreier, J.E. & Castleton, K.J., 2002. *Investigating Uncertainty and Sensitivity in Integrated, Multimedia Environmental Models: Tools for FRAMES-3MRA*. IN: Proceedings of 1st Biennial Meeting of International Environmental Modeling and Software Society, 2**,** 90-95, Lugano, Switzerland.

Beck, M.B., 1987. Water quality modeling: a review of the analysis of uncertainty. *Water Resources Research,* 1987. **23**(8): 1393-1442.

Beck, M., L.A. Mulkey, and T.O. Barnwell, 1994. *Model Validation for Exposure Assessments - DRAFT*, United States Environmental Protection Agency: Athens, Georgia.

Beck, M.E., 2002a. Environmental Foresight and Models: A Manifesto. *Developments in Environmental Modeling*; 22. 2002, Amsterdam: Elsevier. 473.

Beck, B., 2002b. Model evaluation and performance, In: *Encyclopedia of Environmetrics*, Abdel H. El-Shaarawi and W.W. Piegorsch, Editors. John Wiley & Sons: Chichester.

Booch, G., 1994. *Object-Oriented Analysis and Design with Applications*. 2nd ed, Benjamin/Cummings Redwood CA

Borsuk, M.E., C.A. Stow, and K.H. Reckhow, 2002. Predicting the frequency of water quality standard violations: A probabilistic approach for TMDL development. *Environmental Science and Technology*, **36**: 2109-2115.

Cossarini, G., C. Solidoro, and A. Crise. 2002. http://www.iemss.org/iemss2002/proceedings/pdf/volume%20tre/285_cossarini.pdf.

Cox, W.M., and J.A. Tikvart, 1990. A statistical procedure for determining the best performing air quality simulation model. *Atmos. Environ*. **24A**(9):2387-2395.

CREM, 2001. *Proposed Agency Strategy for the Development of Guidance on Recommended Practices in Environmental Modeling*. Draft. The Council on Regulatory Environmental Modeling. US Environmental Protection Agency. July 12, 2001

Cullen, A.C. and H.C. Frey, 1999. *Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*, ed. 326. New York: Plenum Press.

Fox, D.G., 1981. Judging air quality model performance: a summary of the AMS workshop on dispersion model performance. *Bull. Amer. Meteor. Soc*. **62**: 599-609.

Frey, H.C., 2002. Guest Editorial: Introduction to Special Section on Sensitivity Analysis and Summary of NCSU/USDA Workshop on Sensitivity Analysis, *Risk Analysis.* **22**, 539-546.

Hanna S.R., 1988. Air quality model evaluation and uncertainty. *Journal of the Air Pollution Control Association*, **38**: 406-442.

Hanna, S.R., 1993. Uncertainties in air quality model predictions. *Boundary-Layer Met*. **62**:3-20.

Hillborn, R. and M. Mangel, 1997. *The Ecological Detective: Confronting Models with Data*. Princeton, N.J.: Princeton University Press.

Kernigham, B.W. and P.J. Plaugher, 1988. *The Elements of Programming Style*. 2nd ed, June 1988.

Konikow, L.F. and J.D. Bredehoeft, 1992. Ground water models cannot be validated. *Advances in Water Resources*, **15**(1): 75-83.

Levins, S., 1992. The problem of pattern and scale in ecology. *Ecology*. **73**: 1943-1967.

Luis, S.J. and D.B. McLaughlin, 1992. A stochastic approach to model validation. *Advances in Water Resources.* **15**(1): 75-83.

Manno, J., Smardon, R., DePinto, J.V., Cloyd, E.T. and S. Del Grando. 2008. The Use of Models in Great Lakes Decision Making: An Interdisciplinary Synthesis. Randolph G. Pack Environmental Institute, College of Environmental Science and Forestry. Occasional Paper 16, April 2008.

Morgan, G. and M. Henrion, 1990. The Nature and Sources of Uncertainty. In: *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge: Cambridge University Press. pp. 47-72.

NRC, 2001. *Assessing the TMDL Approach to Water Quality Management, Committee to Assess the Scientific Basis of the Total Maximum Daily Approach to Water Pollution Reduction*. Water Science and Technology Board, Division of Earth and Life Studies, National Research Council, National Academy Press, Washington, D.C.

NRC, 2007. *Models in Environmental Regulatory Decision Making*. Committee on Models in the Regulatory Decision Process. Board on Environmental Studies and Toxicology, Division on Earth and Life Studies, National Research Council, National Academy Press, Washington, D.C.

Oreskes, N.M., K. Shrader-Frechette, and K. Belitz, 1994. Verification, validation and confirmation of numerical models in the earth sciences. *Science*, **263**: p. 641-646.

Platt, J.R., 1964. Strong inference. *Science* **146**: 347-352.

Press. W.H., 1992. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge.

Reckhow, K.H., 1994. Water quality simulation modeling and uncertainty analysis for risk assessment and decision making, *Ecological Modeling*, **72**: 1-20.

SAB, 1988. *Review of the Underground Storage Tank (UST) Release Simulation Model*, SAB-EEC-88-029, Environmental Engineering Committee, Science Advisory Board: Washington, D.C.

SAB, 1989. *Resolution on the Use of Mathematical Models by EPA for Regulatory Assessment and Decision-Making*, EPA-SAB-EEC-89-012, Science Advisory Board: Washington, D.C.

SAB, 1993a. *Review of Draft Agency Guidance for Conducting External Peer Review of Environmental Regulatory Modeling*, EPA-SAB-EEC-LTR-93-008, Science Advisory Board: Washington, D.C.

SAB, 1993b. *An SAB Report: Review of MMSoils Component of the Proposed RIA for the RCRA Corrective Action Rule*, EPA-SAB-EEC-94-002, Science Advisory Board: Washington, D.C.

Saltelli, A., S. Tarantola, and F. Campolongo, 2000a. *Sensitivity analysis as an ingredient of modeling.* Statistical Science, **15**: 377-395.

Saltelli, A., K. Chan and M. Scott, eds., 2000b. *Sensitivity Analysis*. New York, John Wiley and Sons.

Saltelli, A., 2002. Sensitivity Analysis for Importance Assessment, Risk Analysis. 2002. 22, 579-590.

Sargent, R.G., 2000. Verification, Validation and Accreditation of Simulation Models. *Proceedings of the 2000 Winter Simulation Conference*, J.A. Joines *et al*. (Eds).

Shelly, A., D. Ford, and B. Beck, 2000. *Quality Assurance of Environmental Models*, 2000, NRCSE Technical Report Series.

Small, M.J. and P.S. Fishbeck, 1999. False precision in Bayesian updating with incomplete models. *Human and Ecological Risk Assessment*. **5**(2): 291-304

Suter, G.W.I., 1993. *Ecological Risk Assessment*. Boca Raton: Lewis Publishers. 528.

U.S. EPA, 1992. *Protocol for Determining the Best Performing Model.* EPA-454/R-92-025. Office of Air Quality Planning and Standards Research Triangle Park, North Carolina, December 1992.

U.S. EPA, 1993. *Review of Draft Agency Guidance for Conducting External Peer Review of Environmental Regulatory Modeling*, EPA-SAB-EEC-LTR-93-008, June 1993.

U.S. EPA, 1994a. *Peer Review and Peer Involvement at the U.S. Environmental Protection Agency*, June 1994.

U.S. EPA, 1994b. *Report of the Agency Task Force on Environmental Regulatory Modeling: Guidance, Support Needs, Draft Criteria and Charter*, EPA 500-R-94-001, U.S. Environmental Protection Agency: Washington, D.C.

USEPA, 1997. *Guiding Principles for Monte Carlo Analysis*, EPA/630/R-97/001, U.S. Environmental Protection Agency: Washington, D.C.

U.S. EPA, 2000a. *Guidance for the Data Quality Objectives Process*, EPA QA/G-4, Office of Environmental Information, US Environmental Protection Agency, August 2000.

U.S. EPA, 2000b. *Guidance for Data Quality Assessment*, EPA QA/G-9, U.S. Environmental Protection Agency: Washington, D.C.

U.S. EPA, 2000c. *Science Policy Council Handbook: Peer Review*, 2nd ed, December 2000.

USEPA, 2000d. *Risk Characterization Handbook*, 2000, Science Policy Council, EPA100-B-00-002, U.S. Environmental Protection Agency: Washington, D.C.

USEPA, 2000e. *Policy and Program Requirements for the Mandatory Agency-Wide Quality System*, May 2000, EPA Order, Classification Number 5360.1 A2.

U.S. EPA, 2000f. *EPA Quality Manual for Environmental Programs*, 5360 A1, May 5, 2000.

U.S. EPA, 2001. *Proposed Agency Strategy for the Development of Guidance on Recommended Practices in Environmental Modeling*, Model Evaluation Action Team, Council for Regulatory Environmental Modeling, U.S. Environmental Protection Agency: Washington, D.C.

U.S. EPA, 2002a. *Information Quality Guidelines.* Office of Environmental Information. US Environmental Protection Agency. Washington, D.C.

U.S. EPA, 2002b. *Quality Assurance Project Plans for Modeling*, EPA QA/G-5M, Office of Environmental Information, US Environmental Protection Agency. December 2002.

U.S. EPA, 2002c. *Guidance on Choosing a Sampling Design for Environmental Data Collection for Use in Developing a Quality Assurance Plan*, EPA QA/G-5S, U.S. Environmental Protection Agency: Washington, D.C.

U.S. EPA, 2002d. *Guidance on Environmental Data Verification and Data Validation*, EPA QA/G-8, U.S. Environmental Protection Agency: Washington, D.C.

U.S. EPA, 2003a. Revision to Guideline on Air Quality Models: Adoption of a Preferred Long Range Transport Model and Other Revisions. *Federal Register*, **68** (72): 18440 – 18482.

U.S. EPA, 2003b. *A Summary of General Assessment Factors for Evaluating the Quality of Scientific and Technical Information*, 2003, Science Policy Council, U.S. Environmental Protection Agency: Washington, DC.

Usunoff, E., J. Carrera, and S.F. Mousavi, 1992. An approach to the design of experiments for discriminating among alternative conceptual models*. Advances in Water Resources*, **15**(3): 199-214.

Weil, J.C., R.I. Sykes, and A. Venkatram, 1992. Evaluating air-quality models: review and outlook. *J. Appl. Meteor.* **31**:1121-1145.

Literature Cited in Boxes in Main Text

ADEC (Alaska Department of Environmental Conservation). 2001. State Air Quality Control Plan, Vol. 2. Analysis of Problems, Control Actions, Section III.C: Fairbanks Transportation Control Program. Adapted July 27, 2001. Alaska Department of Environmental Conservation, Juneau, AK.

Barnwell, T.O., L.C. Brown, and R.C. Whittemore. 2004. Importance of field data in stream water quality modeling using QUAL2E-UNCAS. J. Environ. Eng. 130(6):643-647.

Bredehoeft, J.D. 2003. From models to performance assessment: The conceptualization problem. Ground Water 41(5):571-577.

Bredehoeft, J.D. 2005. The conceptualization model problem-surprise.

Burden, D. 2004. Environmental Decision Making: Principles and Criteria for Groundwater Fate and Transport Models. Presentation at the First Meeting on Models in the Regulatory Decision Process, March 18, 2004, Washington, DC.

Dennis, R.L. 2002. The ozone problem. Pp. 147-169 in Environmental Foresight and Models: A Manifesto, M.B. Beck, ed. New York: Elsevier.

EPA (U.S. Environmental Protection Agency). 1998. Guidelines for Ecological Risk Assessment. EPA/630/R-95/002F. Risk Assessment Forum, U.S. Environmental Protection Agency. April 1998 [online]. Available: http://www.epa.gov/superfund/programs/nrd/era.htm [accessed Nov. 7, 2006].

EPA (U.S. Environmental Protection Agency). 1999c. Description of the MOBILE Highway Vehicle Emissions Factor Model. Office of Mobile Sources, Ann Arbor, MI.

EPA (U.S. Environmental Protection Agency). 2000. Benchmark Dose Technical Guidance Document. EPA/630/R-00/001. External Review Draft. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC [online]. Available: http://www.epa.gov/nceawww1/pdfs/bmds/BMD-External_10_13_2000.pdf [accessed June 12, 2007].
EPA (U.S. Environmental Protection Agency). 2002b. Total Maximum Daily Load for Total Mercury in the Ochlockonee Watershed, GA. Region 4, U.S. Environmental Protection Agency. February 28, 2002 [online]. Available: http://www.epa.gov/Region4/water/tmdl/georgia/ochlockonee/final_tmdls/OchlockoneeHgFinalTMDL.pdf [accessed Nov. 7, 2006].

EPA (U.S. Environmental Protection Agency). 2002c. Perchlorate Environmental Contamination: Toxicological Review and Risk Characterization. External Review Draft. NCEA-1-0503. National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC [online]. Available: http://cfpub.epa.gov/ncea/cfm/recordisplay. cfm?deid=24002 [accessed Nov. 7, 2006].

EPA (U.S. Environmental Protection Agency). 2004b. Final Regulatory Impact Analysis: Control of Emissions from Nonroad Diesel Engines. EPA 420-R-04-007. Assessment and Standards Division, Office of Transportation and Air Quality, U.S. Environmental Protection Agency [online]. Available: http://www.epa.gov/nonroad-diesel/2004fr/420r04007a.pdf [accessed Nov. 9, 2006].

EPA (U.S. Environmental Protection Agency). 2004c. Air Quality Criteria for Particulate Matter, Vol. 1 and 2. EPA 600/P-99/002aF-bF. National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC [online]. Available: http://cfpub.epa.gov/ncea/cfm/partmatt.cfm [accessed Nov. 9, 2006].

EPA (U.S. Environmental Protection Agency). 2005. Guidance on the Use of Models and Other Analyses in Attainment Demonstrations for the 8-hour Ozone NAAQS. Draft Final. Office of Air and Radiation, Office of Air Quality Planning and Standards, Research Triangle Park, NC. February 17, 2005 [online]. Available: http://www.epa.gov/ttn/scram/guidance/guide/draft-final-o3.pdf [accessed April 27, 2007].

EPA (U.S. Environmental Protection Agency). 2005a. Guidelines for Carcinogen Risk Assessment EPA/630/P-03/001F. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC. March 2005 [online]. Available: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=116283 [accessed Nov. 9, 2006].

EPA (U.S. Environmental Protection Agency). 2006a. Peer Review Handbook, 3rd Ed. EPA/100/B-06/002. Prepared for the U.S. Environmental Protection Agency by Members of the Per Review Advisory Group, for EPA's Science Policy Council, Washington, DC [online]. Available: http://epa.gov/peerreview/pdfs/Peer%20Review%20HandbookMay06.pdf [accessed Nov. 10, 2006].

Fenner, K., M. Scheringer, M. MacLeod, M. Matthies, T. McKone, M. Stroebe, A. Beyer, M. Bonnell, A.C. Le Gall, J. Klasmeier, D. Mackay, D. van de Meent, D. Pennington, B. Scharenberg, N. Suzuki, and F. Wania. 2005. Comparing estimates of persistence and long-range transport potential among multimedia models. Environ. Sci. Technol. 39(7):1932-1942.

George, J. 2004. State Perspective on Modeling in Support of the TMDL Program. Presentation at the First Meeting on Models in the Regulatory Decision Process, March 18, 2004, Washington, DC.
Gilliland, A. 2003. Overview of Model Evaluation Plans for CMAQ FY04 Release. Presented at the CMAQ Model Peer Review Meeting, December 17, 2003, Research Triangle Park, NC [online]. Available: http://www.cmascenter.org/r_and_d/first_review/pdf/model_evaluation_plans_for_cmaq04_(gilliland).pdf?temp_id=99999 [accessed Nov. 22, 2006].

Klasmeier, J., M. Matthies, M. MacLeod, K. Fenner, M. Scheringer, M. Stroebe, A.C. Le Gall, T.E. McKone, D. van de Meent, and F. Wania. 2006. Application of multimedia models for screening assessment of long-range transport potential and overall persistence. Environ. Sci. Technol. 40(1):53-60.
Leighton, P.A. 1961. Photochemistry of Air Pollution. New York: Academic Press.

Morales, K.H., L. Ryan, T.L. Kuo, M.M. Wu, and C.J. Chen. 2000. Risk of internal cancers from arsenic in drinking water. Environ. Health Perspect. 108(7):655-661.

Morales, K.H., J.G. Ibrahim, C.J. Chen, and L.M. Ryan. 2006. Bayesian model averaging with applications to benchmark dose estimation for arsenic in drinking water. J. Am. Stat. Assoc. 101(473):9-17.

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

NRC (National Research Council). 2001b. Arsenic in Drinking Water 2001 Update. Washington, DC: National Academy Press.

NRC (National Research Council). 2003. Adaptive Monitoring and Assessment for the Comprehensive Everglades Restoration Plan. Washington, DC: National Academies Press.

NRC (National Research Council). 2005a. Superfund and Mining Megasites. Washington, DC: National Academies Press.

NRC (National Research Council). 2005b. Health Implications of Perchlorate Ingestion. Washington, DC: National Academies Press.

OECD (Organisation for Economic Co-operation and Development). 2002. Report of the OECD/UNEP Workshop on the Use of Multimedia Models for Estimating Overall Environmental Persistence and Long-Range Transport Potential in the Context of PBTs/POPs Assessment. ENV/JM/MONO (2002)15. OECD Series on Testing and Assessment No. 36. Organisation for Economic Co-operation and Development, Paris. August 9, 2002 [online]. Available: http://www.olis.oecd.org/olis/2002doc.nsf/43bb6130e5e86e5fc12569fa005d004c/150147753d5c7f6cc1256c010047d31d/$FILE/JT00130274.PDF [accessed April 27, 2007].

Scheffe, R.D., and R.E. Morris. 1993. A review of the development and application of the Urban Airshed model. Atmos. Environ. B-Urb. 27(1):23-39.

Scheringer, M., M. MacLeod, and F. Wegmann. 2006. The OECD Pov and LRTP Screening Tool, Version 2.0 — Software and Manual. http://www.sust-chem.ethz.ch/downloads/Tool2_0_Manual.pdf [accessed April 27, 2007].

Shoemaker, L. 2004. Modeling and Decision Making Overview. Presentation at the First Meeting on Models in the Regulatory Decision Process, March 18, 2004, Washington, DC.

TCEQ (Texas Commission on Environmental Quality). 2004. Required control strategy elements. Chapter 5 in Revision to the State Implementation Plan (SIP) for the Control of Ozone Air Pollution: Houston/Galveston/Brazoria Ozone Nonattainment Area. Project No. 2004-042-SIP-NR. Texas Commission on Environmental Quality, Austin, TX [online]. Available: http://www.tnrcc.state.tx.us/oprd/sips/june2004hgb_EDrec.html#docs [accessed June 8, 2005].

Volinsky, C.T., D. Madigan, A.E. Raftery, and R.A. Kronmal. 1997. Bayesian model averaging in proportional hazard models: Assessing the risk of the stroke. Appl. Statist. 46(4):433-448.

Weaver, J. 2004. Modeling Leaking Underground Storage Tanks. Presentation at the First Meeting on Models in the Regulatory Decision Process, March 18, 2004, Washington, DC.

Wool, T. 2004. Overview of the TMDL Program and Modeling Approaches. Presentation at the First Meeting on Models in the Regulatory Decision Process, March 18, 2004, Washington, DC.

Literature Cited in Appendix A-2

Birgand, F. 2004. Evaluation of QUAL2E. Pp. 99-106 in Agricultural Non-Point Source Water Quality Models: Their Use and Application, J.E. Parsons, D.L. Thomas, and R.L. Huffman, eds. Southern Cooperative Series Bulletin 398 [online]. Available: http://www3.bae.ncsu.edu/Regional-Bulletins/Modeling-Bulletin/qual2e.html.

Brown, L.C. 1986. Uncertainty Analysis Using QUAL2E. EPA/600/D/86/053. Office of Research and Development, U.S. Environmental Protection Agency.

Brown, L.C., and T. O. Barnwell. 1987. The Enhanced Stream Water Quality Models QUAL2E and QUAL2E-UNCAS: Documentation and User Manual. EPA/600/3-87/007. Environmental Research Laboratory, U.S. Environmental Protection Agency, Athens, GA.

Byun, D.W., and J.K.S. Ching. 1999. Science Algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modeling System. EPA/600/R-99/030. Atmospheric Modeling Division, National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC [online]. Available: http://www.epa.gov/asmdnerl/CMAQ/CMAQscienceDoc.html [accessed June 13, 2007].

Caliper Corporation. 2007. TransCAD [online]. Available: http://www.caliper.com/tcovu.htm [accessed June 13, 2007].

Coulter, C.T. 2004. EPA-CMB8.2 Users Manual. EPA-452/R-04-011. Air Quality Modeling Group, Emissions, Monitoring & Analysis Division, Office of Air Quality Planning & Standards, Research Triangle Park, NC [online]. Available: http://www.epa.gov/scram001/models/receptor/EPACMB82Manual.pdf [accessed June 13, 2007].

Donigian, A.S., Jr. 2002. Watershed Model Calibration and Validation: The HSPF Experience. WEF National TMDL Science and Policy 2002, November 13-16, 2002. Phoenix, AZ [online]. Available: http://hspf.com/TMDL.Nov02.Donigian.Paper.doc [accessed June 13, 2007].

EIA (Energy Information Administration). 1993. Documentation of the DRI Model of the U.S. Economy. [online]. Available: tonto.eia.doe.gov/FTPROOT/modeldoc/m061.pdf [accessed March 31, 2007].

EPA (U.S. Environmental Protection Agency). 1994. Guidance Manual for the Integrated Exposure Uptake Biokinetic Model for Lead in Children. EPA/540/R-93/081. OSWER9285.7-15-1. PB93-963510. Office of Solid Waste and Emergency Response, U.S. Environmental Protection Agency, Washington, DC. February 1994 [online]. Available: http://www.epa.gov/superfund/programs/lead/products.htm [accessed Nov. 2, 2006].

EPA (U.S. Environmental Protection Agency). 1998. BIOPLUME III: Natural Attenuation Decision Support System. User's Manual Version 1.0. EPA/600/R-98/010. U.S. Environmental Protection Agency, Washington, DC.

EPA (U.S. Environmental Protection Agency). 1999. ABEL Model. Environmental Data Registry, U.S. Environmental Protection Agency [online]. Available: http://iaspub.epa.gov/edr/edr_proc_qry.navigate?P_LIST_OPTION_CD=CSDIS&P_REG_AUTH_IDENTIFIER=1&P_DATA_IDENTIFIER=90389&P_VERSION=1 [accessed June 14, 2007].

EPA (U.S. Environmental Protection Agency). 2000. Benchmark Dose Technical Guidance Document. EPA/630/R-00/001. External Review Draft. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC [online]. Available: http://www.epa.gov/nceawww1/pdfs/bmds/BMDExternal_10_13_2000. pdf [accessed June 12, 2007].

EPA (U.S. Environmental Protection Agency). 2001. PLOAD Version 3.0: An ArcView GIS Tool to Calculate Nonpoint Sources of Pollution in Watershed and Stormwater Projects. User's Manual. Office of Water Science, U.S. Environmental Protection Agency. January 2001 [online]. Available: www.epa.gov/waterscience/basins/b3docs/PLOAD_v3.pdf [accessed March 21, 2007].

EPA (U.S. Environmental Protection Agency). 2004. MOVES2004 User Guide. Draft. EPA420-P-04-019. Assessment and Standards Division, Office of Transportation and Air Quality, U.S. Environmental Protection Agency, Washington, DC. November 2004 [online]. Available: http://www.epa.gov/otaq/models/ngm/420p04019.pdf [accessed June 13, 2007].

EPA (U.S. Environmental Protection Agency). 2005a. User's Guide for the Final NONROAD2005 Model. EPA420-R-05-013. Assessment and Standards Division, Office of Transportation and Air Quality, U.S. Environmental Protection Agency, Washington, DC. December 2005 [online]. Available: http://www.epa.gov/otaq/models/nonrdmdl/nonrdmdl2005/420r05013.pdf [accessed June 13, 2007].

EPA (U.S. Environmental Protection Agency). 2005b. PRZM-3, A Model for Predicting Pesticide and Nitrogen Fate in the Crop Root and Unsaturated Soil Zones: Users Manual for Release 3.12.2. EPA/600/R-05/111. U.S. Environmental Protection Agency, Washington, DC.

EPA (U.S. Environmental Protection Agency). 2005c. Integrated Exposure Uptake Biokinetic Model for Lead in Children, Windows Version (IEUBKwin v1.0 build 263). Superfund, U.S. Environmental Protection Agency. December 2005 [online]. Available: http://www.epa.gov/superfund/programs/lead/products.htm#ieubk [accessed June 13, 2007].

EPA (U.S. Environmental Protection Agency). 2005d. Economic Impact Analysis of the Standards of Performance for Stationary Compression Ignition Internal Combustion Engines. EPA-452/R-05-006. Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC. June 2005 [online]. Available: http://www.epa.gov/ttn/atw/nsps/cinsps/ci_nsps_eia_reportfinalforproposal.pdf [accessed June 14, 2007].

EPA (U.S. Environmental Protection Agency). 2006a. Biogenic Emissions Inventory System (BEIS) Modeling. Atmospheric Sciences Modeling Division, Office of Research and Development, U.S. Environmental Protection Agency [online]. Available: http://www.epa.gov/asmdnerl/biogen.html [accessed June 13, 2007].

EPA (U.S. Environmental Protection Agency). 2006b. NONROAD Model (Nonroad Engines, Equipment, and Vehicles). Office of Transportation and Air Quality, U.S. Environmental Protection Agency [online]. Available: http://www.epa.gov/otaq/nonrdmdl.htm [accessed June 13, 2007].

EPA (U.S. Environmental Protection Agency). 2006c. CSMoS Ground-Water Modeling Software. Ground Water Technical Support Center, Risk Management Research, Office of Research and Development, U.S. Environmental Protection Agency [online]. Available: http://www.epa.gov/ada/csmos/models.html [accessed June 13, 2007].

EPA (U.S. Environmental Protection Agency). 2006d. HSPF. Exposure Assessment Models. U.S. Environmental Protection Agency [online]. Available: http://www.epa.gov/ceampubl/swater/hspf/ [accessed June 13, 2007].

EPA (U.S. Environmental Protection Agency). 2006e. Water Quality Analysis Simulation Program (WASP). Ecosystems Research Division, Office of Research and Development, U.S. Environmental Protection Agency [online]. Available: http://www.epa.gov/athens/wwqtsc/html/wasp.html [accessed June 13, 2007].

EPA (U.S. Environmental Protection Agency). 2006f. The Chemical Mass Balance (CMB) Model EPA-CMBv8.2. Receptor Miodeling, Air Quality Models, U.S. Environmental Protection Agency [online]. Available: http://www.epa.gov/scram001/receptor_cmb.htm [accessed June 13, 2007].

EPA (U.S. Environmental Protection Agency). 2006g. AQUATOX. Office of Water Science, U.S. Environmental Protection Agency [online]. Available: http://www.epa.gov/waterscience/models/aquatox/ [accessed June 14, 2007].

EPA (U.S. Environmental Protection Agency). 2006h. BASS. Ecosystems Research Division, Office of Research and Development, U.S. Environmental Protection Agency [online]. Available: http://www.epa.gov/athens/research/modeling/bass.html [accessed June 14, 2007].

EPA (U.S. Environmental Protection Agency). 2007a. Better Assessment Science Integrating Point & Nonpoint Sources (BASINS). Water Quality Models and Tools, Office of Water Science, U.S. Environmental Protection Agency [online]. Available: http://www.epa.gov/waterscience/basins/ [accessed June 13, 2007].

EPA (U.S. Environmental Protection Agency). 2007b. MOBILE6 Vehicle Emission Modeling Software. Office of Transportation and Air Quality, U.S. Environmental Protection Agency [online]. Available: http://www.epa.gov/otaq/m6.htm [accessed June 13, 2007].

EPA (U.S. Environmental Protection Agency). 2007c. Modeling Products. Exposure Assessment, U.S. Environmental Protection Agency [online]. Available: http://www.epa.gov/ceampubl/products.htm [accessed June 13, 2007].

EPA (U.S. Environmental Protection Agency). 2007d. Community Multiscale Air Quality (CMAQ). Atmospheric Science Modeling, Office of Research and Development, U.S. Environmental Protection Agency [online]. Available: http://www.epa.gov/asmdnerl/CMAQ/index.html [accessed June 13, 2007].

EPA (U.S. Environmental Protection Agency). 2007e. Human Exposure Modeling—Air Pollutants Exposure Model (APEX/TRIM.Expo$_{\mathrm{Inhalation}}$). Office of Air and Radiation, U.S. Environmental Protection Agency [online]. Available: http://www.epa.gov/ttn/fera/human_apex.html [accessed June 13, 2007].

EPA (U.S. Environmental Protection Agency). 2007f. Benchmark Dose Software. National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency [online]. Available: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=164443 [accessed June 14, 2007].

EPA (U.S. Environmental Protection Agency). 2007g. SERAFM—Spreadsheet-based Ecological Risk Assessment for the Fate of Mercury. Surface water Models. Exposure Assessment Models, U.S. Environmental Protection Agency [online]. Available: http://www.epa.gov/ceampubl/swater/serafm/index.htm [accessed June 14, 2007].

EPA (U.S. Environmental Protection Agency). 2007h. Program to Assist in Tracking Critical Habitat (PATCH). Western Ecology Division, Office of Research and Development, U.S. Environmental Protection Agency, Corvallis, OR [online]. Available: http://www.epa.gov/wed/pages/models/patch/patchmain.htm [accessed June 14, 2007].

EPA (U.S. Environmental Protection Agency). 2007i. Benefits Analysis Models/Tools. Economics & Cost Analysis Support, U.S. Environmental Protection Agency [online]. Available: http://www.epa.gov/ttnecas1/benmodels.html [accessed June 14, 2007].

Glover, E.L., and M. Cumberworth. 2003. MOBILE6.1 Particulate Emission Factor Model Technical Description—Final Report. EPA420-R-03-001. Assessment and Standards Division, Office of Transportation and Air Quality, U.S. Environmental Protection Agency. January 2003 [online]. Available: http://www.epa.gov/otaq/models/mobile6/r03001.pdf [accessed June 13, 2007].

Hawkins, T. 2005. Critical Evaluation of the Aquatox Model. Carnegie Mellon University [online]. Available:
http://www.ce.cmu.edu/~trh/Professional/Research/Hawkins_CriticalEvaluationOfTheAquatoxModel.pdf [accessed March 31, 2007].

Hayes, J.T., P.A. O'Rourke, W.H. Terjung, and P.E. Todhunter. 1982. A feasible crop yield model for worldwide international food production. Int. J. Biometeorol. 26(3):239-257.

ICF Consulting. 2005. User's Guide to the Regional Modeling System for Aerosols and Deposition (REMSAD): Version 8 [online.] Available:
http://www.remsad.com/documents/remsad_users_guide_v8.00_112305.pdf [accessed March 31, 2007].
ICF International/Systems Applications International. 2006. Regional Modeling System for Aerosols and Deposition (REMSAD) [online]. Available: http://www.remsad.com/ [accessed June 13, 2007].

Knightes, C.D. 2005. SERAFM: An Ecological Risk Assessment Tool for Evaluating Wildlife Exposure Risk Associated with Mercury-Contaminated Sediment in Lake and River Systems. Presentation at EPA Science Forum 2005, May 16-18, 2005, Washington, DC.

Lawler, J.J., D. White, R.P. Neilson, and A.R. Blaustein. 2006. Predicting climate-induced range shifts: Model differences and model reliability. Glob. Change Biol. 12(8):1568-1584.

Lifeline Group, Inc. 2006. Users Manual: LifeLine™ Verson 4.3. Software for Modeling Aggregate and Cumulative Exposures to Pesticides and Chemicals. April 5, 2006 [online]. Available: http://www.thelifelinegroup.org/lifeline/documents/v4.3_UserManual.pdf [accessed March 31, 2007].

Lifeline Group, Inc. 2007. Lifeline Software [online]. Available: http://www.thelifelinegroup.org/index.htm [accessed June 13, 2007].

MBL (Marine Biological Laboratory). 2005. Marine Biological Laboratory General Ecosystem Model (MBL-GEM). The Ecosystem Center, Marine Biological Laboratory, MA [online]. Available: http://ecosystems.mbl.edu/Research/Models/gem/welcome.html [accessed June 13, 2007].

MIT (Massachusetts Institute of Technology). 2006. Natural Emissions Model (NEM). The MIT Integrated Global System Model: Ecosystems Impacts [online]. Available:
http://web.mit.edu/globalchange/www/tem.html#nem [accessed June 13, 2007].

Parkhurst, D.L., D.C. Thoratenson, L.N. Plummer. 1980. PHREEQE: A Computer Program for Geochemical Calculations. U.S. Geological Survey Water Research Investigations 80-96. Reston, VA: U.S. Geological Survey.

Prudic, D.E., L.F. Konikow, and E.R. Banta. 2004. A New Stream- Flow Routing (SFR1) Package to Simulate Stream-Aquifer Interaction with MOD-FLOW-2000. U.S. Geological Survey Open-File Report 2004-1042. U.S. Department of the Interior, U.S. Geological Survey [online]. Available: http://pubs.usgs.gov/of/2004/1042/ [accessed June 13, 2007].

Rashleigh, B. 2007. Assessment of lake ecosystem response to toxic events with the AQUATOX model. Pp. 293-299 in Assessment of the Fates and Effects of Toxic Agents on Water Resources, I.E. Gonenc, V. Koutitonsky, B. Rashleigh, R.A. Ambrose, and J.P. Wolfin, eds. Dordrecht, The Netherlands: Springer.

Richmond, H., T. Palma, G. Glen, and L. Smith. 2001. Overview of APEX (2.0): EPA's Pollutant Exposure Model for Criteria and Air Toxic Inhalation Exposures. Annual Meeting of the International Society of Exposure Analysis, November 4-8, 2001. Charleston, SC.

Ritchie, J.T., and D. Godwin. 2007. CARES Wheat 2.0. Michigan State University [online]. Available: http://nowlin.css.msu.edu/wheat_book/ [accessed June 14, 2007].

Schwarz, G.E., A.B. Hoos, R.B. Alexander, and R.A. Smith. 2006. The SPARROW Surface Water-Quality Model: Theory, Application and User Documentation. U.S. Geological Survey Techniques and Methods 6-B3. U.S. Geological Survey [online]. Available: http://pubs.usgs.gov/tm/2006/tm6b3/PDF.htm. [accessed March 31, 2007].

Shiftan, Y., and J. Suhrbier. 2002. The analysis of travel and emission impacts of travel demand management strategies using activity-based models. Transportation 29(2):145-168.

Sokolov, A.P., C.A. Schlosser, S. Dutkiewicz, S. Paltsev, D.W. Kicklighter, H.D. Jacoby, R.G. Prinn, C.E. Forest, J. Reilly, C. Wang, B. Felzer, M.C. Sarofim, J. Scott, P.H. Stone, J.M. Melillo, and J. Cohen. 2005. The MIT Integrated Global System Model (IGSM) Version 2: Model Description and Baseline Evaluation. Report No. 124. Joint Program on the Science and Policy of Global Change, Massachusetts Institute of Technology [online]. Available: http://web.mit.edu/globalchange/www/abstracts.html#a124 [accessed March 31, 2007].

Systems Applications International, Inc. 1999. User's Guide to the Variable-Grid Urban Airshed Model (UAM-V). Systems Applications International, Inc., San Rafael, CA [online]. Available: http://www.uamv.com/documents/uam-v_1.31_user's_guide.pdf [accessed June 13, 2007].

USGS (U.S. Geological Survey). 2007a. SPARROW Modeling of Surface-Water Quality. U.S. Department of the Interior, U.S. Geological Survey [online]. Available: http://water.usgs.gov/nawqa/sparrow/ [accessed June 13, 2007].

USGS (U.S. Geological Survey). 2007b. MODFLOW-2000 Version 1.17.02. USGS Ground-Water Software. U.S. Department of the Interior, U.S. Geological Survey [online]. Available: http://water.usgs.gov/nrp/gwsoftware/modflow2000/modflow2000.html [accessed June 13, 2007].

Vukovich, J.M., and T. Pierce. 2002. The Implementation of BEIS3 within the SMOKE. 11th International Emission Inventory Conference: Emission Inventories—Partnering for the Future, April 15-18, 2002, Atlanta, GA [online]. Available: http://www.epa.gov/ttn/chief/conference/ei11/modeling/vukovich.pdf [accessed June 13, 2007].

Wilson, J.D., and R.L. Naff. 2004. MODFLOW-2000: The U.S. Geological Survey Modular Ground-Water Model-GMG Linear Equation Solver Package Documentation. U.S. Geological Survey Water Resources

Open-File Report 2004-1261. U.S. Department of the Interior, U.S. Geological Survey [online]. Available: http://pubs.usgs.gov/of/2004/1261/ [accessed June 13, 2007].

Young, T., R. Randolph, and D. Bowman. 1994. Economic Growth Analysis System: Version 2.0. EPA/600/SR-94/139. Air and Energy Engineering, Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park [online]. Available: http://www.p2pays.org/ref/07/0622.pdf [accessed June 13, 2007].