Best Modeling Practices: Model Evaluation

Welcome to CREM's **Best Modeling Practices: Model Evaluation** module!

Table of Contents

PREFACE

EPA's Council for Regulatory Modeling (CREM) aims to aid in the advancement of modeling science and application to ensure model quality and transparency. In follow-up to CREM's Guidance Document on the Development, Evaluation, and Application of Environmental Models (PDF) (99 pp, 1.7 MB, About PDF) released in March 2009, CREM developed a suite of interactive web-based training modules. These modules are designed to provide overviews of technical aspects of environmental modeling and best modeling practices. At this time, the training modules are not part of any certification program and rather serve to highlight the best practices outlined in the Guidance Document with practical examples from across the Agency.

CREM's Training Module Homepage contains all eight of the training modules:

- Environmental Modeling 101
- The Model Life-cycle
- Best Modeling Practices: Development
- Best Modeling Practices: Evaluation
- Best Modeling Practices: Application
- Integrated Modeling 101
- Legal Aspects of Environmental Modeling
- Sensitivity and Uncertainty Analyses
- QA of Modeling Activities *(pending)*

DESIGN

- This training module has been designed with **Tabs** and **Sub-tabs**. The "active" Tabs and Sub-tabs are underlined.
- Throughout the module, definitions for **bold terms** (with the icon) appear in the Glossary.
- The vertical slider feature from the web is annotated with the same image; superscripts have been added for further clarification. The information in the right hand frames (web view) typically appears on next page in the PDF version.

Vertical Slider Feature	Corresponding Figure/Text
[1]**What is a model?**	[1]**Vertical Slider #1**

Image caption.

- Similar to the web version of the modules, these dialogue boxes will provide you with three important types of information:

This box directs the user to additional insight of a topic by linking to other websites or modules

This box directs the user to additional resources (reports, white papers, peer-reviewed articles, etc.) for a specific topic

This box alerts the user to a caveat of environmental modeling or provides clarification on an important concept.

BEST MODELING PRACTICES: EVALUATION

This module builds upon the fundamental concepts outlined in previous modules: Environmental Modeling 101 and The Model Life-cycle. The objectives of this module are to explore the topic of model evaluation and identify the 'best modeling practices' and strategies for the Evaluation Stage of the **model** life-cycle.

Model Evaluation

According to the EPA (2009a) **model evaluation** is defined as:

> "The *process* used to generate information that will determine whether a model and its analytical results are of a sufficient quality to inform a decision."

The process of evaluation is used to address the:

- soundness of the underlying science of the model
- quality and quantity of available data
- degree of correspondence between model output and observed conditions
- appropriateness of a model for a given application

Best Modeling Practices for Model Evaluation:

All models (especially regulatory models) should be continually evaluated at all stages within their life-cycle. The *Guidance on the Development, Evaluation, and Application of Environmental Models* (EPA, 2009a) describes best practices for model evaluation that include the following activities:

- Quality Assurance (QA) project planning
- Peer review
- Model corroboration
- **Sensitivity analysis** (SA)
- **Uncertainty analysis** (UA)

WHY ARE MODELS EVALUATED?

There is inherent **uncertainty** associated with models which can lead to (Sunderland, 2008):

- Extreme and inefficient approaches to problem solving
- Delays in decision making which require model support
- Unintended consequences from misinformed decisions

Model evaluation provides the model development team (developers, intended users, and decision makers) with the level of model corroboration; providing an understanding of how consistent the model is with data.

Model corroboration is defined as the quantitative and qualitative methods for evaluating the degree to which a model corresponds to reality (e.g. measured data). In general, the term corroboration is preferred (rather than **validation**) because it implies a claim of usefulness and not truth (EPA, 2009a).

Model evaluation can provide answers to questions like:

- Does the model reasonably approximate the system?
- Is the model supported by the available data?
- Is the model founded with principles of sound science?
- Does the model perform the specified task?
- What level of uncertainty is attributable to the data or the model?
- Is better data needed for future model applications?

EVALUATION TECHNIQUES

There are many techniques and approaches for model evaluation, as made evident by a review conducted by researchers in the EPA Office of Research and Development. Their efforts (Matott et al., 2009) identified 70 model evaluation tools, classified into seven thematic categories:

- **Data Analysis:** to evaluate or summarize input, response or model output data
- **Identifiability Analysis:** to expose inadequacies in the data or suggest improvements in the model structure
- **Parameter Estimation:** to quantify uncertain model **parameters** using model simulations and available response data
- **Sensitivity Analysis:** to determine which inputs are most significant
- **Uncertainty Analysis:** to quantify output uncertainty by propagating sources of uncertainty through the model
- **Multi-model Analysis:** to evaluate model uncertainty or generate ensemble predictions via consideration of multiple plausible models
- **Bayesian Networks:** to combine prior distributions of uncertainty with general knowledge and site-specific data to yield an updated (posterior) set of distributions

Additional Web Resource:
Further information regarding the model evaluation tools from Matott et al. (2009) can be found at: Model Evaluation

Further Insight:
Evaluating uncertainty in integrated environmental models: A review of concepts and tools. Matott, L. S., J. E. Babendreier and S. T. Purucker 2009. Water Resour. Res. 45: Article Number: W06421.

BEST PRACTICE: A GRADED APPROACH

Model evaluation should be conducted using a **graded approach** that is adequate and appropriate to the model application or decision at hand (EPA, 2009a).

A graded approach recognizes that model evaluation, as an iterative process, cannot be designed in a 'one size fits all approach.' Further, the NRC (2007) recommends that model evaluation be designed to the complexity and impacts (of the model) in addition to consideration of the life-cycle stage of the model and the evaluation history.

For example, a screening model (a type of model designed to provide a "conservative" or risk-averse answer) that is 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.

Ideally, decision makers and modeling staff work together at the onset of new projects to identify the appropriate degree of model evaluation.

Intended Use of Model Results

Increasing Consequence of Model Results

- Regulatory compliance
- Litigation
- Congressional testimony
- Regulatory development
- State Implementation plan attainment
- Verification of model
- Trends monitoring (non-regulatory)
- Proof of principle
- Basic research
- Bench-scale testing

A Graded Approach to Model Evaluation. Model results that have higher consequences our outcomes should be subjected to more rigorous evaluation methods.

A MODEL EVALUATION PLAN

An **evaluation plan** can be a valuable tool for outlining the evaluation exercises that will be carried out during the model life-cycle. The [1]**evaluation plan** should be determined during the earliest stages of model development (i.e. in the required QA project plan) and include all members of the **model development team**.

The overarching goal of model evaluation is to ensure model quality. The EPA defines quality in the [2]**Information Quality Guidelines** (IQGs) (EPA, 2002b). The IQGs apply to all information that EPA disseminates, including models, information from models, and input data.

Additional Web Resource:
Further guidance on model evaluation can be found in another module: Sensitivity and Uncertainty Analyses

(The vertical sliders are on the next page.)

[1]*Vertical Slider #1*

Recommendations of Elements to Include in an Evaluation Plan (NRC, 2007)

- Describe the model and its intended uses
- Describe the relationship of the model to data (for both inputs and corroboration)
- Describe how model performance will be assessed
- Use an outline or diagram that show how the elements and instances of evaluation 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 can be fairly broad and qualitative.
- Identify the responsibilities, accountabilities, and resources needed to ensure implementation of the evaluation plan.

[2]*Vertical Slider #2*

Information Quality Guidelines

Quality has three major components: **integrity**, **utility**, and **objectivity** (EPA, 2002b). In the context of environmental modeling, evaluation aims to ensure the objectivity of information from models by considering their **accuracy**, **bias**, and **reliability**.

Additional Web Resource:

Further guidance on quality assurance for modeling can be found in another module: **QA of Modeling Activities** *(coming soon)*.

CASE STUDY

The Atmospheric Modeling and Analysis Division

(AMAD Website)

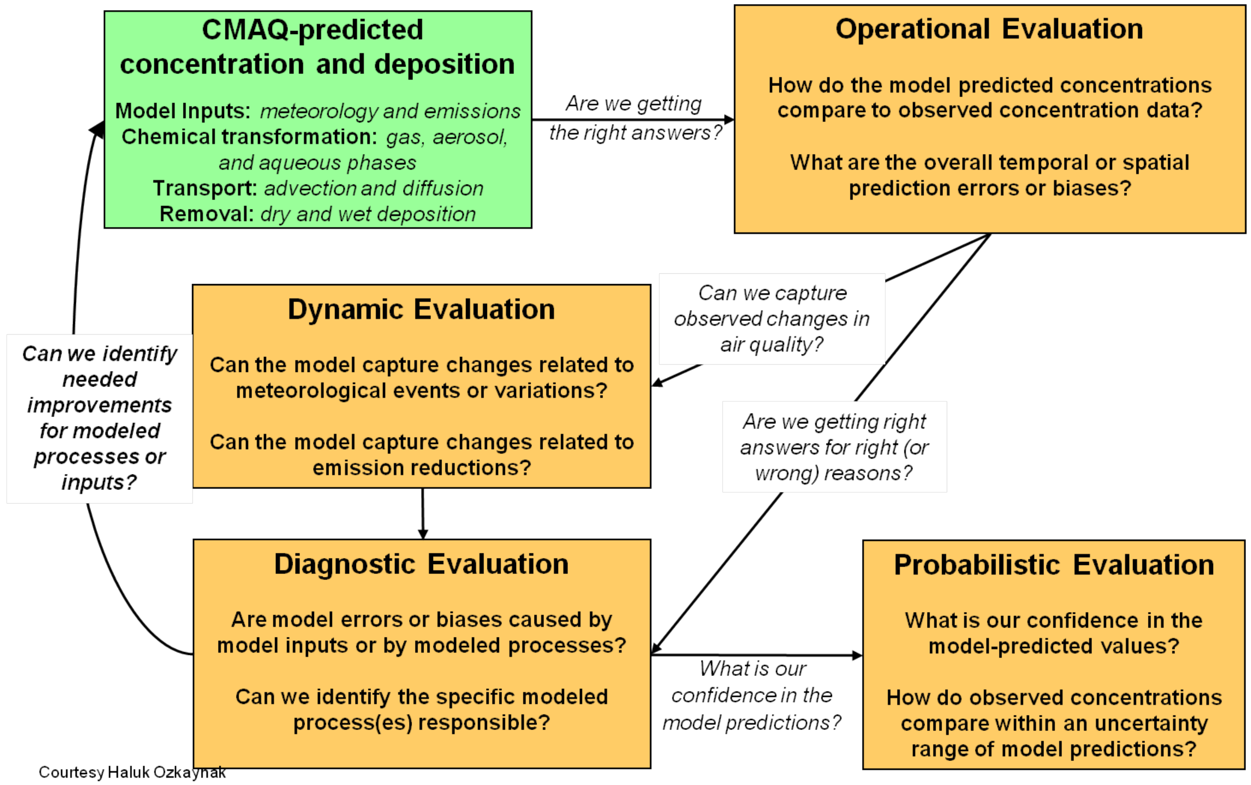
The Atmospheric Modeling and Analysis Division (AMAD) of the Office of Research and Development leads the development and evaluation of predictive atmospheric models on all spatial and temporal scales for assessing changes in air quality and air pollutant exposures. For evaluation of their models, AMAD has developed a

[1]**framework to describe different aspects of model evaluation:**

- [2]**Operational evaluation:** a comparison of model-predicted and routinely measured concentrations of the end-point pollutant(s) of interest in an overall sense.
- Diagnostic evaluation: investigates the atmospheric processes and input drivers that affect model performance to guide model development and improvements needed in emissions and meteorological data.
- Dynamic evaluation: assesses a model's air quality response to changes in meteorology or emissions, which is a principal use of an air quality model for air quality management.
- Probabilistic evaluation: characterizes uncertainty of model predictions for model applications such as predicted concentration changes in response to emission reductions.

(The vertical sliders are on the next two pages.)

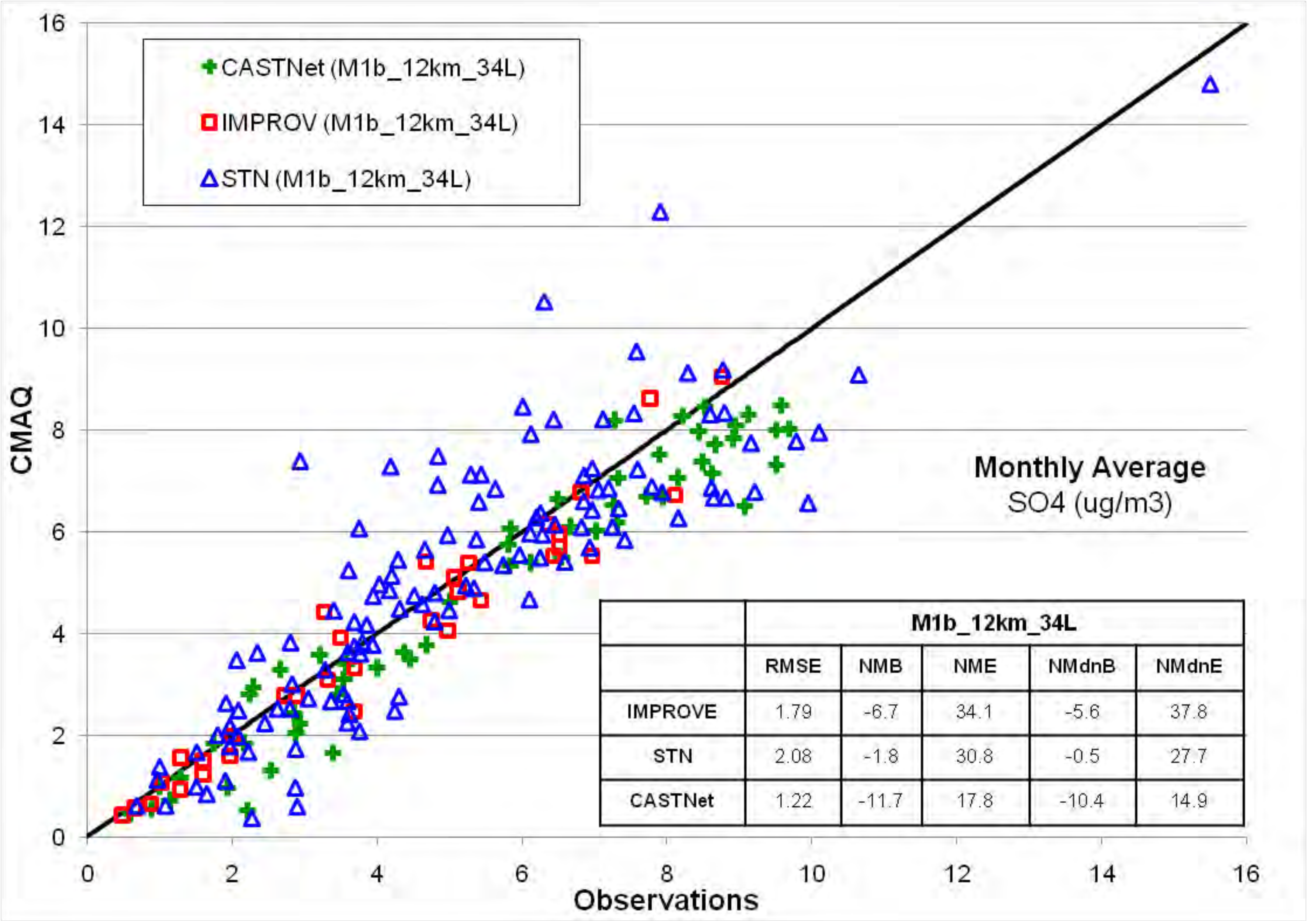
[1]*Vertical Slider #1*



A model evaluation framework example for the CMAQ model. Adapted from the AMAD Website

- The CMAQ model's home page
- Registry of EPA Applications, Models and Databases (READ)

[2]*Vertical Slider #2*



	M1b_12km_34L				
	RMSE	NMB	NME	NMdnB	NMdnE
IMPROVE	1.79	-6.7	34.1	-5.6	37.8
STN	2.08	-1.8	30.8	-0.5	27.7
CASTNet	1.22	-11.7	17.8	-10.4	14.9

An example of **operational evaluation -** a fundamental first phase of any model evaluation. A scatter plot of observed versus CMAQ (an air quality model) predicted sulfate (SO_4) concentrations for August 2006. (Image modified from AMAD)

QUALITY ASSURANCE PLANNING

A well-executed **quality assurance project plan (QAPP)** helps to ensure how model evaluation will be performed and that a model performs the specified task. The objectives and specifications of the model set forth in a quality assurance plan can also be subjected to peer review.

Data quality assessments are an integral component of any QA plan that includes modeling activities. Similar to peer review, data quality assessments evaluate and assure that (EPA, 2002a):

- the data used by model is of high quality
- data uncertainty is minimized
- the model has a foundation of sound scientific principles

The data used to parameterize and corroborate models should be assessed during all relevant stages of a modeling project. These data assessments should be both qualitative and quantitative (i.e. is there enough appropriate data). These assessments ensure that the data sufficiently represent the system being modeled.

Additional Web Resources:

The topic of model documentation (an important component of a complete QA plan) is discussed in other modules as well:

- Best Modeling Practices: Development
- Best Modeling Practices: Application
- *QA of Modeling Activities (Coming Soon)*

Additional information (including guidance documents) can be found at the Agency's website for the Quality System for Environmental Data and Technology.

DRIVERS FOR QA PLANNING

Congress has directed the Office of Management and Budget (OMB) to issue government-wide guidelines that

> *"…provide policy and procedural guidance to Federal agencies for ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by Federal agencies..."*

EPA is dedicated to the collection, generation, and dissemination of high quality information (EPA, 2002b). The EPA states that its quality systems must 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."*

Requirements for QA plans for data collection and modeling activities is one of the EPA's major means to achieve its high quality assurance goals.

Further Insight:
Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by the Environmental Protection Agency. (61 pp, 896 KB, about PDF) 2002. EPA-260R-02-008. Office of Environmental Information. U.S. Environmental Protection Agency. Washington, DC.

QAPP FOR MODELING

The EPA's Quality System for Environmental Data and Technology is in place to manage the quality of its environmental data collection, generation, and use ([1]**including models**).Guidelines provide information about how to document and conduct quality assurance planning for modeling. Specific recommendations include:

- Specifications for developing **assessment criteria**
- Assessments at various stages of the modeling process
- Reports to management as feedback for corrective action
- The process for acceptance, rejection, or qualification of the output for the intended use

A [2]**graded approach** is also practical in the development of quality assurance project plan (QAPP) for modeling (EPA, 2002b). When models are developed or applied, the intended use of the generated results should be known. This information provides guidance for determining the appropriate level of quality assurance.

[1]***Vertical Slider #1***

Additional Web Resource:

Quality assurance planning for modeling is specifically addressed in the **QA of Modeling Activities module** *(coming soon)*.

[2]*Vertical Slider #2*

Increasing Consequence of Model Results

Intended Use of Model Results

- Regulatory compliance
- Litigation
- Congressional testimony

- Regulatory development
- State Implementation plan attainment
- Verification of model

- Trends monitoring (non-regulatory)
- Proof of principle

- Basic research
- Bench-scale testing

Increasing Level of Quality Assurance

Typical QA Issues

- Legal defensibility of data sources
- Compliance with laws and regulatory mandates applicable to data gathering

- Compliance with regulatory guidelines
- Existing data obtained under suitable QA program
- Audits and data reviews

- Use of accepted data-gathering methods
- Use of widely accepted models
- Audits and data reviews

- QA planning and documentation
- Peer review of novel theories and methodology

A Graded Approach: Examples of modeling projects with increasing consequence and the associated level of QA planning.

DATA QUALITY ASSESSMENT

In some instances, the modeling project may utilize data coming from both direct and indirect measurements. A QA plan for data quality should identify:

1. The need and intended use of each type of data or information to be acquired.
2. Requirements on how indirect measurements are to be acquired and used
3. How the data will be identified or acquired, and expected sources of these data.
4. The method of determining the quality of the data.
5. The criteria established for determining the quantity and quality level of data that is acceptable.

⇕[1]Accepted Criteria for Individual Data Values (⇕[2]continued)

A Modeling Caveat
The EPA recommends using the terms ‘precision’ and ‘bias,’ rather than ‘accuracy,’ to convey the information usually associated with accuracy (the closeness of a measured or computed value to its ‘true’ value)

(The vertical sliders are on the next page.)

### [1] *Vertical Slider #1*	### [2] *Vertical Slider #2*
Acceptance Criteria for Individual Data Values (EPA, 2009a) **Representativeness:** • Were the data collected from a population sufficiently similar to the population of interest within the context of model application? **Bias:** • Would any characteristics of the dataset have an unintentional and direct impact the model output? • In probabilistic models, are there adequate data in the upper and lower extremes of the tails to allow for unbiased probabilistic estimates? **Precision:** • How is the uncertainty in the results estimated? • Is the estimate of variability sufficiently small to meet the uncertainty objectives of the modeling project	**Acceptance Criteria for Individual Data Values (EPA, 2009a)** **Qualifiers:** • Have the data met the quality assurances and data quality objectives? • Is the system of qualifying or flagging data adequately documented? **Summarization:** • Is the data summarization process clear and sufficiently consistent with the goals of this project? • Processing and transformation equations should be made available so that the underlying assumptions can be evaluated against the objectives of the project.

PEER REVIEW

The peer review process provides the main mechanism for independent evaluation and review of environmental models used by the EPA. Its purpose is two-fold:

- To evaluate whether the assumptions, methods, and conclusions derived from environmental models are based on sound scientific principles
- To check the scientific appropriateness of a model for informing a specific regulatory decision

Peer review can uncover technical problems, oversights, or unresolved issues in preliminary versions of the model (EPA, 2006b).

Further Insight:
Peer Review Handbook. (190 pp, 1156 KB, about PDF) 2006. EPA/100/B-06/002. Science Policy Council, US Environmental Protection Agency Washington, DC.

BEST PRACTICES

The **peer review** process has an important role in each stage of the model lifecycle. Mechanisms for external peer review could include (EPA, 1994):

1. Using an *ad hoc* panel of scientists
2. Using an established peer review mechanism (e.g. Science Advisory Board, Committee on Environment and Natural Resources Research)
3. Holding a technical workshop

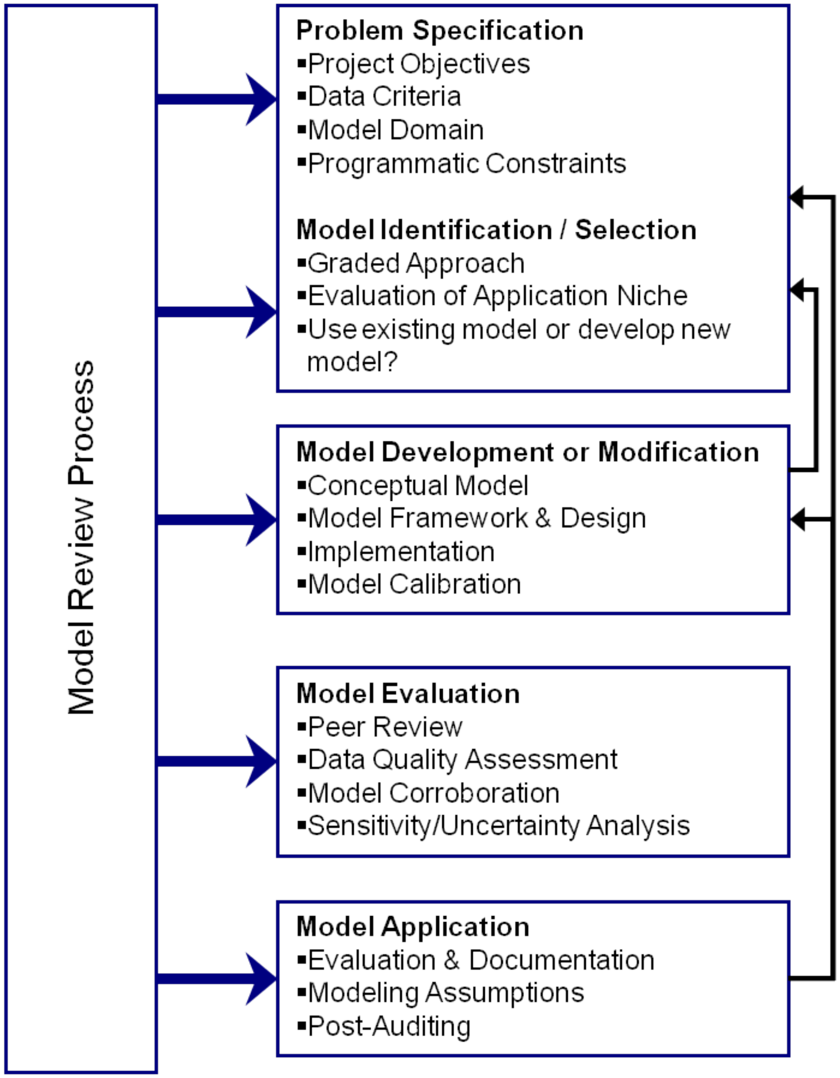
The peer review process should also be well documented in the QA plan. In the earliest stages of model development, the team should identify the expected evaluation events and peer review processes. During any peer review process the review panel would also be presented with charge questions to drive the review.

[1] Peer Review Figure

[2] Peer Review Elements

(The vertical sliders are on the next two pages.)

[1]*Vertical Slider #1*



A detailed diagram of the model life-cycle (EPA, 2009a); including peer review during at every stage of the model life-cycle.

[2] ***Vertical Slider #2***

Critical Elements of the Peer Review Process Individual Data Values (EPA, 1994; 2009a)

- Modeling Purpose/Objectives (context, **application niche**)
- Major Considerations (processes, scales, etc.)
- Theoretical Basis of the Model (shortcomings, algorithms)
- **Parameter** Estimation (methods, boundary conditions)
- Data Quality/Quantity (data adequacy, selection process)
- Key Assumptions (basis of, sensitivity to)
- Model Performance Measures (criteria, relative performance)
- Model Documentation (comprehensive)
- Retrospective (were intentions realized, robustness, uncertainty quantification)

EXAMPLE CHARGE QUESTIONS

Examples of charge questions for a peer review process are provided to the right. For additional information on these models please see the citations and links provided below.

[1] Second Generation Model (SGM)

- Registry of EPA Applications, Models and Databases (READ)
- SGM Homepage
- Science Advisory Board Report on SGM (46 pp, 403 KB, about PDF) (EPA, 2007a)

[2] EPI Suite™ Model

- Registry of EPA Applications, Models and Databases (READ)
- EPI Suite™ Homepage
- Science Advisory Board Report on EPI Suite™ (60 pp, 472 KB, about PDF) (EPA, 2007b)

(The vertical sliders are on the next page.)

[1] *Vertical Slider #1*

The Science Advisory Board Review of EPA's Second Generation Model (EPA, 2007a)

The Second Generation Model (SGM) is a computable general equilibrium model designed specifically to analyze issues related to energy, economy, and greenhouse gas emissions. These questions represent general and overarching questions charged to the peer review group. *See EPA (2007a) for the full report and more specific questions.*

- Is the SGM appropriate and useful for answering questions on the economic effects of climate policies?
- Are the model's structure and fundamental assumptions reasonable and consistent with economic theory?
- Are the **parameter** values employed in the model (e.g., elasticities of substitution and of demand, price and income) within the range of values in the literature?
- Are the model's parameterizations logical?
- Are the model's projections of future energy use and efficiency reasonable, given fundamental physical constraints and rates of technological change?
- In what areas is the model in need of further development?

[2] *Vertical Slider #2*

The Science Advisory Board Review of the Estimation Programs Interface Suite (EPI Suite™) (EPA, 2007b)

The EPI Suite™ is suite of physical/chemical property and environmental fate estimation models developed by the EPA's Office of Pollution Prevention Toxics and Syracuse Research Corporation (SRC). These questions represent general and overarching questions charged to the peer review group. *See EPA (2007b) for the full report and more specific questions.*

- Are there additional properties that should be included in upgrades to the model for its various specified uses?
- Are there places where EPI Suite™'s user guide (and other program documentation) does not clearly explain EPI's design and use? How can these be improved?
- Are there aspects of the user interface that need to be corrected, redesigned, or otherwise improved?
- Are there other features that could enhance convenience and overall utility for users?
- Are property estimates expressed in correct/appropriate units?
- Is adequate information on accuracy/validation conveyed to the user by the program documentation and/or the program itself?

DEFINITIONS

Assessing the degree to which a model represents a defined system is not simply a matter of comparing model results and empirical data. During the Development Stage, the modeling team determined an acceptable degree of total **uncertainty** within the context of specific model applications. Ideally, this determination should be informed by decision makers and stake holders; and described in a quality assurance plan.

Model corroboration assesses the degree to which a model corresponds to reality, using both quantitative and qualitative methods. The modelers may use a graded approach, as mentioned earlier, to determine the rigor of these assessments which should be appropriately defined for each model application.

Qualitative methods, like **expert elicitation**, can provide the development team with beliefs about a system's behavior in a data-poor situation. Utilizing the expert knowledge available, qualitative corroboration is achieved through *consensus* and *consistency* (EPA, 2009a).

Additional Web Resource:
Further information regarding the Development Stage can be found in the Best Modeling Practices: Development module.

CLARIFICATION ON MODEL EVALUATION, VALIDATION AND VERIFICATION

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 (EPA, 2009a).

Validated models are those that have been shown to successfully perform a specific task (model application scenario) of site-specific field data. Model validation is essentially problem specific since there are few 'universal' models (Beck et al., 1994). The *Guidance Document* (EPA, 2009a) focuses on the processes and techniques for model evaluation rather than model validation or invalidation. **For a case study – the validation of AQUATOX v1.66 for Lake Onondaga, NY EPA (2000) – please see the "Case Study" subtab.**

Verification is another term commonly applied to the evaluation process. However, model verification typically refers to model code. Verification is an assessment of the algorithms and numerical techniques used in the computer code to confirm that they work correctly and represent the conceptual model accurately – a process typically applied during the **Development Stage** (EPA, 2009a).

Further Insight:

Guidance on the Development, Evaluation, and Application of Environmental Models. (99 pp, 1717 KB, about PDF) 2009. EPA/100/K-09/003. Washington, DC. Office of the Science Advisor, US Environmental Protection Agency.

Models in Environmental Regulatory Decision Making. 2007. National Research Council. Washington, DC. National Academies Press.

Standard Guide for Statistical Evaluation of Atmospheric Dispersion Model Performance (D 6589). ASTM. 2000. Available: http://www.astm.org

Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. 1994. Oreskes, N., K. Shrader-Frechette and K. Belitz. Science 263 (5147): 641-646.

QUANTITATIVE METHODS

Quantitative measures assess the **robustness** of a model – the capacity of a model to perform equally well across the full range of environmental conditions for which it was designed (EPA, 2009a). These assessments rely upon statistical measures to calculate various measures of fit between modeled results and measured data.

Model performance measures assess how close modeled results are to the measured data through deviances. Each method (Janssen and Heuberger, 1995) has strengths and weaknesses that should be considered when choosing an assessment measure. For example, modeling efficiency (ME) is a dimensionless statistic which directly relates model predictions to observed data and root mean square error (RMSE) is a method which is sensitive to outliers, but can accurately describe relationships between modeled data and noise-free (measured) data.

(Formulas are on the next page.)

Deviance Measures Between Modeled (P) and Observed/Measured (O) Values
From Janssen and Heuberger (1995) n = number of observations.

Average Error:

$$AE = \frac{\sum_{i=1}^{n} (P_i - O_i)}{n}$$

Mean Absolute Error:

$$MAE = \frac{\sum_{i=1}^{n} |(P_i - O_i)|}{n}$$

Modeling Efficiency:

$$\frac{\sum_{i=1}^{N} (O_i - \overline{O})^2 - \sum_{i=1}^{N} (P_i - \overline{P})^2}{\sum_{i=1}^{N} (O_i - \overline{O})^2}$$

Root Mean Square Error:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (P_i - O_i)^2}{n}}$$

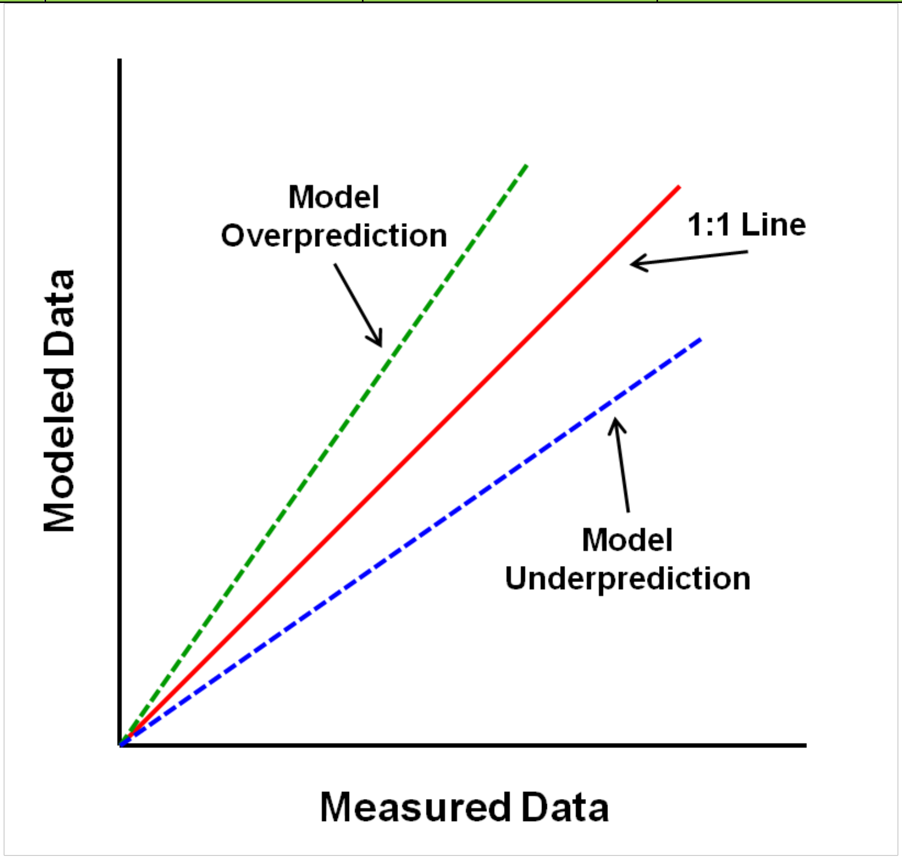
Mean Square Error:

$$MSE = \frac{\sum_{i=1}^{n} (P_i - O_i)^2}{n}$$

GRAPHICAL METHODS

Simple plots between modeled and measured data can reveal qualitative assessments of model performance (i.e. time series, scatter plots, quantile-quantile (Q-Q) plots, spatial concentration plots, etc.).

Plotting modeled data against measured data is a simple way to assess model performance, as depicted in the figure to the right. The 1:1 line represents a model that is accurately predicting measured data.



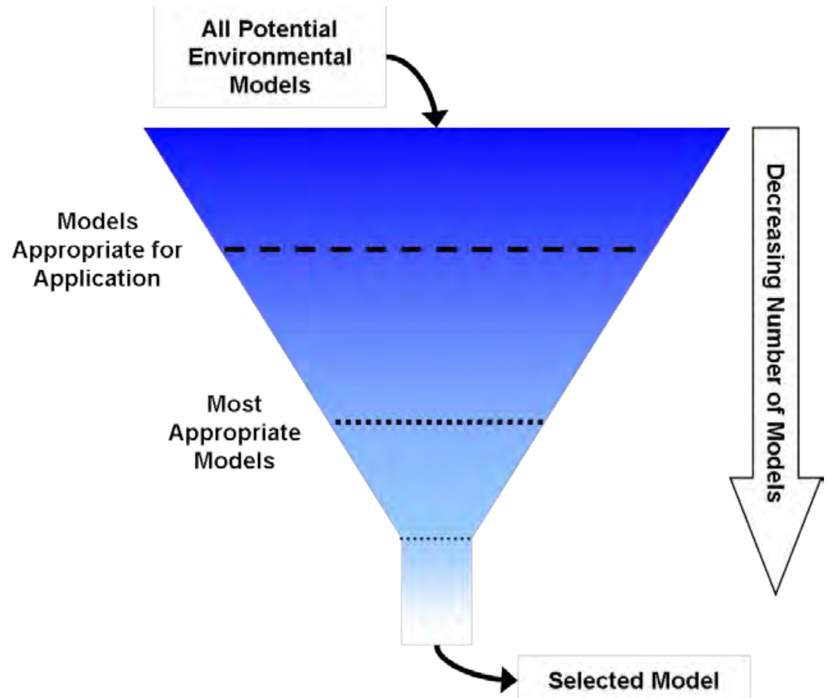
Comparisons between Measured and Modeled Data along the '1:1 Line.' This simple comparison can be useful in early stages of model evaluation as a qualitative way to assess model performance.

MODEL SELECTION

In many scenarios, there may be a number of models suited for a particular application and the project team uses both quantitative and qualitative methods of model evaluation to select the best model for their modeling application.

Ranking models on the basis of their statistical performance against measured data can aid in the process of model selection. When quantitative measures of models performance do not distinguish one model from another, model selection can shift to a more qualitative nature. Past use, public familiarity, cost or resource requirements, and availability can all be useful metrics to help determine the most suitable model.

Additional Web Resource:
Further discussion of model selection can be found in the [Best Modeling Practices: Application] module.



During a model selection process, all potential environmental models are examined to determine to the most appropriate models and the selected model(s).

CASE STUDY: VALIDATION

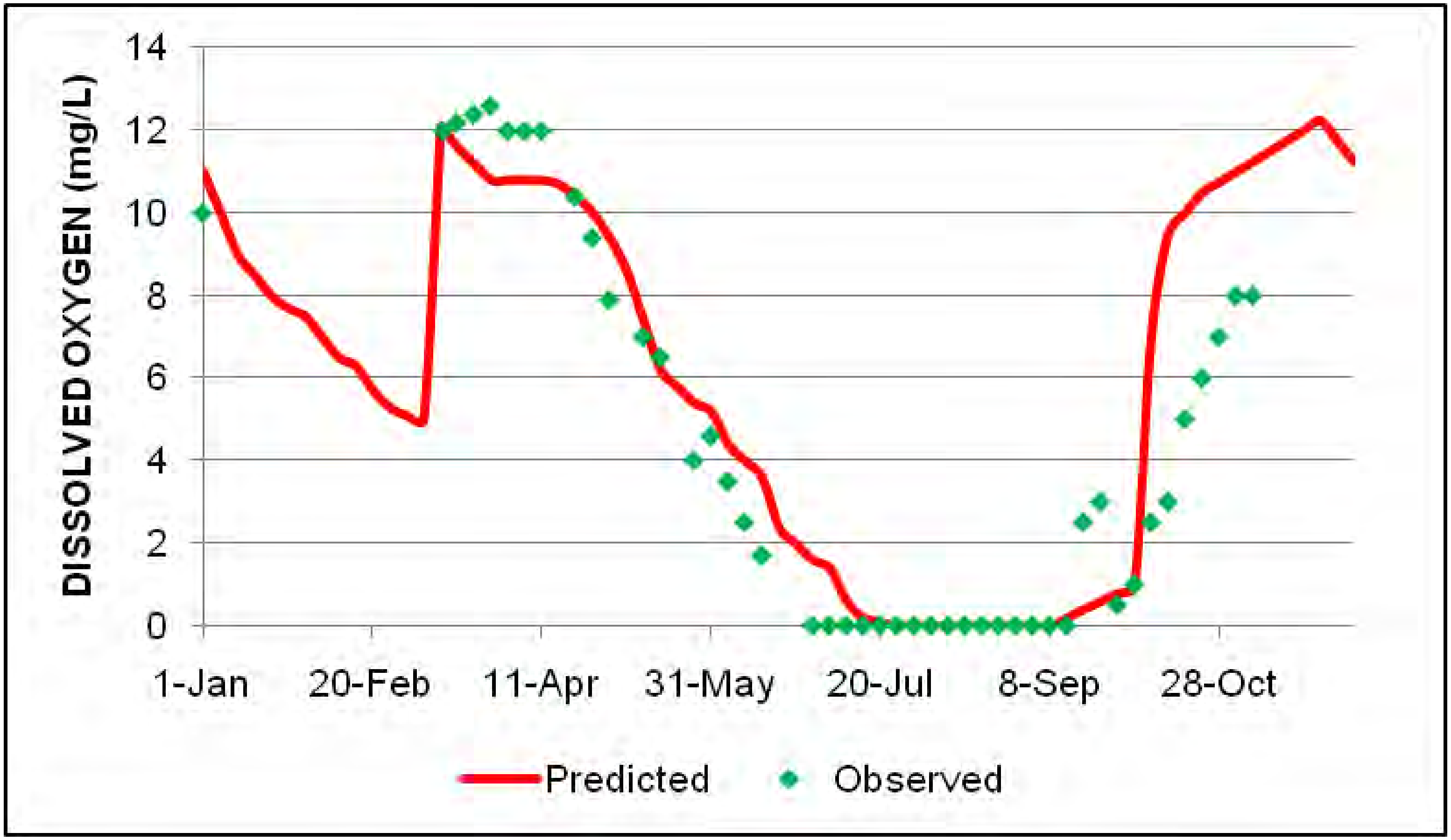
Application of AQUATOX v1.66 to Lake Onondaga, NY

- Registry of EPA Applications, Models and Databases (READ)
- AQUATOX homepage
- Validation reports (EPA, 2000)

The aquatic ecosystem model AQUATOX is one of the few general ecological risk models that represents the combined environmental fate of various pollutants and their effects on the ecosystem, including fish, invertebrates, and aquatic plants (EPA, 2000).

In this example, AQUATOX was validated for use at Lake Onondaga, NY. The validation report (EPA, 2000) highlights multiple levels of validation with calibrated versions of the model.

(Figure is on the next page.)



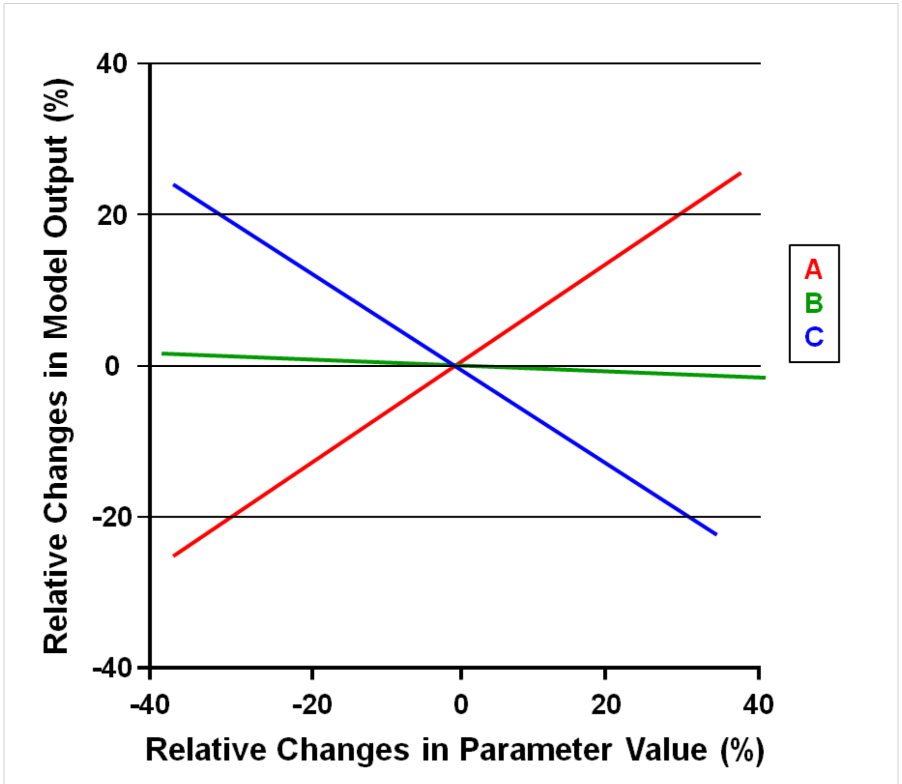
Dissolved oxygen concentrations in the Lake Onondoaga hypolimnion in 1990. AQUATOX predictions indicate anoxic conditions in the middle of summer and the episode is remarkably close to the observed conditions (EPA, 2000). Image adapted from EPA (2000).

SENSITIVITY ANALYSIS

The purpose of a sensitivity analysis (SA) can be two-fold. First, SA computes the effect of changes in model inputs on the outputs. Second, SA can be used to study how uncertainty in a model output can be systematically apportioned to different sources of uncertainty in the model input.

Sensitivity analysis is defined as the computation of the effect of changes in input values or assumptions (including boundaries and model functional form) on the outputs. In other terms, *how sensitive are the results to changes in the inputs, parameters, or model assumptions.*

A non-intensive sensitivity analysis can first be applied to identify the most sensitive inputs. By discovering the 'relative sensitivity' of model **parameters**, the model development team is then aware of the relative importance of parameters in the model and can select a subset of the inputs for more rigorous sensitivity analyses (EPA, 2009a). This also ensures that a single parameter is not overly influencing the results.



A spider diagram used to compare relative changes in model output to relative changes in the parameter values can reveal sensitivities for each parameter (Addiscott, 1993). In this example, the effects of changing parameters A, B, and C are compared to relative changes in model output. The legs represent the extent and direction of the effects of changing parameter values.

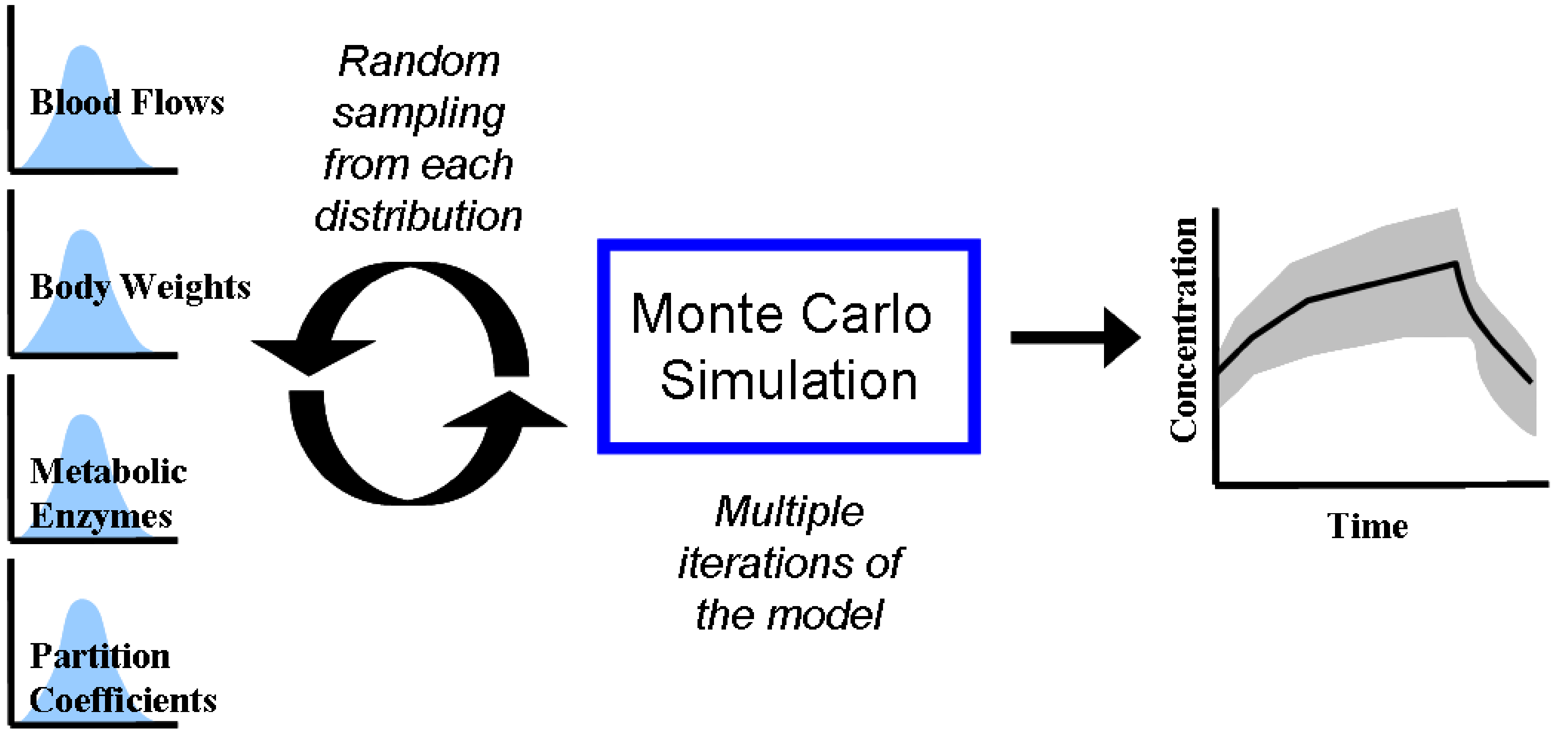
SENSITIVITY ANALYSIS METHODS

There are many methods for sensitivity analysis (SA), a few of which were highlighted in the *Guidance on the Development, Evaluation, and Application of Environmental Models* (EPA, 2009a). The chosen method is dependent upon assumptions made and the amount of information needed from the analysis. Those methods are categorized into:

- Screening Tools
- Morris's One-at-a-Time
- Differential Analysis Methods
- Methods Based on Sampling
- Variance Based Methods

For many of the methods it is important to consider the geometry of the response plane and potential interactions among parameters and/or input variables. Depending on underlying assumptions of the model, it may be best to start SA with simple methods to initially identify the most sensitive inputs and then apply more intensive methods to those inputs.

(Figure is on the next page.)



Physiologically based pharmacokinetic (PBPK) models represent an important class of dosimetry models that are useful for predicting internal dose at target organs for risk assessment applications (EPA 2006a). This figure is an example of the Monte Carlo simulation method for Sensitivity Analysis. The distribution of internal concentration versus time (*output*) is simulated by repeatedly (often as many as 10,000 iterations) sampling input values based on the distributions of individual parameters (blood flow rate, body weight, metabolic enzymes, partition coefficients, etc.) in a population. Adapted from EPA (2006a).

FURTHER INSIGHT

Sensitivity analysis (SA) is an important component of Model Evaluation. When combined with uncertainty analysis (UA) the contribution of input parameters to total uncertainty can be revealed. Further, knowing which inputs to focus further analyses on saves the research team valuable time.

There are many methods for SA, each coming with a set of caveats and features that can be used to select the best SA for a specific model application.

Additional Web Resource:

Specific methodologies are explored in the Sensitivity and Uncertainty Analyses module.

Further Insight:

Guiding Principles for Monte Carlo Analysis. (39 pp, 170 KB, about PDF) 1997. EPA-630-R-97-001. Risk Assessment Forum. U.S. Environmental Protection Agency. Washington, DC.

Multimedia, Multipathway, and Multireceptor Risk Assessment (3MRA) Modeling System Volume IV: Evaluating Uncertainty and Sensitivity. 2003. EPA530-D-03-001d. Office of Research and Development. US Environmental Protection Agency. Athens, GA.

Guidance on the Development, Evaluation, and Application of Environmental Models. (99 pp, 1717 KB, about PDF) 2009. EPA/100/K-09/003. Washington, DC. Office of the Science Advisor, US Environmental Protection Agency.

UNCERTAINTY ANALYSIS

Uncertainties (i.e. a lack of knowledge) are present and inherent throughout the modeling process. However, models can continue to be valuable tools for informing decisions through proper quantification and communication of the associated uncertainties (EPA, 2009a).

Uncertainty analysis (UA) investigates the effects of lack of knowledge or potential errors on model output. When UA is conducted in combination with sensitivity analysis; the model user can become more informed about the confidence that can be placed in model results (EPA, 2009a).

Model Uncertainty (EPA, 2009a)

- **Application niche uncertainty** – uncertainty attributed to the appropriateness of a model for use under a specific set of conditions (i.e. a model application scenario)
- **Structure/framework uncertainty** –incomplete knowledge about factors that control the behavior of the system being modeled; limitations in spatial or temporal resolution; and simplifications of the system.
- **Input/data uncertainty** – resulting from data measurement errors; inconsistencies between measured values and those used by the model; also includes parameter value uncertainty

UNCERTAINTY AND VARIABILITY

Uncertainty represents lack of knowledge about something that is true. It is a general term that is often applied in a number of contexts. In environmental modeling, it may describe a lack of knowledge about models, **parameters**, constants, data, or the underlying assumptions.

The nature of uncertainty can be described as (Walker et al., 2003; Pascual 2005; EPA, 2009b):

- **Stochastic uncertainty** – resulting from errors in empirical measurements or from the world's inherent stochasticity
- **Epistemic uncertainty** – uncertainty from imperfect knowledge
- **Technical uncertainty** – uncertainty associated with calculation errors, numerical approximations, and errors in the model algorithms

Variability vs. Uncertainty

Variability is a special instance of uncertainty – often called **data uncertainty**. Variability of environmental data is a product of the inherent randomness and heterogeneity of the environment.

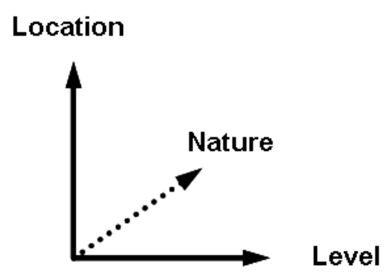
Variability can be better characterized, but hard to reduce, with further study.

Separating variability and uncertainty is necessary to provide greater accountability and transparency (EPA, 1997).

UNCERTAINTY MATRIX

Many of the modeling uncertainties identified have limited room for improvement. However, there are methods to resolving model uncertainty; even some of the qualitative uncertainties can be addressed.

Uncertainty analysis begins with characterizing the associated modeling uncertainties; often accomplished using a framework or uncertainty matrix (i.e. Walker et al., 2003; Refsgaard et al., 2007). Walker et al. (2003) discuss uncertainty as a 3-Dimensional relationship:



Location: Where the uncertainty manifests itself within the model complex (application niche, framework, or input uncertainty).
Level: The degree of uncertainty along the spectrum between deterministic knowledge and total ignorance.
Nature: whether the uncertainty comes from epistemic uncertainty, or the inherent variability of the phenomena being described.

(Table is on the next page.)

Location		Level			Nature	
		Statistical Uncertainty	Scenario Uncertainty	Recognized Uncertainty	Epistemic Uncertainty	Stochastic Uncertainty
Context	Natural, technological, economic, social, and political representation					
Model	Model structure					
	Technical model					
Inputs	Driving forces					
	Systems Data					
Parameters						
Model Outcomes						

An example of an uncertainty matrix (Walker et al., 2003). Frameworks like this assist the model development team in identifying the sources of uncertainty and where efforts for resolving uncertainty may be best applied.

UNCERTAINTY ANALYSIS PRIORITIES

Reducing **application niche uncertainty** should be a first priority during a modeling exercise (EPA, 2009a). The **application niche** ❓ of a model should be used to determine whether the use of a given model is appropriate for the situation. Other priorities include:

- Mapping the **model attributes**❓ to the problem statement
- Confirming the degree of certainty needed from model outputs
- Determining the amount of reliable data available or the resources available to collect more
- The quality of the scientific foundations of the model
- The technical competence of the model development / application team

FURTHER INSIGHT

The EPA has produced a number of resources and guidance documents on uncertainty analysis that are specific to a variety of environmental modeling fields. A few of those resources are identified to the right.

Additional Web Resource:
These methods are discussed further in the Sensitivity and Uncertainty Analyses module.

Further Insight:
Uncertainty and Variability in Physiologically Based Pharmacokinetic Models: Key Issues and Case Studies (10 pp, 69 KB, about PDF)2008. EPA/600/R-08/090 Office of Research and Development. US Environmental Protection Agency. Washington, DC.

Guidance on the Development, Evaluation, and Application of Environmental Models. (99 pp, 1717 KB, about PDF) 2009. EPA/100/K-09/003. Office of the Science Advisor. US Environmental Protection Agency Washington, DC.

Using Probabilistic Methods to Enhance the Role of Risk Analysis in Decision-Making With Case Study Examples DRAFT (92 pp, 712 KB, about PDF) 2009. EPA/100/R-09/001. Risk Assessment Forum. US Environmental Protection Agency. Washington, DC.

SUMMARY

The purpose of this module is to explore the topic of model evaluation and identify the 'best modeling practices' and strategies for the Evaluation Stage of the model life-cycle. In summary:

- Model evaluation is a process of many activities that should include:
    - Peer review
    - Quality Assurance (QA) project planning
    - Model corroboration
    - Sensitivity analysis
    - Uncertainty analysis
- QA project planning promotes model transparency.
- The peer review process provides the main mechanism for independent evaluation and review of environmental models used by the EPA.
- There are many techniques and approaches for model corroboration. An appropriate method should be determined at the beginning of the model life-cycle.
- When practiced together, sensitivity and uncertainty analyses can used to study how uncertainty in a model output can be systematically apportioned to different sources of uncertainty in the model input.
- Model evaluation should be conducted using a graded approach that is adequate and appropriate to the objectives of the modeling exercise.

Additional Web Resources:

- SuperMUSE Website: Ecosystems Research Division's Supercomputer for Model Uncertainty and Sensitivity Evaluation (SuperMUSE) is a key to enhancing quality assurance in environmental models and applications.
- Model Evaluation Tools: A compilation of nearly 70 model evaluation tools.
- Sensitivity and Uncertainty Analyses Module

YOU HAVE REACHED THE END OF THE BEST MODELING PRACTICES: EVALUATION MODULE.

REFERENCES

Addiscott, T. M. 1993. Simulation modelling and soil behaviour. Geoderma 60(1-4): 15-40.

Beck, B., L. Mulkey and T. Barnwell. 1994. Model Validation for Exposure Assessments. DRAFT. Athens, GA: US Environmental Protection Agency.

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*. New York. Plenum Press

EPA (U.S. Environmental Protection Agency). 1994. Report of the Agency Task Force on Environmental Regulatory Modeling (PDF). (86 pp, 3.4 MB, about PDF) EPA 500-R-94-001. Solid Waste and Emergency Response.

EPA (U.S. Environmental Protection Agency). 1997. Guiding Principles for Monte Carlo Analysis (PDF). (39 pp, 170 KB, about PDF) EPA-630-R-97-001. Washington, DC. Risk Assessment Forum.

EPA (US Environmental Protection Agency). 2000. AQUATOX For Windows: A Modular Fate and Effects Model for Aquatic Ecosystems: Release 1 Volume 3: Model Validation Reports. (69 pp, 3.4 MB, about PDF) EPA-823-R-00-008. Washington, DC. Office of Water.

EPA (U.S. Environmental Protection Agency). 2002a. Guidance on Environmental Data Verification and Data Validation EPA QA/G-8 (PDF). (96 pp, 387 KB, about PDF) EPA-240-R-02-004. Washington, DC. Office of Environmental Information.

EPA (U.S. Environmental Protection Agency). 2002b. Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by the Environmental Protection Agency (PDF). (61 pp, 896 KB, about PDF) EPA-260-R-02-008. Washington, DC. Office of Environmental Information.

EPA (US Environmental Protection Agency). 2003. Multimedia, Multipathway, and Multireceptor Risk Assessment (3MRA) Modeling System Volume IV: Evaluating Uncertainty and Sensitivity. EPA530-D-03-001d. Athens, GA. Office of Research and Development.

EPA (US Environmental Protection Agency). 2006a. Approaches for the Application of Physiologically Based Pharmacokinetic (PBPK) Models and Supporting Data in Risk Assessment (PDF). (123 pp, 725 KB, about PDF) EPA/600/R-05/043F. Washington, DC. Office of Research and Development.

EPA (US Environmental Protection Agency). 2006b. Peer Review Handbook (PDF). (190 pp, 1.1 MB, about PDF) EPA/100/B-06/002. Washington, DC. Science Policy Council.

EPA (US Environmental Protection Agency). 2007a. SAB Advisory on EPA's Second Generation Model (PDF). (46 pp, 403 KB, about PDF) EPA-SAB-07-006. Washington, DC. Science Advisory Board

REFERENCES (CONTINUED)

EPA (US Environmental Protection Agency). 2007b. Science Advisory Board (SAB) Review of the Estimation Programs Interface Suite (EPI SuiteTM) (PDF). (60 pp, 472 KB, about PDF) EPA-SAB-07-011. Washington, DC. Science Advisory Board

EPA (US Environmental Protection Agency). 2009a. Guidance on the Development, Evaluation, and Application of Environmental Models (PDF). (99 pp, 1.7 MB, About PDF). EPA/100/K-09/003. Washington, DC. Office of the Science Advisor.

EPA (US Environmental Protection Agency). 2009b. Using Probabilistic Methods to Enhance the Role of Risk Analysis in Decision-Making With Case Study Examples. *DRAFT* (PDF). (92 pp, 722K, About PDF) EPA/100/R-09/001 Washington, DC. Risk Assessment Forum.

Hanna, S. R. 1988. Air quality model evaluation and uncertainty. Journal of the Air Pollution Control Association 38(4): 406-412.

Jakeman, A. J., R. A. Letcher and J. P. Norton 2006. Ten iterative steps in development and evaluation of environmental models. Environmental Modeling & Software 21(5): 602-614.

Janssen, P. H. M. and P. S. C. Heuberger 1995. Calibration of process-oriented models. Ecol. Model. 83(1-2): 55-66.

Matott, L. S., J. E. Babendreier and S. T. Purucker 2009. Evaluating uncertainty in integrated environmental models: A review of concepts and tools. Water Resour. Res. 45: Article Number: W06421.

NRC (National Research Council) 2007. *Models in Environmental Regulatory Decision Making*. Washington, DC. National Academies Press.

Pascual, P. 2005. Wresting Environmental Decisions From an Uncertain World. Environmental Law Reporter 35: 10539-10549

Refsgaard, J. C., J. P. van der Sluijs, A. L. Højberg and P. A. Vanrolleghem 2007. Uncertainty in the environmental modeling process - A framework and guidance. Environ. Model. Software 22(11): 1543-1556.

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

Sunderland, E. 2008. Addressing Model Uncertainty & Best Practices for Model Evaluation. Presentation at Region 1 Regional Science Council Models Training Course. October 23, 2008.

Walker, W. E., P. Harremoës, J. Rotmans, J. P. van der Sluijs, M. B. A. van Asselt, P. Janssen and M. P. Krayer von Krauss 2003. Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. Integrated Assessment 4(1): 5-17.

GLOSSARY

Accuracy: The closeness of a measured or computed value to its “true” value, where the “true” value is obtained with perfect information.

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.

Bias: Systematic deviation between a measured (i.e., observed) or computed value and its “true” value.

Integrity: The protection of information from unauthorized access or revision to ensure that it is not compromised through corruption or falsification.

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.

Model Attributes: The processes (chemical, biological, physical); variables; scale; and outputs described or contained within the model.

Model Development Team: Comprised of model developers, users (those who generate results and those who use the results), and decision makers; also referred to as the project team.

Objectivity: Determines 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.

Parameter: 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.

Precision: the quality of being reproducible in amount or performance.

Reliability: A function of the performance record of a model and its conformance to best available, practicable science.

Representativeness: the measure of the degree to which data accurately and precisely represent a characteristic of a population, parameter variations at a sampling point, a process condition, or an environmental condition.

Sensitivity Analysis: The computation of the effect of changes in input values or assumptions (including boundaries and model functional form) on the outputs. The study of how uncertainty in a model output can be systematically apportioned to different sources of uncertainty in the model input.

Uncertainty: Describes a lack of knowledge about models, parameters, constants, data, and beliefs.

Uncertainty Analysis: Investigates the effects of lack of knowledge or potential errors on the model (e.g, the “uncertainty” associated with parameter values or the model framework) 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.

Utility: The usefulness of the information to the intended users.

Validation: Validated models are those that have been shown to successfully perform a specific task (model application scenario) of site-specific data.