Environmental Modeling 101

Welcome to CREM's **Environmental Modeling 101** 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.

ENVIRONMENTAL MODELING 101

The U.S. Environmental Protection Agency (EPA) uses a variety of models to inform decisions that support its mission of protecting human health and safeguarding the natural environment — air, water, and land — upon which life depends.

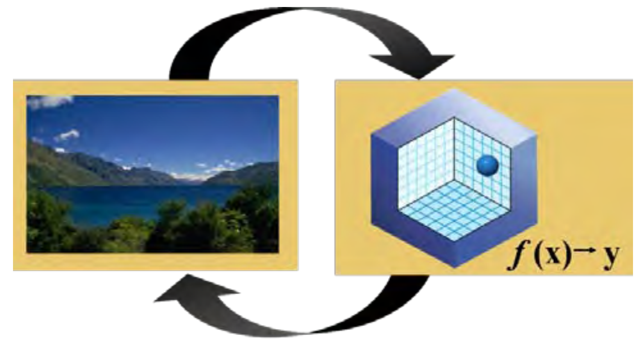
This module has four main objectives:

1. Provides a basic introduction to environmental modeling
2. The definition and types of environmental models
3. How and why models are used in environmental sciences
4. The model "life-cycle"

⬍ [1]**Figure**

⬍ [2]**What is a model?**

[1]Vertical Slider #1



Models are representations of the environment that can be used to inform regulation or management decisions.

[2]Vertical Slider #2

What is a model?

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

"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."

DEFINITION

The term **model** can be an ambiguous word used to describe an 'abstraction (or parameterization) of reality.' Models can take on many forms, the most common and relevant forms are computational and conceptual models.

In a broader sense, there can be many kinds of models (EPA, 2009a):

- [1]**Computational models**
  - **Analytical models** are special computational models that can be solved mathematically in terms of analytical functions.
- [2]**Conceptual models**
- Physical models*
- [3]**Analogous models***

**While the last two types of models are not conventional models, the statistical models used to extrapolate from these abstractions to the 'real' system are. They are included here to distinguish among the types of models.*

[1]Vertical Slider #1

The Tier 1 Rice Model – A computational model

$$C_w = \frac{m_{ai}}{0.00105 + 0.00013K_d}$$

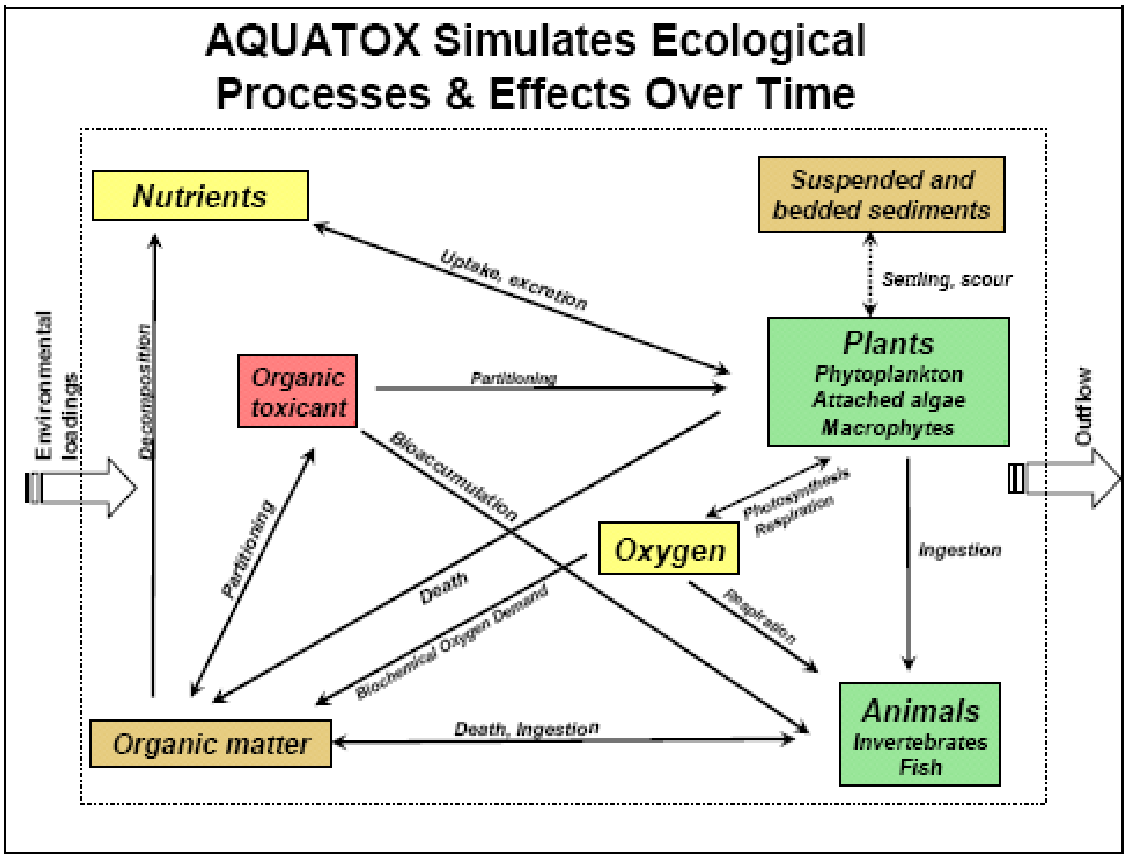
Tier 1 Rice Model Website

[3]Vertical slider #3

A mouse can serve as an analogous model of human physiology.

[2] *Vertical slider #2*

Conceptual model of the AQUATOX model



[Diagram courtesy of the AQUATOX website]()

WHY ARE MODELS USED?

Models have a long history of helping to explain scientific phenomena and predict outcomes and behavior in settings where empirical observations are limited or not available (EPA, 2009a).

Models are based on simplifying assumptions of environmental processes and cannot completely replicate the inherent complexity of the entire environmental system. Despite these limitations, models are essential for a variety of purposes; described in two broad categories:

- To **diagnose** (i.e., assess what happened) and examine causes and precursor conditions (i.e., why it happened) of events that have taken place
- To **forecast** outcomes and future events (i.e., what will happen).

The NRC (2007) describes 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."*

Models can be used to inform a variety of activities including:

- Research
- Toxicity screening
- Policy analysis
- National regulatory decision making
- Implementation applications

MODEL STRUCTURE

In any modeling exercise, the **system** of interest should be defined. This definition is not only used to identify the boundaries of the model, but also serves to define *how* the model can be applied and to *which* systems/situations.

Model developers should answer the following questions:

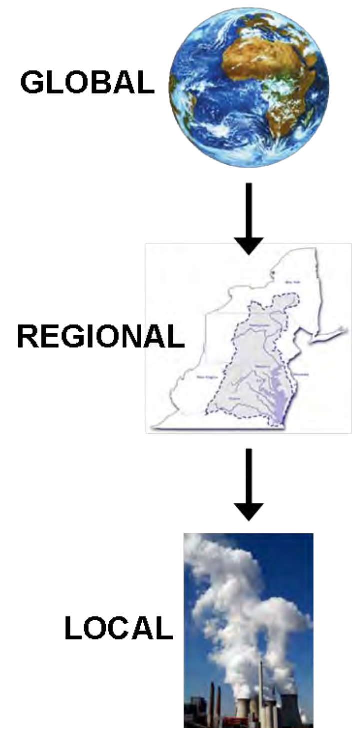
1. What processes is the model attempting to reproduce and include?
2. At what time scale are the included processes occurring?
3. At what spatial scale are the included processes occurring?

Therefore, model structure can be described two ways:

1. Included Processes (chemical, physical, or biological)
2. Scope / Scale (time or space)

A Modeling Caveat

Models are typically (and should be) developed for a well defined system and a set of conditions under which the use of the model is scientifically defensible – the **application niche**. The identification of application niche is a key step during model development and helps guide future application of the model.



Examples of decreasing scale for generic air quality models.

TYPES OF COMPUTATIONAL MODELS

The remainder of this module will focus on computational models. The types of computational models are determined by the available data, the intended use, and the interpretation of model generated results. However, the types of models are not mutually exclusive (see Summary Table slide).

- [1] **Empirical vs. Mechanistic models**
- [2] **Deterministic vs. Probabilistic models**
- [3] **Dynamic vs. Static models**
- [4] **Generic equations by model type**
- [5] **Other relevant modeling terms**

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

Empirical models – include very little information on the underlying mechanisms and rely upon the observed relationships among experimental data. These can be thought of as ‘best-fit’ models whose **parameters** may or may not have real-world interpretation.

Mechanistic models explicitly include the mechanisms or processes between the **state variables**; unlike empirical models. The parameters in mechanistic models should be supported by data and have real-world interpretations (EPA, 2009b).

A Modeling Caveat

When data quality is otherwise equivalent, extrapolation from mechanistic models (e.g., biologically based dose-response models) often carries higher confidence than extrapolation using empirical models (EPA, 2009b).

[2] *Vertical Slider #2*	[3] *Vertical Slider #3*
Deterministic models – provide a solution for the state variable(s) rather than a set of probabilistic outcomes. 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, the boundary conditions, or initial conditions (EPA, 2009a). Therefore, repeated simulations under constant conditions will result in consistent results. **Probabilistic models** – utilize the entire range of input data to develop a probability distribution of model output (i.e. exposure or risk) rather than a single point value. Probabilistic models are sometimes referred to as **statistical or stochastic models**. Probabilistic models can be used to evaluate the impact of variability and uncertainty in the various input parameters, such as environmental exposure levels, fate and transport processes, etc.	**Dynamic models** – make predictions about the way a system changes with time or space. Solutions are obtained by taking incremental steps through the model domain. For most situations, where a differential equation is being approximated, the simulation model will use a finite time step (or spatial step) to estimate changes in state variables over time (or space). **Static models** make predictions about the way a system changes as the value of an independent variable changes.

[4] Vertical Slider #4

Generic Equations by Model Type

Type	Equation
Deterministic	$N = e^{K\left(\alpha - \beta^2\right)}$ *"N is a function of K, α, and β"*
Probabilistic	$P\left(N \middle| \alpha, \beta\right)$ *"The probability of N given α and β"*
Dynamic	$N_{t+1} = N_0 + \alpha \cdot N_t$
Static	$N = \sigma^2 + {}^{\mu}\!/_{\rho}$

[5] Vertical Slider #5

Other Relevant Modeling Terms

The **model framework** is defined as the system of governing equations, parameterization and data structures that represent the formal mathematical specification of a conceptual model (EPA, 2009a).

Mode (of a model): The 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 (EPA, 2009a).

SUMMARY TABLE OF MODEL TYPE

	Probabilistic Models	Deterministic Models	Empirical Models	Mechanistic Models
Also Known As:	Statistical or Stochastic Models	---	'Best Fit' Models	---
Input Data:	Measured Values or Estimated Distributions	Measured Values	Measured Values or Estimated Distributions	Measured Values or Estimated Distributions
Model Output:	Probability Distribution	Single Point Value	Probability Distributions or Single Point Value	Probability Distributions or Single Point Value
Description:	Utilize the entire range of input data to develop a probability distribution of model output	Provide a solution for the state variables rather than a set of probabilistic outcomes	Rely upon the observed relationships among experimental data	Explicitly include the mechanisms or processes between the state variables

THE ROLE OF MODELING

The use of models has increased significantly. Although, models do not generate "truth", they can provide analyses and information used to inform the EPA's decision making process. Policy decisions should be informed by the best information and data. However, researchers are confronted with many constraints when obtaining data [e.g. time, access, and resources (funding, equipment, staff)].

Where there is a shortage of data and information, models can be used to provide useful insight. In general, models can help users study the behavior of ecological systems, design field studies, interpret data, and generalize results (EPA, 2009a). Models are used to make long- and short-term forecasts to extrapolate from the past and answer "what-if" questions. Models can also be used to provide concise summaries of data, in both diagnostic and regulatory contexts (NRC, 2007).

The relationship between data and models is changing. The increasing availability of data may promote new model development or application of existing models to new data. However, this requires that data are used appropriately with models. The limitations from uncertainties and assumptions associated with any model must be considered – as with observational data – before model generated results are applied in any context.

"Fundamentally, the reason for modeling is a lack of full access, either in time or space, to the phenomena of interest. In areas where public policy and public safety are at stake, the burden is on the modeler to demonstrate the degree of correspondence between the model and the material world it seeks to represent and to delineate the limits of that correspondence."
– Oreskes et al. 1994

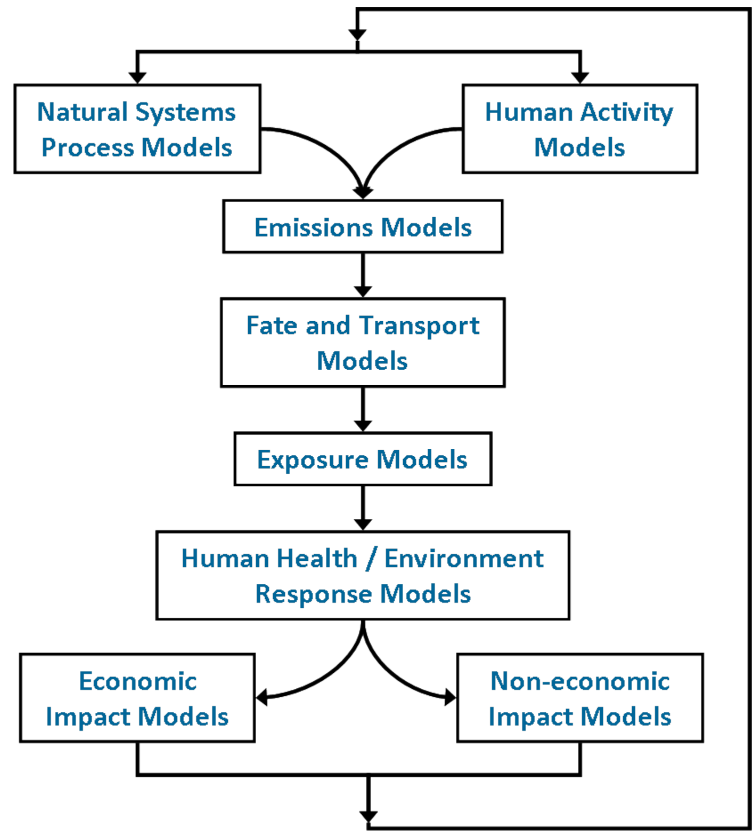
ENVIRONMENTAL MODELS USED BY EPA

Environmental models are categorized into groups representing a continuum of processes which translate the interactions between human activities and natural processes into human health and environmental impacts. The CREM *Guidance Document* (EPA, 2009a) identifies the classes of environmental models used by the EPA:

- **Human Activity Models**
- **Natural Systems Process**
- **Emissions Models**
- **Fate and Transport Models**
- **Exposure Models**
- **Human Health Effects Models**
- **Ecological Effects Models**
- **Economic Impact Models**
- **Noneconomic Impact Models**

Additional Web Resource:

Registry of EPA Applications, Models and Databases (READ) houses ~ 150 models used, developed, or funded by the EPA. It serves as the central repository of the Agency's models, across all disciplines.



Classes of Environmental Models: These classes represent a research continuum from human activities and natural system processes to environmental and economic impacts. Modified from NRC (2007).

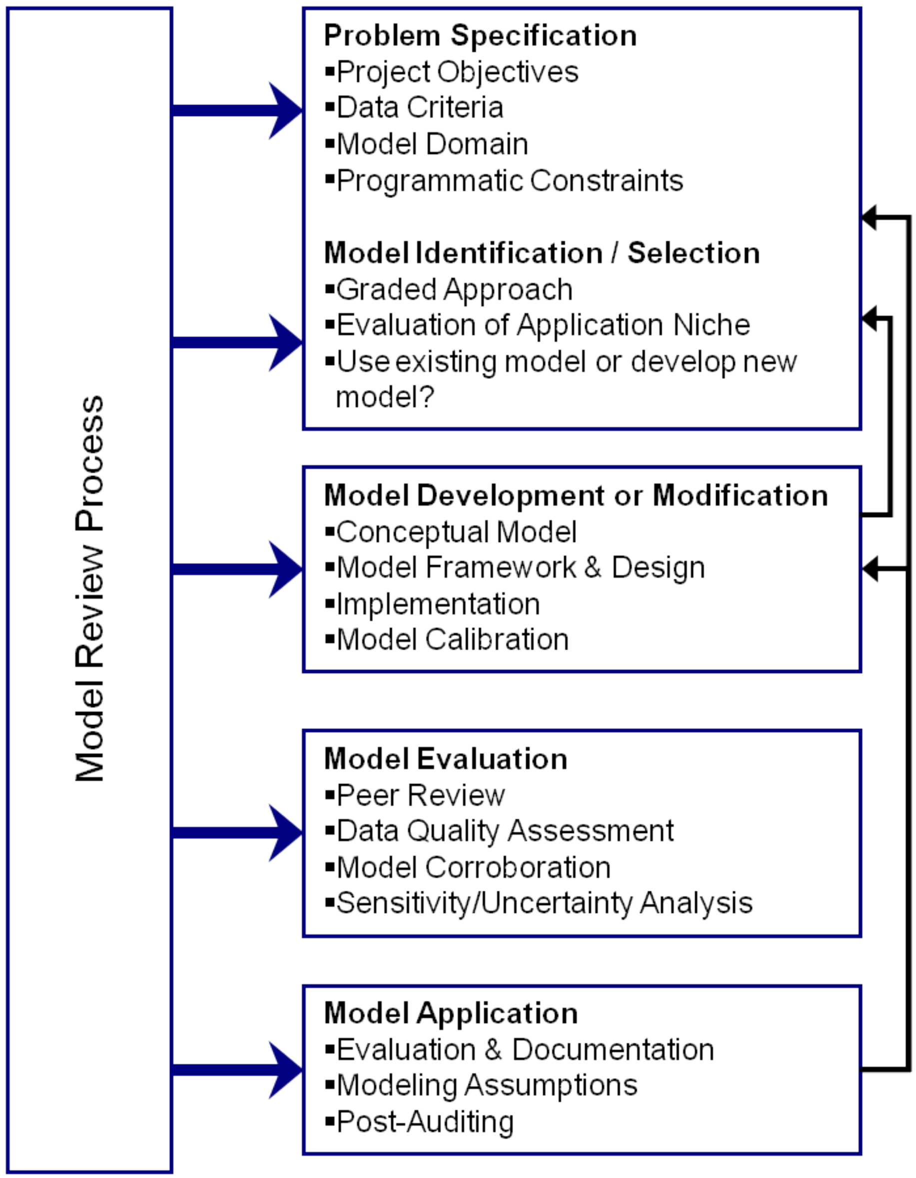
THE MODEL LIFE-CYCLE

The model life-cycle is ongoing, and there are many instances when earlier stages are revisited to refine the model. The life-cycle follows a general iterative progression shown in the figure to the right and described below (from EPA, 2009a):

- **Identification**
  - Determine correct decision-related questions and establish modeling objectives
  - Define the purpose of the modeling activity
  - Specify the model application context
- **Development**
  - Develop the **conceptual model** that reflects the underlying science of included processes
  - Derive the mathematical representation of that science and then encode into a computer program
- **Evaluation**
  - Peer Review
  - Conduct formal testing to ensure model expressions have been encoded correctly
  - Test model outputs by comparisons with empirical (and independent) data
- **Application**
  - Run the model and analyze outputs to inform a decision

(The figure and caption are on the next page.)

Additional Web Resource:
Further information regarding the model life-cycle can be found in The Model Life-cycle module.



The Primary Stages of the Model Life-cycle: Identification/Selection, Development, Evaluation, and Application. Modified from EPA (2009a).

AN ALTERNATIVE MODELING LIFE-CYCLE

Not every project requires the full development of a new model; often there are existing models which can be applied to a specific situation. In these instances, there is an alternative model life-cycle; which involves model evaluation, application, and as needed, **post-auditing**.

In the modified life-cycle, a model is selected that meets the requirements of the specified problem. Once selected, a model may require **calibration** or site-specific parameter values. Likewise, other qualitative evaluations of the model may further corroborate its application. **([1]Example of Site Specific Calibration)**

After the model has been applied, post-auditing can determine whether the predicted model outcome(s) were observed. The model post-audit process involves monitoring the modeled system, after implementing a remedial or management action, to determine whether the actual system response concurs with that predicted by the model. Post-audits can also be used to evaluate how well stake-holder and decision-making roles were integrated during the development stages (Manno et al., 2008; EPA, 2009a).

(The pop-out window is located on the next page; the figure and caption are on the page after that.)

[1]Pop-out Window

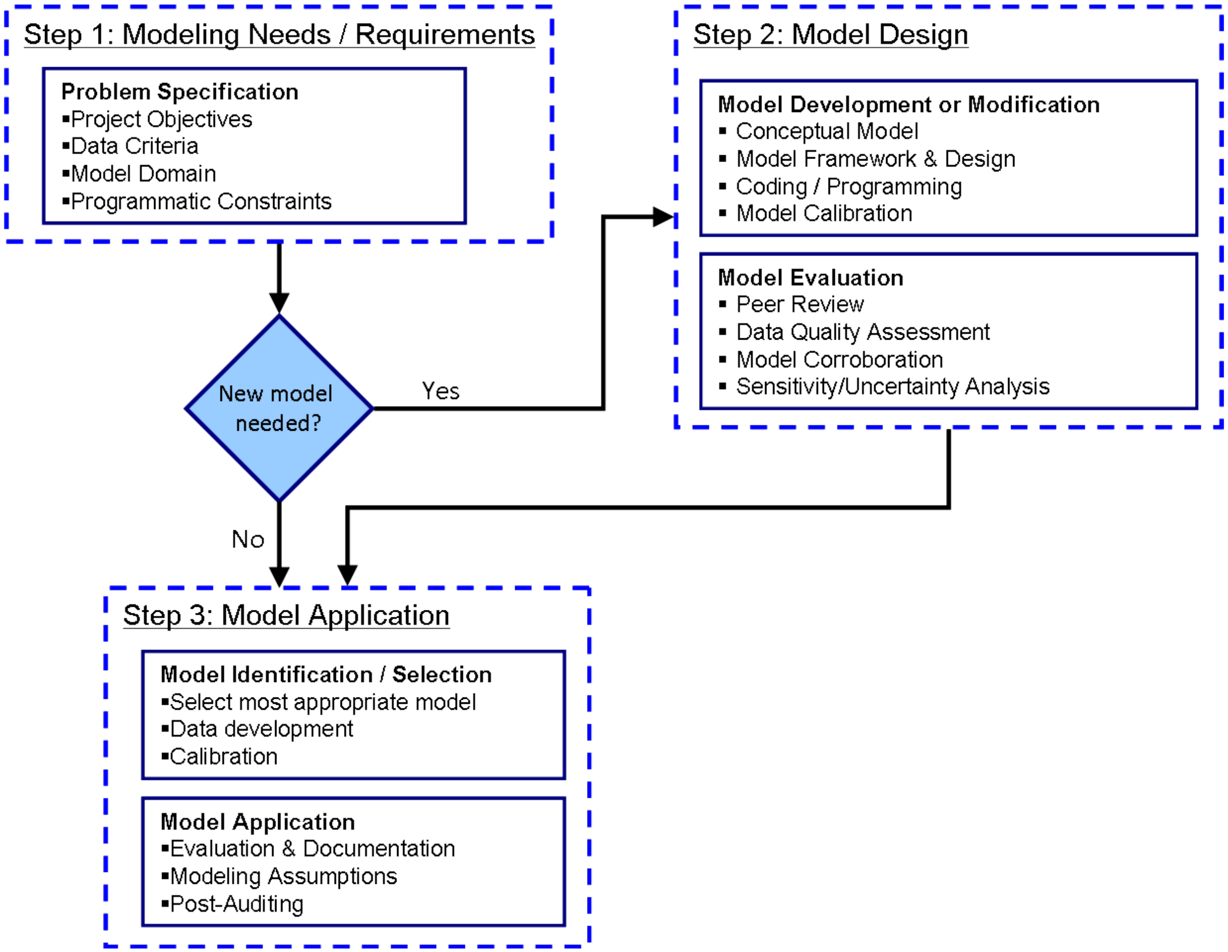
Site Specific Calibration (EPA, 2009a)

When data for quantifying one or more parameter values are limited, calibration exercises can be used to find solutions that result in the 'best fit' of the model. However, these solutions will not provide meaningful information unless they are based on measured physically defensible ranges. Therefore, this type of calibration should be undertaken with caution.

The use of **calibration** to improve model performance varies because of the many concerns associated with it. Often, the appropriateness of calibration may be a function of the modeling activities undertaken.

For example, the EPA's Office of Water's standard practice is 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).

Additional Web Resource:
Registry of EPA Applications, Models and Databases (READ)



An Alternate Version of the Model Life-cycle: When model development is not required a modified version of the life-cycle is appropriate. If an existing model will work for the specified problem, model development (and design) is circumvented; leaving three steps to the life-cycle (shown above with dashed lines). The stages of the life-cycle defined by EPA (2009a) appear in the solid boxes. Recall that model evaluation occurs during the Development and Application Stages.

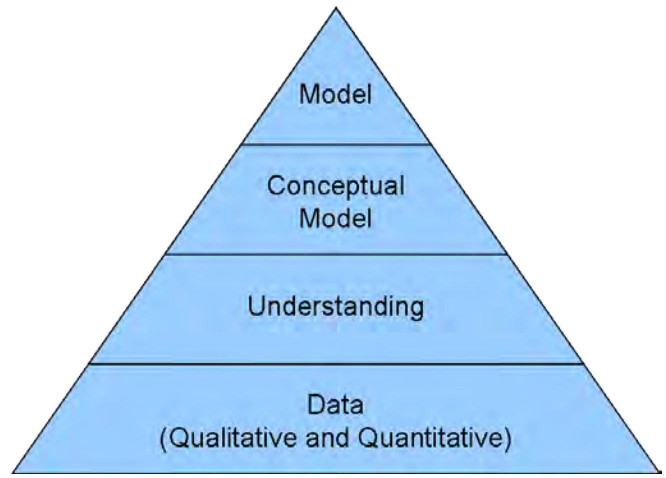
THE IMPORTANCE OF DATA QUALITY

The [1]**quality of the data is fundamental to environmental modeling;** and pertinent not only during model application, but throughout the modeling life-cycle.
The quality of a model is also governed by model structure, scientific understanding, evaluation, etc. Quality assurance is therefore necessary throughout the stages of the modeling life-cycle.

[2]**Indicators of data quality** include the quantitative and qualitative measures of principal quality attributes (EPA, 2009a).

[3]**Quality assurance (QA),** quality control, and **peer review** also play important roles in the Agency's modeling efforts. The data are subject to data quality objectives and other QA measures. Similarly, Quality Assurance Project Plans help guide model development, evaluation, and application. Together, quality assurance requirements are the means to overall **transparency**.

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



A Foundation of Data Quality: Data provide the foundation for our understandings which motivate the development and application of environmental models. Data are used during parameter estimation events, calibration processes, and ultimately model application. Model developers and users should consider:

"what goes in is equal to what comes out"

that is to say, data which is poor in quality will not yield model results with higher quality.

[2] *Vertical Slider #2*

Indicators of data quality

- **Precision** – the quality of being reproducible in amount or performance
- **Bias** – systematic deviation between a measured (i.e., observed) or computed value and its “true” value.
- **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
- **Comparability** – a measure of the confidence with which one data set or method can be compared to another
- **Completeness** – a measure of the amount of valid data obtained from a measurement system
- **Sensitivity** – The degree to which the model outputs are affected by changes in a selected input parameters.

[3] *Vertical Slider #3*

Data and Model Quality Assurance

Additional Web Resource:
CREM’s training module on **QA of Modeling Activities** is coming soon!

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

LEGAL ASPECTS WHEN EPA USES MODELS

A number of laws serve as EPA's foundation for protecting the environment and public health. The Administrative Procedure Act (5 U.S.C. § 553) requires EPA to provide the public notice and an opportunity to comment on its rulemakings.

If a rule is supported by a model, this legal obligation means the Agency must provide the public notice of the Agency's use of the model and an opportunity to comment on the assumptions and **algorithms** that are built into the model, along with the other scientific components of the regulation or rule-making.

Further, it must be clear how a particular model may be used, and the Agency must provide sufficient information about the model for public comment. The legal challenges to the Agency's actions in enforcing those laws could be classified into two categories identified in the adjacent panel (adapted from McGarity and Wagner, 2003).

Process Challenges
Procedural challenges are usually directed at the overall transparency of the modeling exercise and the adequacy of any notice and opportunity for public comment that the agency might be required to provide.

- [1] **Example of a legal challenge to the review process of a model**

Substantive Challenges
These challenges are mounted against areas of technical disagreements with assumptions of the model or the context in which the model was applied.

- [2] **Example of a legal challenge to the scientific components of a model**

Additional Web Resource:

In the Legal Aspects of Environmental Modeling module, we explore how the Agency's regulatory actions (related to modeling) have been challenged and **point to best modeling practices related to those challenges.**

[1] *Pop-out Window #1*

Legal Challenges to the Scientific Components of a Model

***American Forest & Paper Assn v. U.S. EPA*, 294 F.3d 113 (D.C. Cir. 2002)**
American Forest and Paper Association challenged EPA's reliance on conservative assumptions regarding how to extrapolate from toxicity studies on animals to humans. These assumptions were pivotal to EPA's refusal to delete methanol from the list of hazardous air pollutants under the Clean Air Act. The court rejected this challenge, finding that EPA's assumptions were well supported and fully justified and therefore not arbitrary or capricious.

***Appalachian Power Co. v. U.S. EPA (II)*, 249 F.3d 1032 (D.C. Cir. 2001)**
Appalachian Power Company successfully challenged the Agency's use of a model for predicting growth rates of electricity usage in setting emissions controls. The court found that the assumptions of the model – and the subsequent predictions of a decrease in power consumption – were arbitrary because they were not supported by the available evidence.

However, the court did note that EPA had the authority to develop generic, abstracted models for such predictions but the assumptions need to be based on the best available evidence.

[2] *Pop-out Window #2*

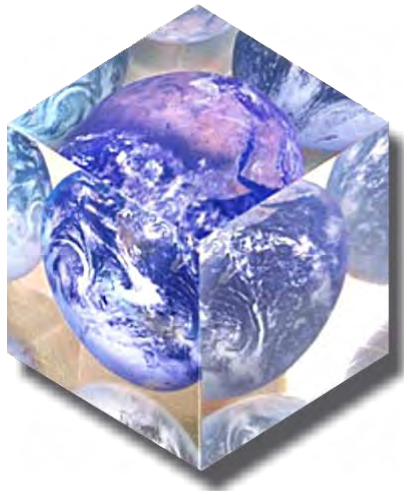
Legal Challenges to the Validation and Review Process of a Model

***McLouth Steel Products Corp. v. Thomas*, 838 F.2d 1317 (D.C. Cir. 1988)**
The McLouth Steel Products Corporation (McLouth) petitioned EPA to de-list a waste stream from its list of hazardous wastes. EPA had used a vertical and horizontal spread model (VHS) to predict the leachate levels of the hazardous components of McLouth's waste.

McLouth argued that EPA had never subjected the model to public notice and comment and challenged the use of the model in this very limited rulemaking proceeding. The court agreed, rejecting EPA's contention that the model [use] was just a policy statement and not a legislative rule. The court remanded the matter to the EPA and held that EPA gave the effect of a rule to its VHS model without having exposed the model to the comment process required for rules.

SUMMARY

- According to the EPA (2009a) a **model** is defined as:
  "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."
- The types of the environmental models used by the EPA include fate and transport models, emissions and activities models, exposure models, and impact models.
- The mode life-cycle includes problem identification, development, evaluation, and application. Iterative peer reviews are an important component throughout a model's life-cycle.
- Models can provide meaningful data to inform the decision making process when the appropriate actions and precautions have taken place during the life-cycle of the model.
- Models can not improve the data that goes into them. Model results should not be considered truths.



Transparency: In the past, models have been considered a 'black box' of the research or regulatory process (Pascual, 2004). Through better understandings of the model life-cycle and best modeling practices, models can be built from plexiglass!

YOU HAVE REACHED THE END OF
THE ENVIRONMENTAL MODELING 101 MODULE.

REFERENCES

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.

Manno, J., R. Smardon, J. V. DePinto, E. T. Cloyd and S. Del Granado. 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.

McGarity, T. O. and W. E. Wagner 2003. Legal Aspects of the Regulatory Use of Environmental Modeling. Environmental Law Reporter 33(10): 10751-10774.

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

Pascual, P. 2004. Building The Black Box Out Of Plexiglass. Risk Policy Report 11(2): 3.

Van Waveren, R. H., S. Groot, H. Scholten, F. Van Geer, H. Wösten, R. Koeze and J. Noort. 2000. Good Modelling Practice Handbook (PDF) (165 pp, 1Mb, About PDF). STOWA report 99-05. Leystad, The Netherlands. STOWA, Utrecht, RWS-RIZA, Dutch Department of Public Works

GLOSSARY

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

Analogous Models: When nonhuman species are used to demonstrate the potential health effects of chemicals on humans

Calibration: The process of adjusting model parameters within physically defensible ranges until the resulting predictions give the best possible fit to the observed data. In some disciplines, calibration is also referred to as “parameter estimation”.

Computational models: Computational models express the relationships among components of a system using mathematical representations (Van Waveren et al., 2000).

Conceptual Models: 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.

Ecological Effects Models: Provide a statistical relationship between a level of pollutant exposure and a particular ecological indicator.

Economic Impact Models: Used in rulemaking, priority setting, enforcement; model output as a monetary value.

Emissions Models: Estimate the rate or amount of pollutant emissions to water bodies and atmosphere.

Exposure Models: Estimate the dose of pollutant which humans or animals are exposed.

Fate and Transport Models: Calculate the movement of pollutants in the environment. Further classified into Subsurface Water Quality Models, Surface Water Quality Models, and Air Quality Models.

Human Activity Models: Simulate human activities and the behaviors that result in emission of pollutants.

Human Health Effects Models: Provide a statistical relationship between a dose of a chemical and an adverse human health effect.

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.

Natural Systems Process: Simulate dynamics of ecosystems that give rise to fluxes of nutrients and/or emissions.

Noneconomic Impact Models: Evaluate the effects of contaminants on a variety of noneconomic parameters (e.g. crop yields).

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.

Peer Review: Performed by independent and objective experts, a review of and judgment on a model's underlying science, the process through which it was developed, and its overall "trustworthiness" and "reliability" for prediction.

Post-auditing: Assesses a model's ability to provide valuable predictions of future conditions for management decisions.

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

System: A collection of objects or variables and the relations among them.

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.

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

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).