Best Modeling Practices: Model Development

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

Table of Contents

PREFACE 3
DESIGN 4
INTRODUCTION 5
- Introduction 5
- Model Development 7

PROBLEM 9
- Modeling Objectives 9
- Model Type 10
- Model Scope 11
- Data Criteria 13
- Domain of Applicability 14

SELECTION 15
- Model Selection 15
- Case Study 17

FRAMEWORK 19
- Conceptual Model 19
- Model Design 22
- Model Complexity 23
- Model Code 25

APPLICATION TOOL 26
- Application Tool 26
- Input Data 27
- Calibration 29
- Version Control 31
- Documentation 33
- Case Study 35

SUMMARY 36
- Development 36
- Recommendations 37
- End of Module 38

REFERENCES 39
- Page 1 39
- Page 2 40

GLOSSARY 41

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

BEST MODELING PRACTICES: DEVELOPMENT

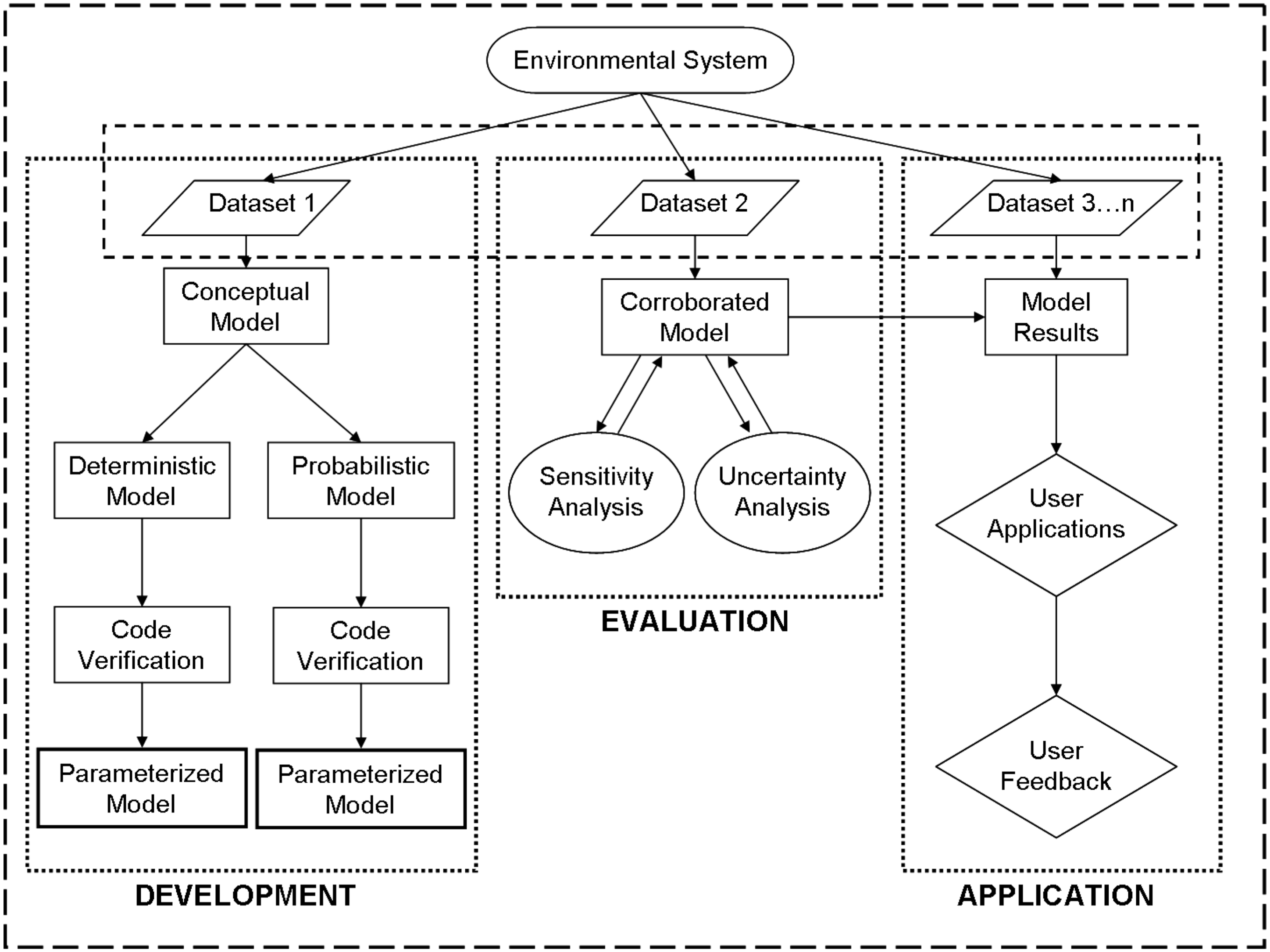
This module will continue on the fundamental concepts outlined in previous modules: Environmental Modeling 101 and The Model Life-cycle. The objectives of this module are to:

- Identify the 'best modeling practices' and strategies for the Development Stage of the model life-cycle
- Define the steps of model development

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** (EPA, 2009a).

The modeling practices within each life-cycle stage are identified in the diagram to the right from EPA's *Guidance on the Development, Evaluation, and Application of Environmental Models* (EPA, 2009a).

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



Three primary stages of the model life-cycle: Development, Evaluation, and Application. Modified from EPA (2009a). An important feature of this diagram is the independent data sets that are used during each stage.

MODEL DEVELOPMENT

During the Development Stage, our best interpretations of real-world processes are translated into **computational models** through a **conceptual model**. During later stages of the model's life-cycle, the model development team (developers, intended users, and decision makers) will consult the objectives and other decisions defined during the Development Stage.

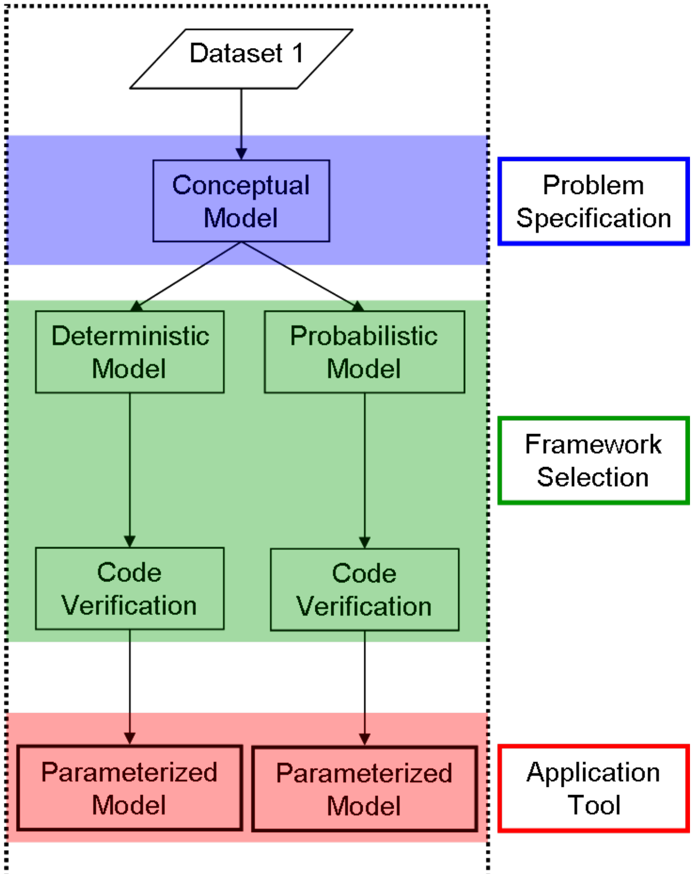
The [1]**incremental steps of model development** include (EPA, 2009a):

1. Problem Specification and Conceptual Model Development
2. Model Selection / Development
3. Application Tool

Each step the development team makes during the life-cycle stages of a model should be clearly described and defined. One technique is to keep a [2]**model journal** (Van Waveren et al., 2000; EPA, 2010).

(The vertical sliders are on the next page.)

[1]*Vertical Slider #1*	[2]*Vertical Slider #2*



DEVELOPMENT

The incremental steps of model development; modified from EPA (2009a).

A Model Journal

The model journal can be used to organize and catalog changes to the model or justifications of chosen procedures for model evaluation. Questions that could be answered by a well kept model journal include:

- What was the pattern of thought followed?
- Which concrete activities were carried out?
- Who carried out which work?
- Which choices were made?

Additional Resource:
[Modeling Journal: Checklist and Template](). 2010. US Environmental Protection Agency. New England Office, New England Regional Laboratory's Office of Environmental Measurement and Evaluation, Quality Assurance Unit

MODELING OBJECTIVES

Initial steps of the Development Stage help to define the project team's goals for an identified problem. Collaboration among the **model development team** helps to ensure that the model is designed to be applied in an appropriate context that can provide useful information to the decision making process.

In some cases, the identified problem may best be supported by an existing model and new model development may not be necessary.

Before formal model development occurs, the project team should (EPA, 2009a):

- Define the modeling objectives
- Define the scope and type of model needed
- Define the criteria applied for the data selection process
- Define the **application niche**
- Define the required level of model performance
- Identify programmatic constraints
- Develop a conceptual model
- Define the evaluation objectives and procedures
- Define model selection criteria (if multiple models exist)

Examples of questions that should be defined by the modeling objectives:

- Is the model to be used in regulation or research?
- Does this problem require the development of a new model?
- Are there existing models (with varying degrees of complexity) that are sufficient?
- Who are the intended users of the model?
- What information from the model is useful for decision makers?

MODEL TYPE

The type of model developed is determined by the needs of the model development team and other characteristics defined in the modeling objectives.

Model types (definitions to the right) are not always mutually exclusive. Complex or integrated models could have components representing each of the different types (i.e. a mechanistic model could have empirical relationships within it or a probabilistic model may have **parameters** with unknown distributions).

The **mathematical 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).

Additional Web Resource:
Model types are also discussed in the Environmental Modeling 101 module.

Review of Model Type (EPA, 2009a; 2009c)

Probabilistic vs. Deterministic

Probabilistic models are sometimes referred to as statistical or stochastic models. These models utilize the entire range of input data to develop a probability distribution of model output for the state variable(s).
Deterministic models provide a solution for the **state variable(s)** rather than a set of probabilistic outcomes.

Dynamic vs. Static

Dynamic models makes predictions about the way a **system** changes with time or space.
Static models make predictions about the way a system changes as the value of an independent variable changes.

Empirical vs. Mechanistic

Empirical models include very little information on the underlying mechanisms and rely upon the observed relationship among experimental data.
Mechanistic models explicitly include the mechanisms or processes between the state variables. The parameters in mechanistic models should be supported by data and have real-world interpretations.

MODEL SCOPE

Models are often developed to inform a decision, but can also be developed in a more research focused context. Models should be applied to the **system** for which they were originally intended.

The model **application niche** is bounded and defined by the scope of the model; it is defined as the set of conditions under which the use of a model is scientifically defensible. Examples of the types of information that should be defined by the model scope are to the right. Important characteristics of the model scope include:

- Spatial and Temporal scales
- [1]**A description of "Process Level Detail"**

Defining the scope of the model is important to do early in the Development Stage of the model life-cycle because it has important influence over *how* and *when* the model is applied.

Examples of questions that should guide the determination of model scope:

- What scale should the model represent (i.e. temporal and spatial)?
- What level of process detail is required?
- What are the stressors to the system?
- What are the **state variables** of concern?
- What are the boundaries of the system of interest?

[1] ***Pop-out Window***

Process Level Details of a Model:

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 (EPA, 2009a).

Environmental models, in part, simulate processes that relate variables (measureable quantities, often model outputs) in the model. These processes are described as (with examples):

- **Physical Processes** – Advection, Diffusion, Settling, Heat Transfer
- **Chemical Processes** – Oxidation, Reduction, Hydrolysis, Equilibria
- **Biological Processes** – Photosynthesis, Respiration, Grazing

These processes occur at various scales (temporal and spatial) – the model development team will, through some degree of simplification, determine the appropriate level of detail necessary to best capture each process of the model. The assumptions of a model often simplify the system to a level of detail that can be modeled.

DATA CRITERIA

Another early step in the Development Stage is to define the necessary data to run the model and the acceptable level of **uncertainty** associated with that data (and model output data).

Data Quality Objectives assist project planners to generate specifications for quality assurance planning. These specifications will partially determine the appropriate boundary conditions and complexity for the model being developed.

Additional Web Resource:

Quality assurance planning is also discussed in the Best Modeling Practices: Evaluation module and the **QA of Modeling Activities module** (*coming soon*).

Data Quality Objectives should contain information that will (EPA, 2000; 2009a):

- Determine the acceptable level of uncertainty that enables the model to be used for the intended purpose(s)
- Provide specifications for model and data quality; and the associated checks
- Guide the design of monitoring plans
- Guide the model development process
- State requirements of data that will limit decision errors

DOMAIN OF APPLICABILITY

Each model has a domain of applicability that is defined during the Development Stage. The model's domain of applicability can be thought of as the range of appropriate modeling scenarios defined by a set of specific conditions – this is the definition of a model's **application niche** (EPA, 2009a). It is under these conditions where the model is scientifically defensible and most relevant to the system being modeled.

The model development team must identify the environmental domain to be modeled and then specify the processes and conditions within that domain.

Guidelines for Defining a Domain of Applicability (EPA, 2009a)

- Identify the transport and transformation processes relevant to the project objectives
- Define the important time and space scales inherent in the aforementioned processes within the model's domain. These should be compared to the time and space scales of the project objectives
- Identify any unique conditions of the domain that will affect model selection or new model construction.

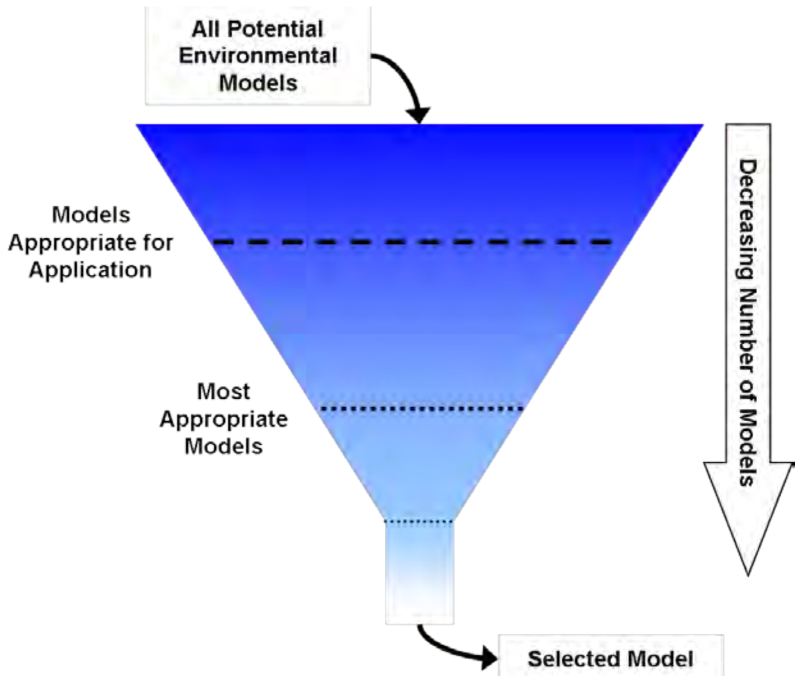
MODEL SELECTION

Before the **model development team** continues the development process, it is important to make sure that applicable models do not already exist. When there are existing models, the team is faced with the challenge of ⇕[1] **selecting an appropriate model**. The team may use quantitative and qualitative assessments of model performance to select the most appropriate model for their application scenario. Information about models that are used, developed or funded by the EPA are housed in the ⇕[2] **Models Knowledge Base.**

When choosing an appropriate model, the research team should also consider the following:

- Does sound science support the underlying hypothesis?
- Is the model's complexity appropriate for the problem at hand?
- Do the quality and quantity of the data support the choice of model?
- Does the model structure reflect all the relevant components of the conceptual model?
- Has the model code been developed and verified?

(The vertical sliders are on the next page.)

[1]*Vertical Slider #1*	[2]*Vertical Slider #2*
 During a model selection process, all potential environmental models are examined to determine to the most appropriate models and the selected model.	**Additional Web Resource:** The EPA has established the Council for Regulatory Environmental Modeling (CREM) to aid in inter- and intra-agency modeling efforts. CREM has built a Registry of EPA Applications, Models and Databases (READ) which stores close to 150 environmental models that are developed, used, or funded by EPA. The database has models from across the Agency, each record containing defining

MODEL SELECTION – A CASE STUDY

The success of a modeling effort to support the development of a **TMDL** is highly dependent on an understanding of the complexity of the water quality problems. This understanding will assist the **model development team** in defining the required accuracy, analyzing the implication of various simplifying assumptions, and eventually selecting an appropriate modeling strategy and modeling tools (EPA, 1995).

The selected model should provide a balance between simplicity and inclusion of the important processes affecting water quality in the stream or river.

If an appropriate model does not exist, the model development team should continue the development process.

Additional Web Resource:
For further information on TMDLs see:
Impaired Waters and Total Maximum Daily Loads

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

Models	Multiple Point Sources of BOD	Distributed Sources of CBOD	Benthic Oxygen Demand	Net Algal Production	Longitudinal Dispersion	BOD Settling	Time-variable Waste Loads and Water Quality	Time-variable Flow
Multi-SMP	X	▼	◊			***		
QUAL2E	X	X	◊	D	X	●	***	
CE-QUAL-RIV1	X	X	◊		◊		X	X
ASP5	X	X	◊	Δ	–	X	X	X
HSPF	X	X	X	X		X	X	X

X	Available feature
◊	Specified (i.e. input to the model as a forcing function)
Δ	Simulated in a nutrient-algal cycle
▼	Can be simulated approximately by input of load at the beginning of each multiple segment
●	Can be simulated by making $K_r > K_d$
***	Meteorology only

During the model selection process, EPA (1995) outlines strategies for comparing and selecting models for the development of a TMDL. In this example, four models are compared by common features among the models. Though qualitative, the expert knowledge of the team allows for the selection of a model that fits their needs the best. If ‘Net Algal Production’ was an important trait for the specific modeling effort, the CE-QUAL-RIV1 and Multi-SMP models may not be selected because those processes are omitted from the model.

CONCEPTUAL MODEL

Another step of the Development Stage is to formalize a **conceptual model**. The modeling team should also identify any potential constraints (time, budget, work allocation, etc.) that would hinder model development.

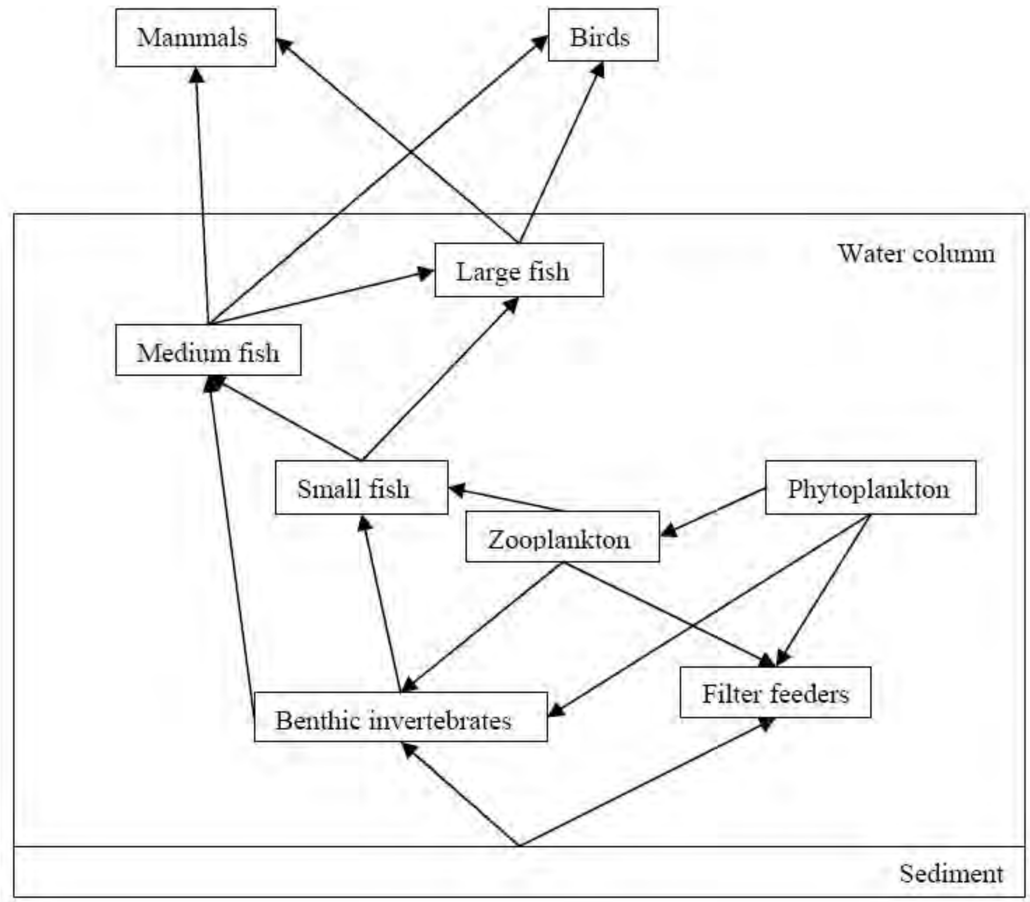
At this step it is also important to identify the skill sets required for the programming and application of the model; which should then be compared with the skill sets that are available within the project team.

The conceptual model should include the most important behaviors of the systems, object, or process relevant to the problem of interest. It is important for the developers to also make note of the science (experiments, mechanistic evidence, peer-reviewed literature, etc.) behind each element and assumption of the conceptual model. When relevant, the strengths and weaknesses of each constituent hypothesis should be described (EPA, 2009a).

- [1] **An example of a conceptual model diagram**
- [2] **Components and benefits and of conceptual models**
- [3] **Interactive Conceptual Diagrams**

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

[1]*Vertical Slider #1*



Example Conceptual Model: The aquatic food web of the KABAM model. (CREM's Models Knowledge Base). Arrows depict direction of trophic transfer of bioaccumulated pesticides from lower levels to higher levels of the food web (EPA, 2009b).

[2]*Vertical Slider #2*

Conceptual Model Components and Benefits (EPA, 1998)

Conceptual models consist of two principal components:

- A set of hypotheses that describe predicted relationships within the system and the rationale for their selection
- A diagram that illustrates the relationships presented in the risk hypotheses.

The process of creating a conceptual model is a powerful learning tool. The benefits of developing a conceptual model include:

- Conceptual models are easily modified as knowledge increases.
- Conceptual models highlight what is known and not known and can be used to plan future work.
- Conceptual models provide an explicit expression of the assumptions and understanding of a system for others to evaluate.
- Conceptual models provide a framework for prediction and are the template for generating more hypotheses.

[3]*Vertical Slider #3*

Interactive Conceptual Diagrams (ICDs)

The ICD application is a tool designed to help users create conceptual diagrams for causal assessments, and then link supporting literature to those diagrams (EPA, 2010).

These conceptual diagrams illustrate hypothesized pathways by which human activities and associated sources and stressors may lead to biotic responses in aquatic systems.

Useful links:

- An introduction to ICD
- ICD User Guide

MODEL DESIGN AND FRAMEWORK

The design and functionality of a model is determined by the model development team. Typically, the model is designed with an appropriate degree of complexity in order to produce output that can properly inform a decision.

The **mode** (of a model) is 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).

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).

Additional Web Resource:

The Environmental Modeling 101 module defines and discusses the different **types of models** and **model structure.** Collectively, the model type, mode, and structure comprise the **model framework.**

MODEL COMPLEXITY

The optimum level of complexity should be determined by making tradeoffs between competing objectives of overall model simplicity and including all the relevant processes. When possible, simple competing conceptual models should be tested against one another to determine the appropriate degree of complexity. The NRC (2007) recommends that models used in the regulatory process should be no more complicated than is necessary to inform regulatory decisions.

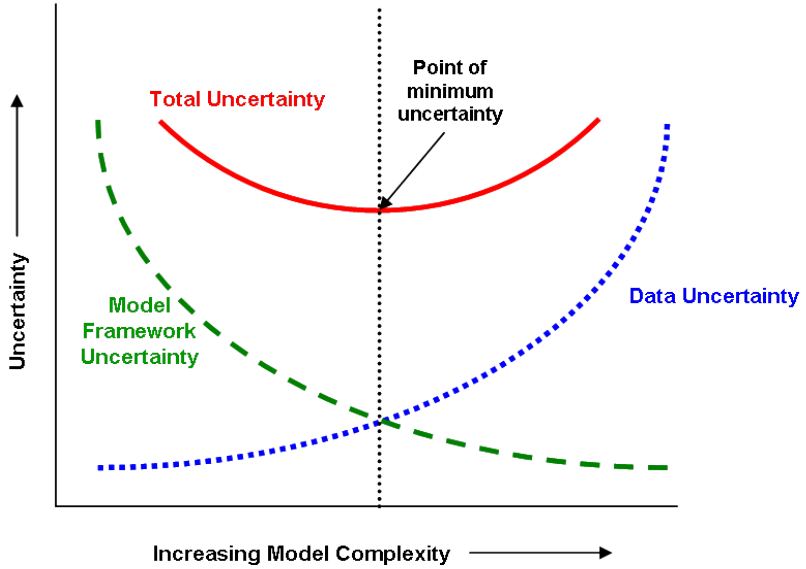
The associated **uncertainty** for each model should be considered in the context of complexity. Two types of uncertainty and their [1] **relation to total uncertainty and model complexity are shown:**

Model Framework Uncertainty – limitations in the mathematical model or technique used to represent the system of interest (i.e. Structure Uncertainty)

Data Uncertainty – measurement, collection, and treatment errors of the data used to characterize the model parameters (i.e. Input Uncertainty)

Uncertainty is typically characterized by conducting [2] **sensitivity and uncertainty analyses.**

(The vertical sliders are on the next page.)

[1]*Vertical Slider #1*	[2]*Vertical Slider #2*
 **Relating Model Uncertainty and Complexity.** Total model uncertainty (solid line) increases as the model becomes increasingly simple or increasingly complex. Data uncertainty (dotted line) increases with increasing model complexity; contrary to model framework uncertainty (dashed line) which decreases with increased model complexity. Modified from Hanna (1988) and EPA (2009a).	**Sensitivity and Uncertainty Analyses (EPA, 2009a):** **Sensitivity analysis** is a useful technique that evaluates the effect of changes in input values or assumptions (including boundaries and model functional form) on the outputs. Sensitivity analysis can also be used to study how uncertainty in a model output can be systematically apportioned to different sources of uncertainty in the model input. **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). **Additional Web Resources:** For elaboration on uncertainty and sensitivity analyses please see these other modules: • Best Modeling Practices: Evaluation • Sensitivity and Uncertainty Analyses

MODEL CODE

Model coding translates the mathematical equations that constitute the model framework into functioning computer code.

Code verification is an important practice which determines that the computer code has no inherent numerical problems with obtaining a solution. Code verification also tests whether the code performs according to its design specifications.

Best Modeling Practices for Model Code from EPA (2009a)

- **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** when the conceptual model is developed and before coding begins helps 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.
- **Use of generic algorithms for common tasks** can often save time and resources, allowing efforts to focus on developing and improving the original aspects of a new model.

THE APPLICATION TOOL

The next step in the development process is the design of an Application Tool. In terms of its applicability and functionality, this is the 'model'. The complexity of the 'tool' can range from a simple spreadsheet application to a stand-alone executable file with a full graphical user interface and data input/output processes.

Once a **model framework** has been selected or developed, the modeler populates the framework with the specific system characteristics needed to address the problem, 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 that can be used to address a specific problem (EPA, 2009a).

Other activities at this point could include the development of model documentation products that could include a database of suitable data, a registry of model versions, or a user's manual.

Components of Application Tool Development:

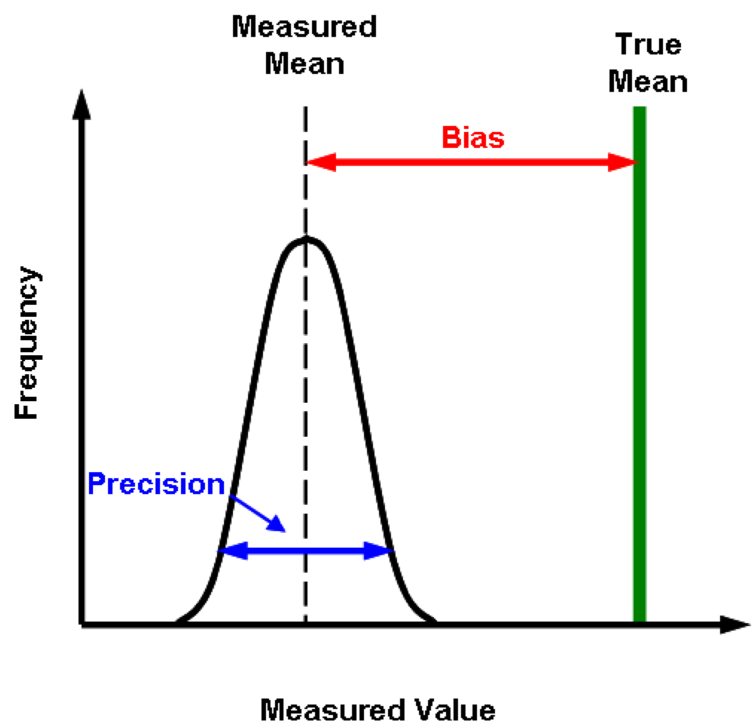
- Input Data – a major source of uncertainty, should meet data quality objectives
- Calibration – should be carefully considered and guidelines should be established
- Version Control – significant changes or improvements to the model should be documented
- Documentation – various means to increase overall **model transparency**

INPUT DATA

Models cannot generate output data that is better in quality than the data that went into the model. As models become more complex, they often require not only more data, but data representing complex processes that may be difficult to acquire.

Data quality objectives should be defined in the quality assurance (QA) plan and data should be obtained by following QA protocols for field sampling, data collection, and analysis (EPA, 2009a). Data quality is described by a set of [1]**indicators.** They are also [2]**represented in this figure**. Additionally, the acceptable level of uncertainty should be determined for input data.

(The vertical sliders are on the next page.)

[1] *Vertical Slider #1*	[2] *Vertical Slider #2*
Data Quality Indicators (EPA, 2002; 2009a) • **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.	 A frequency distribution of measured values compared to the **true mean**. **Bias** represents systematic error in measurement; whereas **precision** is random error in measurement.

CALIBRATION

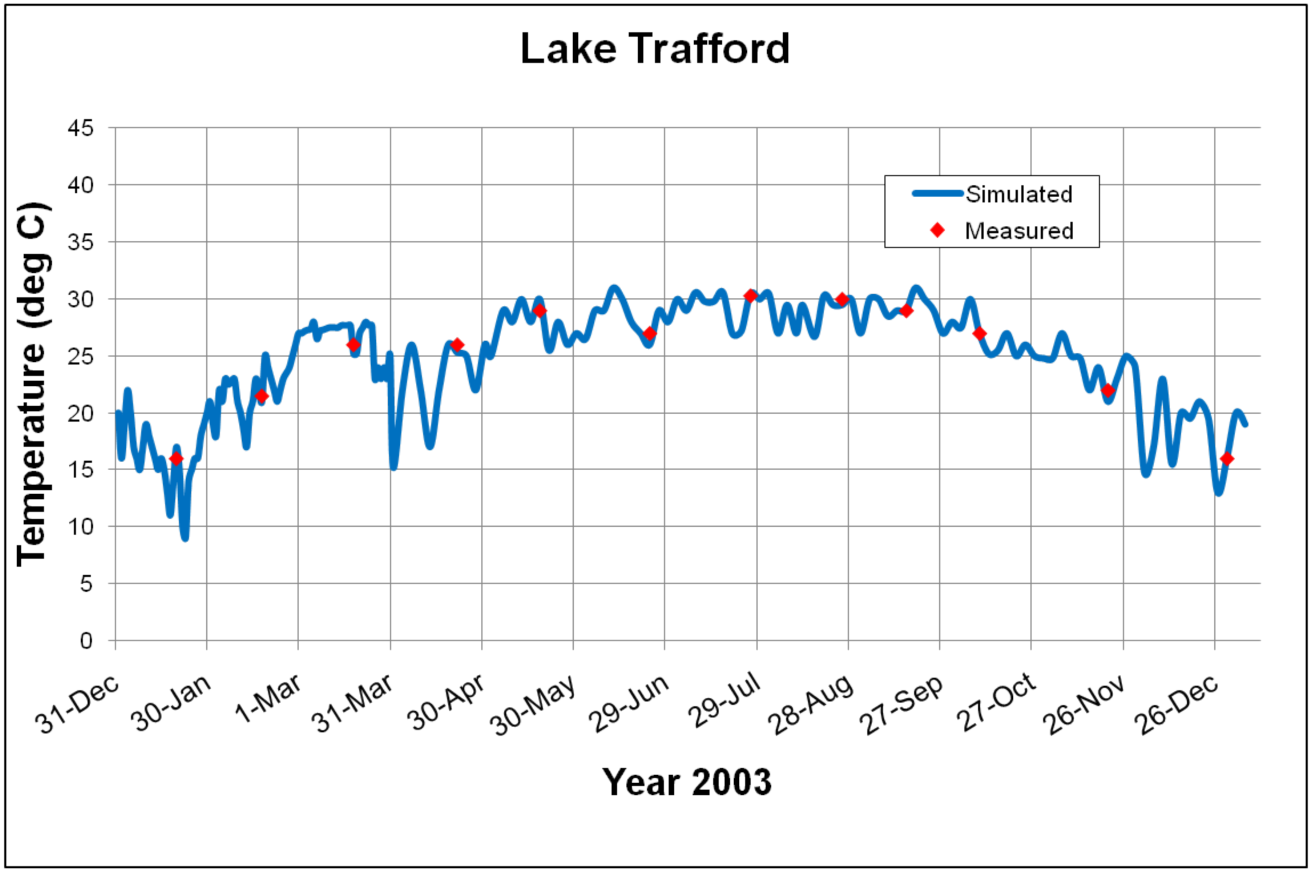
Where possible, model parameters should be characterized using direct measurements from sample populations within the defined scope of the modeling problem. Model parameters are fixed terms during any model simulation or run, but can change for **sensitivity analysis** or **uncertainty analysis** (EPA, 2009a).

Calibration is the process of adjusting model parameters within physically defensible ranges until the resulting predictions give the best possible fit to the observed data (EPA, 2009a). In some disciplines, calibration is also referred to as ‘parameter estimation’. Prior to any calibration routine the model development team should:

- Define objectives of the calibration activity
- Define the acceptance criteria
- Justify the calibration approach and acceptance criteria
- Identify independent data sets that can be used for calibration or evaluation

The practice of calibration varies between each modeling domain. In some applications calibration is avoided at all costs; in other domains calibration is done for each application of the model and can be site-specific (NRC, 2007; EPA, 2009a).

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



In a calibration exercise of the Hydrological Simulation Program-FORTRAN (HSPF) model for Lake Trafford in Florida, the team calibrated simulated annual water temperature by adjusting key parameters; clearly identified and discussed in the report, (FDEP, 2008). *Image adapted from FDEP (2008).*

VERSION CONTROL

Managing the evolving versions of a model through its life-cycle is an important contribution to overall model transparency. Updated versions reflect new scientific findings, acquisition of data, and improved algorithms (NRC, 2007).

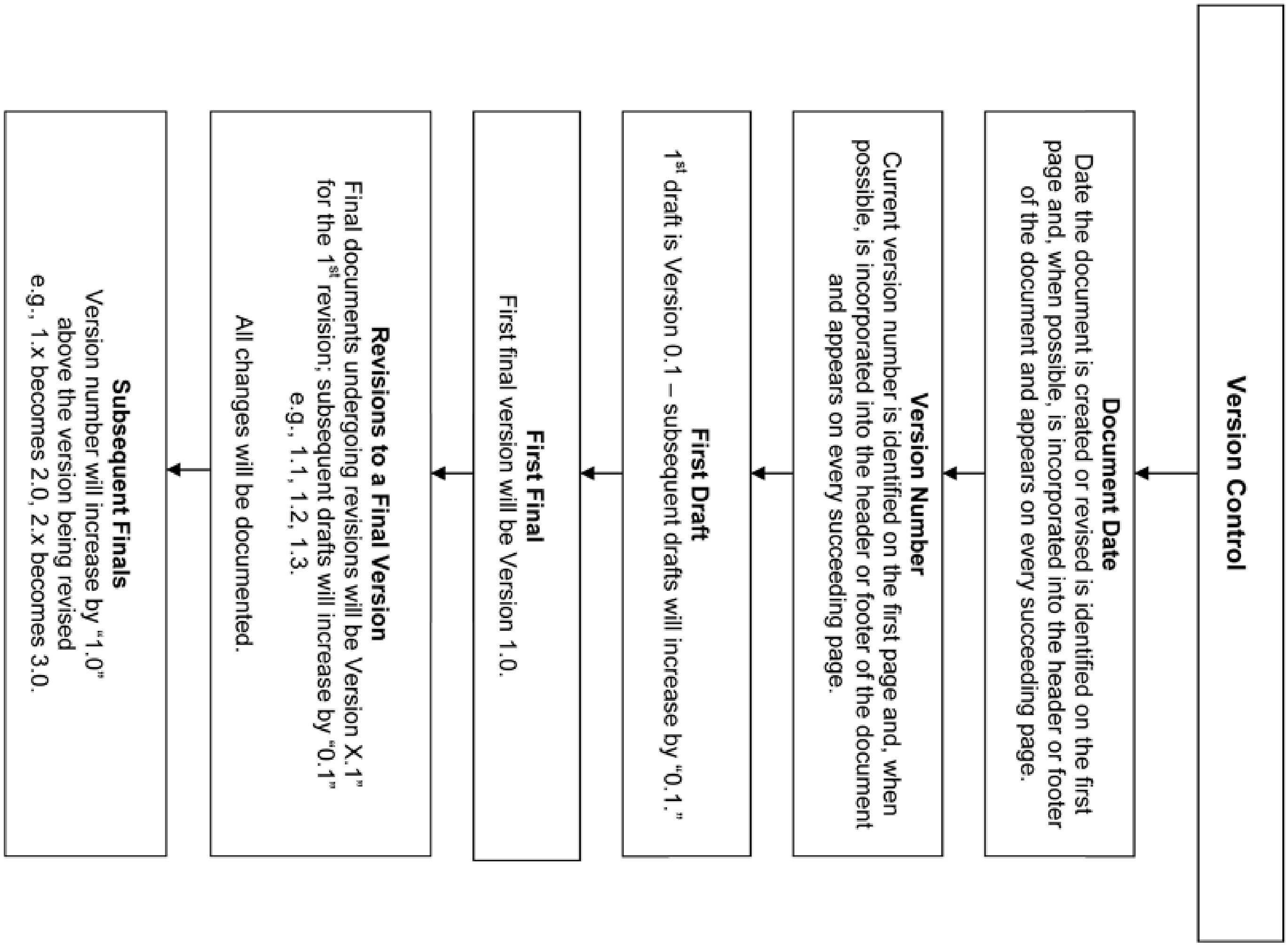
Unless there are changes in model structure, framework, or coding, it is up to the model development team to decide when the changes are substantive enough to justify a new version. During early stages of development, there may be many minor changes to the model but it is important to keep a record of changes that were made, who made them and why.

The National Institute of Health (NIH, 2010) has developed guidelines for version control of written manuscripts. The same principles can be applied to models. The initial draft of the model is "Version 0.1." Subsequent versions are increased by "0.1" until a final version is complete (Version 1.0). Revisions to the final versions are documented as "Version X.1", X.2, etc. [1]**A diagram of the versioning process from the NIH (2010)** or an **example of versioning from the EPA on the "Case Study" subtab.**

Guidelines for Version Control:

- Identify Potential Roles
  - o Version Master – responsible for final version control
  - o Developer(s) – work(s) closely on model development/coding, keep track of changes made to the model
- Appropriate reasons for version updates could be:
  - o Substantive changes in model structure or code
  - o Changes to governing equations
  - o Changes to scope
  - o Changes to input/output sources
- Documentation for each change should include at least:
  - o Date modified
  - o Person responsible for change
  - o Description of change
  - o Justification of change

[1]*Pop-out Window*



Version control guidelines from NIH (2010).

DOCUMENTATION

The longevity of a model is determined, in part, by its ability to evolve as scientific understandings are expanded. ⬍[1]**A well documented model** ensures that the origin and history of the model are not lost when as the core members of the model development team may change or when new user groups begin to use the model.

The decisions and definitions made early in the Development Stage are important to follow so the model is not incorrectly applied. Historical documentation can also provide insight to stakeholders who are interested in understanding a model.

Finally, proper documentation provides a record of critical decisions and their justifications (NRC, 2007). Model documentation can often include an ⬍[2]**evaluation plan,** a database of suitable data, a registry of model versions, or a ⬍[3]**user's manual**.

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

Recommendations for Transparent Documentation (NRC, 2007; EPA, 2009a)

- Documentation should include names of members of the model development team and their affiliation
- Elements of the documentation are written in 'plain English'
- Decisions and justification of critical design elements
- Limitations of the model
- Key peer reviews
- Records should include information surrounding software version releases
- Documentation for any model should clearly state why and how the model can and will be used

[2]*Vertical Slider #2*

Evaluation Plan

An **evaluation plan** outlines the evaluation exercises that will be carried out during the model life-cycle. It is important to determine the appropriate level of evaluation before the model is applied.

Additional Web Resource:

For more discussion of an **evaluation plan** please see:
Best Modeling Practices: Evaluation

[3]*Vertical Slider #3*

Items That Could Be Included In a User's Manual:

- Scientific background of the model
- Definition of application niche
- Description / figure of the conceptual model
- Methods used during development and evaluation
- Installation instructions
- Guidance for model application
- Required input data
- Guidance on interpretation of model results
- Example scenarios / Test cases
- Documentation standards

An example of a user's manual: APEX / TRIM

Registry of EPA Applications, Models and Databases

(READ)

VERSION CONTROL: CASE STUDY

The Exposure Analysis Modeling System (EXAMS) is an interactive software application for formulating aquatic ecosystem models and rapidly evaluating the fate, transport, and exposure concentrations of synthetic organic chemicals including pesticides, industrial materials, and leachates from disposal sites.

Significant releases and a model summary are provided on the EXAMS model website.

EXAMS Releases:

Version	Release Date
2.98.04.06	April 2005
2.98.04.03	May 2004
2.97.5	June 1997
2.94	February 1990

The above table is a good example of how version information is documented and released on a website. For detailed information about each release please check the Release Notes for the EXAMS model. Table was adapted from the EXAMS model website.

Registry of EPA Applications, Models and Databases (READ)

MODEL DEVELOPMENT

The Development Stage of the model life-cycle may be the most important stage because of the many decisions and definitions of the modeling project that occur. The three steps of this stage:

1. Problem Specification and Conceptual Model Development
2. Model Framework Selection / Development
3. Application Tool Development

During the Evaluation and Application Stages, the model development team will often refer to, and rely upon, the documentation set forth during Development Stage.

When models are used to inform the decision making process (as per their intended purpose); the models are not just products of theory and data but are also shaped by the priorities of the decision-makers who are deploying them (Fisher et al., 2010).

Additional Web Resource:

There are additional modules that build upon some of the topics presented in this module. Please see:

- Best Modeling Practices: Evaluation
- Best Modeling Practices: Application
- Sensitivity and Uncertainty Analyses
- QA of Modeling Activities *(coming soon)*

SUMMARY OF RECOMMENDATIONS

Recommendations for model development from EPA (2009a) and NRC (2007):

- Communication between model developers and model users is crucial during model development.
- Each element of the conceptual model should be clearly described (in words, functional expressions, diagrams, and graphs, as necessary), and the science behind each element should be clearly documented.
- Sensitivity analysis should be used early and often.
- When possible, simple competing conceptual models/hypotheses should be tested.
- Capabilities (i.e. complexities) that do not improve model performance substantially should be omitted.
- The optimal level of model complexity should be determined by making appropriate tradeoffs among competing objectives and requirements of the regulatory decision.
- 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.

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

REFERENCES

EPA (US Environmental Protection Agency). 1995. Technical Guidance Manual for Developing Total Maximum Daily Loads (PDF) (258 pp, 11MB, about PDF). Book II: Streams and Rivers Part 1: Biochemical Oxygen Demand/Dissolved Oxygen and Nutrients/Eutrophication. EPA 823-B-95-007. Washington, DC. Office of Water.

EPA (US Environmental Protection Agency). 1996. BIOSCREEN Natural Attenuation Decision Support System User's Manual Version 1.3. EPA/600/R-96/087 (PDF) (100 pp, 1.2MB, about PDF). Washington, DC. Office of Research and Development.

EPA (U.S. Environmental Protection Agency). 1998. Guidelines for Ecological Risk Assessment. EPA-630-R-95-002F. Washington, DC. Risk Assessment Forum.

EPA (US Environmental Protection Agency). 2000. Guidance for the Data Quality Objectives Process EPA QA/G-4. EPA/600/R-96/055. Washington, DC. Office of Research and Development.

EPA (U.S. Environmental Protection Agency). 2002. Guidance on Environmental Data Verification and Data Validation EPA QA/G-8 (PDF) (96 pp, 336K, about PDF). EPA/240/R-02/004. Washington, DC. Office of Environmental Information.

EPA (US Environmental Protection Agency). 2009a. Guidance on the Development, Evaluation, and Application of Environmental Models (PDF) (99 pp, 1.6MB, about PDF). EPA/100/K-09/003. Washington, DC. Office of the Science Advisor.

EPA (US Environmental Protection Agency). 2009b. User's Guide and Technical Documentation KABAM version 1.0 (PDF) (123 pp, 870K, about PDF). Washington, DC. Environmental Fate and Effects Division, Office of Pesticide Programs.

EPA (US Environmental Protection Agency). 2009c. Using Probabilistic Methods to Enhance the Role of Risk Analysis in Decision-Making With Case Study Examples (PDF) (92 pp, 721K, about PDF). *DRAFT.* EPA/100/R-09/001 Washington, DC. Risk Assessment Forum.

EPA (US Environmental Protection Agency). 2010. Modeling Journal: Checklist and Template (16 pp, 310K, word doc). New England Office, New England Regional Laboratory's Office of Environmental Measurement and Evaluation, Quality Assurance Unit.

REFERENCES (CONTINUED)

FDEP (Florida Department of Environmental Protection) 2008. TMDL Report Nutrient, Un-ionized Ammonia, and DO TMDLs for Lake Trafford (WBID 3259W) (PDF) (88 pp, 1.8MB, about PDF). Division of Water Resource Management, Bureau of Watershed Management.

Fisher, E., P. Pascual and W. Wagner. 2010. Understanding Environmental Models in Their Legal and Regulatory Context. Journal of Environmental Law 22(2): 251-283.

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

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

NIH (National Institute of Health). 2010. Version Control Guidelines. Bethesda, MD. National Institute of Dental and Craniofacial Research. *Accessed July 13, 2011*.

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.

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.

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

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.

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.

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.

Model 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.

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.

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.

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.

TMDL: A Total Maximum Daily Load, or TMDL, is a calculation of the maximum amount of a pollutant that a waterbody can receive and still safely meet water quality standards.

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.