

## Glossary of Frequently Used Modeling Terms

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

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

**Analytical Models:** Models that can be solved mathematically in closed form. For example, some model algorithms that are based on relatively simple differential equations can be solved analytically to provide a single solution.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

**Deterministic Model:** A model that provides a single solution for the state variables. Because this type of model does not explicitly simulate the effects of data uncertainty or variability, changes in model outputs are solely due to changes in model components.

**Domain (spatial and temporal):** The limits of space and time that are specified within a model’s boundary conditions (see boundary conditions).

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

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

**Environmental Data:** Information collected directly from measurements, produced from models, and compiled from other sources such as databases and literature.

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

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

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

**Expert Elicitation:** a process for obtaining expert beliefs about subjective quantities and probabilities. Typically, structured interviews and/or questionnaires are used to elicit the necessary knowledge. Expert elicitation may also include "coaching" techniques to help the expert conceptualize, visualize, and quantify the knowledge being sought.

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

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

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

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

**Function:** A mathematical relationship between variables.

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

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

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

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

**Measurement Errors:** Errors in the observed data that are a function of human or instrumental error during collection. Such errors may be independent or random. When a persistent bias or miscalibration is present in the measurement device, measurement errors may be correlated among observations. In some disciplines, measurement error may be referred to as observation error.

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

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

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

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

**Model:** A representation of the behavior of an object or process, often in mathematical or statistical terms. Models can also be physical or conceptual.

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

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

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

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

**Objectivity:** One of three main components of quality in EPA's *Information Quality Guidelines*. Objectivity includes whether disseminated information is being presented in an accurate, clear,

complete and unbiased manner. In addition, objectivity involves a focus on ascertaining accurate, reliable and unbiased information.

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

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

**Parametric Variation:** When the value of a parameter itself is not a constant and includes natural variability. Consequently, the parameter should be described as a distribution.

**Parameter Uncertainty:** See Data Uncertainty.

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

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

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

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

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

**Quantitative Assessments:** The uncertainty in some sources—such as some model parameters and some input data—can be estimated through quantitative assessments involving statistical uncertainty and sensitivity analyses. In addition, comparisons can be made for the special

purpose of quantitatively describing the differences to be expected between model estimates of current conditions and comparable field observations.

**Quality:** A broad term that includes notions of integrity, utility, and objectivity.

**Reducible Uncertainty:** Uncertainty in models that can be minimized or even eliminated with further study and additional data. See [data uncertainty](#).

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

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

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

**Sensitivity:** The degree to which the model outputs are affected by changes in a selected input parameters.

**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. By investigating the “relative sensitivity” of model parameters, a user can become knowledgeable of the relative importance of parameters in the model.

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

**Simulation Models:** Simulation models are used to obtain solutions for more models that are too complex to be solved analytically. In general, simulation models provide approximations of the mathematical solutions. For most situations, where a differential equation is being approximated, the simulation model will use finite time step (or spatial step) to “simulate” changes in state variables over time (or space).

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

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

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

**Statistical Models:** Simple linear or multivariate regression models obtained by fitting observational data to a mathematical function.

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

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

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

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

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

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

**Utility:** One of three main components of quality in EPA’s Information Quality Guidelines. Utility refers to the usefulness of the information to the intended users.

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

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

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