Protocols for Water and Environmental Modeling

Draft Final: April 7, 2021

CWEMF
California Water & Environmental Modeling Forum

Project Team

This work was guided by the Modeling Protocols Committee of the California Water and Environmental Modeling Forum (CWEMF). Committee members are: Rich Satkowski (California State Water Resources Control Board, Retired; Committee Chair), Jamie Anderson (California Department of Water Resources), Will Anderson (Contra Costa Water District), Mike Deas (Watercourse Engineering), John DeGeorge (Resource Management Associates), Ben Geske (Delta Stewardship Council), Tariq Kadir (California Department of Water Resources), Josué Medellín-Azuara (University of California, Merced), George Nichol (California State Water Resources Control Board, Retired), Nicky Sandhu (California Department of Water Resources), Tad Slawecki (LimnoTech), Ali Taghavi (Woodard & Curran), and Chuching Wang (Metropolitan Water District of Southern California).

The consultant team at Tetra Tech consisted of Sujoy Roy, Paul Hutton, Katherine Heidel, John Rath, and Arushi Sinha.

Acknowledgements

Funding for the consultant team was provided by the California Water and Environmental Modeling Forum (CWEMF). This work benefited from a series of focused meetings with practitioners in early 2020 to discuss the state of modeling in various water resource-related disciplines. Meeting participants are identified in Appendix A. This work also benefited from related modeling documents recently produced by the California Delta Stewardship Council. Bethany Robinson and Jonathan Herman of the University of California at Davis provided a summary of literature and case studies under an initial phase of this work.

Contents

Figures

Tables

Foreword

In 2000, the Bay-Delta Modeling Forum (the predecessor organization of CWEMF) developed the document "Protocols for Water and Environmental Modeling," which provided protocols and guidelines for the development and use of mathematical models for water planning and management in California. This document is an update, more than 20 years after the original, reflecting changes in the practice of water and environmental modeling. These changes have been driven by the nature of questions being asked today, technological advancements in hardware and software tools for modeling, and the increasing role of stakeholders and decision-makers in the modeling process. This update of the 2000 document reflects the current practice of modeling by addressing these issues.

Mathematical modeling is a central part of the decision-making framework for virtually all water resources questions addressed in California today. CWEMF believes that acceptance and implementation of modeling protocols by California's water community will result in better and more defensible models and modeling studies by:

- Improving the performance and reliability of models;
- Providing better documentation of models and modeling studies;
- Providing easier professional and public access to models and modeling studies;
- Making models and modeling studies more easily understood, transparent, and amendable to examination and reproduction by others; and
- Increasing confidence in models and modeling studies.

CWEMF accepted "Protocols for Water and Environmental Modeling" on [DATE] and is assisting CWEMF members and other interested parties in implementing the modeling protocols. Since this report is a "living document," it will be updated periodically, as the need arises. The authors recommend that CWEMF reconvene its Ad Hoc Modeling Protocols Committee at least once every three years to ascertain whether a partial or full update is needed. As specified in CWEMF's bylaws, it should be noted that this report does not necessarily represent the views of the governing bodies of the represented organizations or the individual members of CWEMF.

Executive Summary

Mathematical modeling using computers—comprising a variety of mechanistic, statistical, optimization-based and other emerging techniques—has become indispensable for managing water in California. Such modeling is used for a variety of essential tasks, including supporting compliance with regulations, managing water rights, planning for future changes due to growth and climate change, designing of new infrastructure and planning for environmental restoration. Only through the use of computer simulations can the large amounts of data and the complex interactions involving water, ecosystems, and human systems be adequately understood and predicted. Thus, major water-related projects are rarely undertaken in California without the support of some model-based analyses. This dependence on models raises important questions about quality control among stakeholders and decision-makers.

This work builds on a prior set of modeling protocols, developed in 2000 by the Bay-Delta Modeling Forum (the predecessor organization of CWEMF), and reflects changes in the practice of modeling, key technological developments, and applications addressing problems relevant today. These protocols are intended to serve modelers, i.e. technical specialists who develop and/or run models, as well as the broader community of model sponsors and stakeholders, who have an interest in the quality and reliability of a modeling study.

This document describes the modeling approaches used to address water resources problems, including analytical/numerical models that simulate solutions to singular and/or integrated physical processes over a defined domain; statistical/empirical models that are based on relationships between observed data but typically contain little to no process representation; optimization-based models that seek to meet key objectives subject to a set of defined constraints; machine learning-based models, a sub-class of statistical/empirical models with a wider range of algorithms and capacity to handle disparate data sets; and agent-based models that represent behavior of organisms or populations (animal or human) in response to external factors. The focus of these protocols is on general approaches and methodologies that apply to different model domains, rather than on specific models.

The term "model" is used throughout this document as a general term for a quantitative and simplified representation of a physical system. However, a further distinction is made through the use of model framework, another general term for the theoretical implementation of a process-oriented model. A model framework will usually need to be configured for application to different geographic settings, and is termed a model application. Model applications are a greater, although not the only, focus in this protocol document.[1]

This work provides a summary of actions to be taken at various stages of the modeling process, divided into the four broad phases (Figure ES-1), including:

- Preliminary analyses,
- Framing the modeling study,
- Application of the model, and

[1] For example, the Central Valley Hydrologic Model (CVHM) is a model application of the MODFLOW model framework. See Appendix B for more examples of model frameworks and model applications.

- Communicating and documenting results.

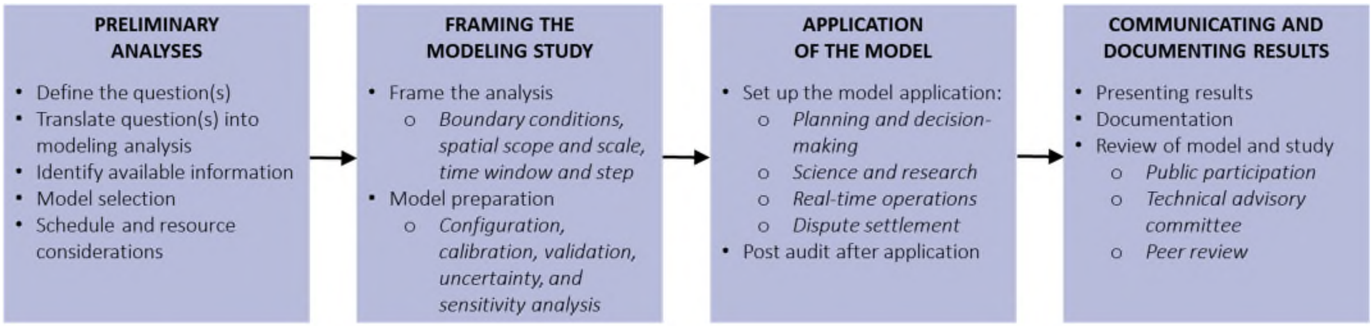

Figure ES-1. Visual representation of the phases of a typical modeling study.

Preliminary analyses begin with a clear statement of the question to be answered through modeling, and assess whether a model is, in fact, a useful tool to address the problem. If so, the study participants should determine what type of model would be most useful, and then whether an existing model can be used for this purpose or whether a new model needs to be developed. The level of model complexity chosen is often informed by the scale and difficulty of the problem to be addressed, budget and schedule considerations, and the type and amount of available data. A clear statement of the goals of the model study, best developed in a transparent, open, and collaborative environment with stakeholders, will establish the long-term vision of model development to most efficiently and best serve the needs of the project. Based on this evaluation, a suitable model may be selected. Early thinking about these processes and communicating the path forward with other study participants will avoid surprises during the performance of the study.

Framing the modeling study refers to the process of setting up and running the model and obtaining useful results. The first step in this phase involves defining the analysis, where important details are addressed such as the geographic and temporal scope of the study, setting problem domain, boundary and initial conditions available to run the model, the questions to be asked which lead to type of scenarios to be evaluated, the time frame of the analysis, and whether or not multiple models need to be integrated for the analysis. The next step involves preparation and operation of the model, which are often considered to be the key activities that modeling includes, such as configuration of the model with background properties, compiling the data needed for modeling, and calibrating the model. Model calibration is an essential step in most environmental models, and includes the modification of adjustable parameters to achieve reasonable match between the model generated results and observed and/or reported data. Once calibrated, a model is evaluated against an independent set of data, to assess its general applicability. Often the term validation is used for this step in some scientific literature. In many cases, model validation may result in additional adjustments to the model parameters, which leads some to considering this step as part of the calibration process. In many cases, additional analyses may be performed to better understand the overall behavior of a model, including an uncertainty analysis which attempts to describe model behavior with inputs and parameters varied to represent imperfect knowledge, and sensitivity analysis, which explores the sensitivity of the model to individual inputs and parameters (uncertainty and sensitivity analysis are related concepts).

Application of the model is performed to ask the questions it was designed for, which often involve conditions that are different from what is directly observable today. This phase should occur only after the model has been successfully evaluated, thus establishing the credibility of the model. The relevant decisions to be made and processes to be followed during the model application phase depend on the four unique classifications of applications: i) planning and decision support, ii) science support and research, iii) real-time operations support, and iv) dispute settlement support.

Communicating and documenting results is the final phase of modeling and includes obtaining peer reviews, an element that is especially important for water resources and environmental models because such modeling is often driven by stakeholder interest. Good practices in these areas, notably the allocation of enough time for these tasks to be performed adequately, are described.

These protocols present an overview of two additional related areas. The first area includes a review of a variety of emerging technologies that are being embedded in modeling, and have the prospect of changing the future practice of modeling. The second addresses long-term beneficial activities around specific modeling themes that transcend the concerns of an individual study. These include management activities around the life cycle of a model and the development of a community around a model or a problem theme.

Many of the concepts identified in this document are perhaps known to most modelers but are not consistently adopted. This may be due to time and resource limitations associated with virtually all modeling studies; this may also be due to the lack of specific expectations in the broader community of modelers and model users. Thus, model users may not know what specific and reasonable requests to make of modelers to guide a model study toward greater credibility and usefulness. Toward this end, this work concludes with specific outreach actions to model specialists, model sponsors, general stakeholders who may be non-modelers, and to students and non-modelers.

To succinctly describe the concepts identified in this work and encourage adoption of these best practices, we provide three summary checklists, corresponding to different phases of modeling. Checklist 1 is designed to be employed at the inception of a modeling effort, and to enable various participants to agree on the basic features of the work to be done. The purpose of Checklist 2 is to evaluate and score a modeling exercise upon completion. Checklist 3 is useful for assessing the management approach for the overall life cycle of a modeling framework.

Checklist 1. Model Study Initial Appraisal Prior to Study Inception

Item	Description	Response
1	Is the problem or question to be addressed well defined?	Yes/No
2	Do we know how the model results will be used?	Yes/No
3	Is the model to be used specified?	Yes/No
4	Has a conceptual framework been developed?	Yes/No
5	Have the criteria for selecting the model been defined?	Yes/No
6	Is an existing model going to be modified?	Yes/No
7	Is a new model to be developed?	Yes/No
8	Are the time frames known for initial model development, calibration, testing, and review?	Yes/No
9	Are data associated with intended model inputs available?	Yes/No
10	Does the model need calibration?	Yes/No
11	Are data associated with intended model outputs available (to support model calibration)?	Yes/No
12	Are time frames of the input and output data known and consistent with one another?	Yes/No
13	Are the errors in data measurements known?	Yes/No
14	Is the level of error in the expected results known?	Yes/No
15	Are the model stakeholders known?	Yes/No
16	Will stakeholders be part of the modeling process?	Yes/No
17	Have users of the model output met together?	Yes/No
18	Will documentation be prepared upon completion of the model?	Yes/No
19	Will the information embedded in the above questions be used to prepare a memo describing the model's purpose?	Yes/No

Checklist 2. Model Study Post-Completion Appraisal

Item	Description	Response (Numeric Score or narrative)
1	Is the model a new formulation or the application of an existing code? If a new formulation, what has been done to test and verify the code?	
2	Has a conceptual framework been developed for this effort and has it been updated following completion?	
3	Are observed data used in the modeling exercise (input and output data) documented and available for review?	
4	Has the calibration approach been described?	
5	Has the model performance following calibration been adequately evaluated using test data?	
6	Has the sensitivity of major variables been evaluated?	
7	Has model output uncertainty been evaluated?	
8	Were any novel approaches used to evaluate the sensitivity and uncertainty of the model response to inputs?	
9	Were the model results compared and contrasted with other models (if available)?	
10	Does the model study documentation adequately explain the approach, assumptions, and findings?	
11	Was a peer review performed and responded to?	
12	What were the stakeholder's reactions to the model results?	
13	Are the model summary documents easily understandable by a variety of audiences?	

Checklist 3. Model Framework Life Cycle Evaluation

Item	Description	Narrative Response
1	Are all source codes and supporting files stored in a single location and archived in a manner that enables future access?	
2	Are the source codes documented, even if this documentation is not in the public domain?	
3	Is the model development dependent on a single individual? What is the long-term transition plan for the expertise in this model?	
4	Is the model framework applied by a community or by a single team? Is there a mechanism to share knowledge about the model application over time, such as a virtual community, trainings, etc.?	
5	Is there a defined plan for making updates to the model framework?	
6	For a public-domain model framework, is there a funding mechanism to support staff that would work on the model?	
7	For a proprietary model framework, what is the mechanism to support the code development over the long-term?	

Glossary of Terms and Acronyms

Term	Definition
Aleatory uncertainty	Aleatory or stochastic uncertainties originate from inherent variability and stochasticity of natural phenomena (e.g., climatic variability). Aleatory uncertainties cannot be reduced by collection of more data.
All-at-a-time (AAT)	A sensitivity analysis approach where all parameters can be varied at each iteration. This approach is typically used with global sensitivity analysis.
Application Programming Interface (API)	Software intermediary that allows two applications to communicate with one other.
Boundary condition	A condition that is required to be satisfied at all or part of the boundary of a region in which a set of differential equations is to be solved.
Calibration	The process of changing values of model parameters to achieve the best match or "fit" of the model results to field observations or reported values.
Code	Representation of the theoretical formulation of a model in computer language that serves as the basis for developing an executable model. In many cases, even for public-domain models, the underlying codes are not in the public domain.
Code verification	The process of testing the accuracy of the model's computer representation of the theoretical formulation. This process includes code examination, testing bounding cases, and comparison against analytical solutions of underlying equations (when available).
Conceptual framework	Often referred to as the conceptual model. In this document, the term conceptual framework is used to avoid confusion with other uses of the word 'model'. A high-level representation of inputs, interacting physical processes to be modeled and the drivers, and outputs for any kind of process (e.g., physical, biological, economic, etc.). Although a conceptual framework may include quantitative information, it is often presented in non-quantitative form and serves to communicate the nature of system to be modeled and model structure in a transparent manner. A conceptual framework may be developed as a communication tool following the completion of a modeling study, or, during the initiation of the project, the conceptual framework can serve as the basis for selection of or development of a quantitative model.
Crowdsourced data	Data obtained for a particular task or project by enlisting the services of a large number of people, either paid or unpaid, typically via mobile devices.
DevOps	This term refers to processes for version control, continuous integration, artifact management, automated testing, continuous delivery, and system monitoring that work together to both reduce the time to develop and deploy software and to improve reliability.

Term	Definition
Distributed parameter model	Provides greater spatial and/or temporal resolution. A lumped parameter model, in contrast, aggregates variable information over time and/or space for simplification or to remedy limited data availability.
Domain	In this work, a specialized field of study.
Driver	A model process or parameter that has the greatest impact on key model outputs.
Empirical/statistical model	A mathematical formulation of inputs and outputs with limited process representation. Model parameters are typically calibrated with observed data.
Emulator	A computationally simplified model representation that uses relationships between inputs and outputs. Emulators are typically developed to reduce the computational cost of model exploration.
Epistemic uncertainty	Epistemic uncertainties stem from our lack of knowledge and they can be reduced with additional collection of data.
Equifinality	Non-uniqueness of a model fit, i.e. when multiple parameter combinations can provide equivalent fits during the calibration process. This may occur with complex models with numerous fitting parameters.
Evaluation	A general term for a sequence of steps taken to understand the performance of a model following calibration. Evaluation may include comparison against independent input and output data sets, sensitivity analysis for key parameters, or uncertainty analysis.
Global Sensitivity Analysis (GSA)	A sensitivity analysis approach that analyzes the variability of model responses across the full parameter space.
Initial condition	The solution of a differential equation or iterative process over time requires the definition of values at the inception of the solution, termed the initial conditions. Other types of formulations, such as time series models, may also require initial conditions.
Local Sensitivity Analysis (LSA)	A sensitivity analysis approach that analyzes model responses around a well-defined region of interest in the input parameter space.
Lumped parameter model	A model that aggregates variable information over time and/or space for simplification or to remedy limited data availability. A distributed model, in contrast, provides greater spatial and/or temporal resolution.
Metadata	A set of data or narrative descriptors for a data set. Often termed "data about data."
Model configuration	The process of specifying background characteristics for a model simulation, e.g. the physical representation of a water body. Model configuration is performed once the theoretical framework of a model has been developed and implemented.

Term	Definition
Model application	A model framework configured for application to a specific setting.
Model framework	A general term for the theoretical implementation of a process-oriented model. A model framework will usually need to be configured for application to a specific geographic setting. Many models in common use are general purpose frameworks that can be configured to represent the same set of processes in different regions (for example, watershed models), whereas others are developed from the ground up as applicable to a single location, and the configuration is embedded within the general setup. Other terms that might be used here include generalized model software, model executable program, computational engine, or simply, model. However, in this document we do not use the term "model" alone because it is not specific enough.
Model integration	An approach where two or more models, typically with different areas of focus, are used together in an integrated framework, such that certain information and data may be exchanged between the models to provide analysis across multiple domains.
Model life cycle	A term referring to the entire timeframe from conceptualization of a mathematical model to implementation in computer code and to multiple cycles of application, revision, and reuse in one or many different domains. Models of complex environmental systems generally require large investments and a lifecycle of many decades.
Model structure	The representation of model inputs, key processes and interactions, and outputs. A conceptual framework may graphically communicate the model structure, but even where a conceptual framework is not available, all process-based models require an underlying model structure. In the case of empirical models, internal processes are generally not represented, and model structure refers to the inputs that are selected *a priori* to influence the outputs.
Model training	Similar to model calibration and parameter estimation, but typically used in the context of machine learning. The process of adjusting empirical model constants to match model outputs and field observations. In the context of machine learning, the model constants may have no physical meaning.
Monte Carlo simulation	A general solution approach in modeling analysis where key values (e.g., parameter values in a model) are sampled randomly over a defined space to provide a range of conditions for testing.
Numerical model	Many quantitative models are represented by mathematical formulations, including partial differential equations that cannot be solved exactly (i.e., analytically) because of spatial and/or temporal extent or mathematical complexity. Numerical techniques (e.g., finite elements or finite differences) are commonly used approaches to estimate solutions to partial differential equations. Models that employ such numerical solutions are particularly common in the representation of physical and chemical processes and are termed numerical models.

Term	Definition
One-at-a-time (OAT)	A sensitivity analysis approach where one parameter is changed at each iteration. Commonly used with local sensitivity analysis.
Open source model	Open source models are those where the underlying source code of the model is available for anyone to examine and modify, potentially creating a new executable version of the model.
Parameter estimation	Similar to calibration. The process of adjusting parameter values in a model such that the model output matches field observations within an acceptable error range. Some parameter values may be obtained from independent experiments, in addition to model runs.
Parameters	Numeric constants associated with key processes that typically represent a natural system feature (e.g., reaction rates or hydraulic conductivities). Parameter values may be known to lie within an expected range. The process of parameter estimation is to find a value or set of values that enables the model to fit observed or reported data within an acceptable range.
Peer review	A process where independent outside experts evaluate a modeling exercise, including the model code, framework, data, calibration process and results, and application.
Probabilistic form	A distribution of values based on a statistical method, as opposed to a single value.
Proprietary model	Proprietary models are owned by a non-public entity, the code and application data may not be available for review, and there may be a cost for leasing and applying the model.
Public-domain model	Public-domain models are those where the executable version of a model is available for use, although the source code may not necessarily be available.
Quantitative	Numerical or calculable, as opposed to qualitative. The quantitative model can be described by its mathematical approach. The conceptual framework is often presented in non-quantitative form and serves to communicate the nature of system to be modeled and model structure in a transparent manner.
Sensitivity analysis	The process of adjusting model parameters or inputs within a realistic range to explore the effect on, or sensitivity of, model outputs. Model sensitivity in a multi-parameter model may depend on the states of other parameters, and individual model outputs may be more or less sensitive to different parameters. A common goal of sensitivity analysis is to identify parameter(s) that have the greatest impact on key model outputs. See also: AAT, GSA, LSA, and OAT in this table.
Stakeholder	Stakeholders are participants who have an interest in the outcome of a modeling study.
Statistical model	See “Empirical/statistical model” above.

Term	Definition
Uncertainty analysis	Model inputs or parameter values are presented in a probabilistic form (i.e., as a distribution of values) to a calibrated model, and the effects on model output evaluated. Given that model inputs and parameters typically and inherently include different degrees of error or uncertainty, the goal of uncertainty analysis is to quantify the range of outputs that reflect the range of errors or uncertainties in the model input data or parameters in a modeling study.
Validation	A set of steps used to demonstrate that, within its domain of applicability, a model possesses a satisfactory range of accuracy consistent with the intended application of the model. In more limited settings, particularly in the context of machine learning, validation refers to the process of applying a fitted model to an independent set of observed data to evaluate model fit.

1 Introduction

Computer modeling—comprising a variety of mechanistic, statistical, optimization-based and other emerging techniques—has become indispensable for managing water in California and elsewhere. This water could be on the land surface, in rivers, lakes, estuaries, underground, or runoff from watersheds. Such modeling is used for a variety of essential tasks, including compliance with regulations, managing water rights, planning for future changes due to growth and climate change, design of new infrastructure and for environmental restoration. Only through the use of computer models can the large amounts of data and the complex interactions involving water, ecosystems, and human systems be adequately understood, and the effects of change be estimated. Thus, major water-related projects are rarely undertaken in California without the support of at least some model-based analyses. This dependence on California's water and environmental models raises unavoidable, but healthy questions of quality control among the diverse water stakeholders and decision-makers.

In 2000, the Bay-Delta Modeling Forum (the predecessor organization of CWEMF) developed the document "Protocols for Water and Environmental Modeling," which provided protocols and guidelines for the development and use of models for water planning and management in California. This document is an update, approximately 20 years after the original, reflecting changes in the practice of modeling in the area of water resources and the types of questions being asked, with a new focus on climate change and new ways of addressing groundwater sustainability. While many of the essential features of modeling have evolved gradually over this time, in certain technological areas there has been more significant change, including advances in geospatial data availability, online data and code sharing, cloud-based computing, the use of smart phones as computer platforms, machine learning, sensor-based data sources and telemetry, and data visualization. Furthermore, large-scale modeling exercises may be viewed as significant investments, to be sustained and preserved in a directed manner, much like other assets. Finally, there have also been cultural shifts related to modeling in the broader community. More specifically, a greater recognition of the need for collaborative modeling with stakeholder participation exists, as opposed to modeling being performed as a technocratic exercise directed by experts. This update of the 2000 protocols document reflects the current practice of modeling by addressing these issues.

This document is written to serve the needs of model developers as well as model users who may sponsor and direct studies and may need to make decisions based on model results. This document is also relevant to stakeholders who are affected by model outcomes and subsequent decisions and need more transparency into what is reported in a model study.

1.1 What are Models?

Within this document, a model is defined as a quantitative representation of a real-world system comprising physical, chemical, biological, economic, and social systems. These various systems often interact; thus, an individual model may encompass more than one system. Although economic and social systems are represented by varied modeling frameworks, for these disciplines our focus is on models where the natural environment is a major focus. Models can be used to ask "what if" questions—e.g., what might happen if a particular project is built or if a regulation is implemented—and are especially important for water resources and environmental problems over large spatial scales because no practical way exists to answer these questions through an experimental approach.

1.1.1 Terminology

Even though the term "model" is used throughout this document as a generic term for a quantitative representation of a system, further distinction is important (**Figure 1**). Typically, at the core of a model is an underlying theory or a conceptual framework (or a conceptual model) representing a process or set of interacting processes. Some processes, especially physical processes, are conceived in mathematical forms, whereas others (such as biological processes) may first be proposed in narrative form and then translated into a mathematical form. The mathematical representation needs to be solved, using different analytical or statistical methods, and is implemented through computer code. A model framework is a general term for the computer implementation of a theoretical process-oriented model. A model framework is a general tool; configuration refers to the incorporation of numeric values from a specific geographic setting. In this form it is termed a model application, and often model parameters are selected for a specific setting using a calibration and validation process. Many commonly available models are in fact general purpose model frameworks that can be configured to represent the same set of processes in different regions (for example, watershed models or groundwater hydrology models). In some cases, models are developed initially as an application for a single region, and the configuration is embedded within the general setup. In these instances, the model framework and application are one and the same. Typically, models that deal with problems that represent physical and chemical processes often have general purpose model frameworks, whereas models in fields such as biology or economics are more commonly developed initially to serve a specific need. The term model study typically refers to a specific set of scenarios or questions that are addressed with a model application. Some other important steps that occur during modeling, include peer review and preparation of supporting documentation, which can occur from the inception of a modeling effort to the completion of model study results. These terms are described in further detail in the rest of this document.

Another commonly-used term is code, which refers to the computer instructions that underlie a model framework. The term simulator is also used and, depending on context, may be a reference to a model framework or an application.

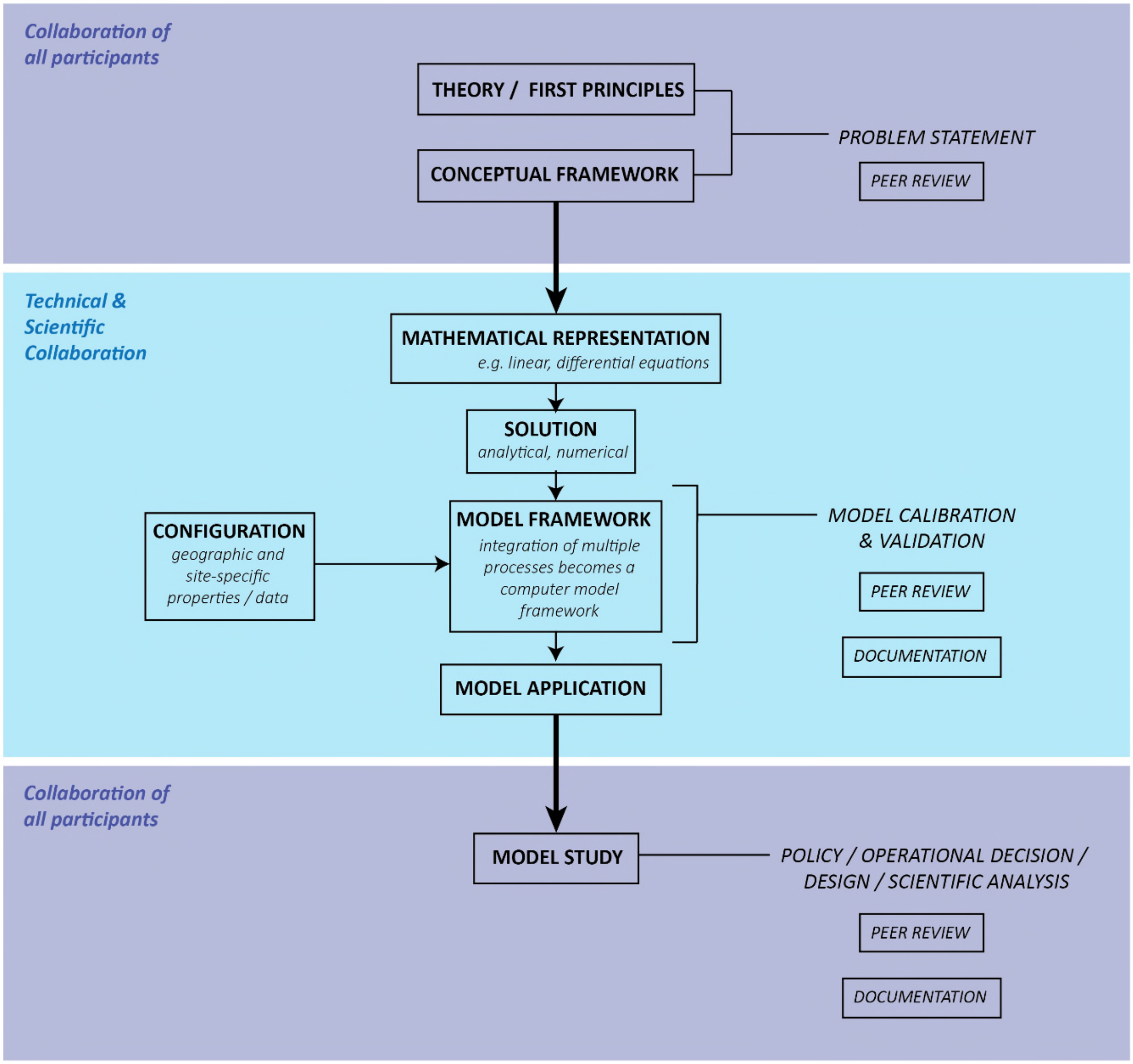

Figure 1. Visual representation of common terms used in this document.

1.1.2 Types of Models

Several different mathematical approaches may be applied in the development of quantitative water and environmental models as shown in **Table 1**. The broad classes of mathematical approaches in use include analytical/numerical solutions of process equations over a defined extent; statistical/empirical models that are based on relationships between observed data but typically contain little to no process representation; optimization based models that seek to meet key objectives subject to a set of defined constraints; machine learning based models, a sub-class of statistical/empirical models with a wider range of algorithms and capacity to handle disparate data sets; and agent-based models that represent behavior of organisms or populations (animal or human) in response to external factors. Several of these approaches may be combined within a single modeling system, resulting in a "hybrid" model. As described in the following chapters, the underlying approach adopted within a particular modeling framework affects the applicable best approaches for development.

Table 1. Typical Model Types Used for Water and Environmental Modeling

Model type	Feature
Simulation Models	Simulates the state of the system by solving a framework of process equations, either in closed analytical form or numerically; model parameters are typically calibrated with observed data.
Optimization Models	Optimizes a set of key objectives that define the state of the system under a range of input conditions, and subject to a set of constraints.
Statistical/Empirical Models	Represents limited processes; model parameters are typically calibrated with observed data.
Machine-learning Models	Trained to find patterns or relationships in available data, but with minimal process-oriented representation. These are an extension of the statistical/empirical models, but with a greater variety of emerging algorithms to represent increasingly complex data sets.
Agent Models	Represents individual behavior of organisms or populations (animal or human) in response to external factors over time and space.

1.1.2.1 Simulation Models

Simulation models are typically based on analytical and/or numerical solutions to the mathematical representation of the system. . Analytical models normally consist of closed-form solutions to differential equations and have been used for relatively simple domains combined with a need for efficiency. Numerical models, which solve differential equations over space and/or time, are intended to simulate the state of the singular and/or integrated system that are defined by the mathematical representations over time and space. Numerical models are in widespread use, especially in the water flow and water quality domains. In recent years, such models have tended to grow more complex, with greater spatial and/or temporal resolution, and associated computational demands. For simplified domains, analytical solutions are important for testing the computer implementation of numerical models (termed model verification), which are prone to solution errors.

1.1.2.2 Optimization Models

Optimization models are intended to optimize an objective function (such as maximizing profit or yield of the system) subject to certain constraints, such as physical, hydrologic, operational, economic, climatic, or infrastructure. For hydrologic systems, the outcomes are the water allocations to different users constrained by water availability, environmental flow requirements, operating rules, and the hierarchy of water rights.

1.1.2.3 Statistical/Empirical Models

Statistical/empirical models are usually based on development of statistical relationships among the various observed or recorded data sets with limited underlying process representation. Larger datasets often improve performance of statistical models.

1.1.2.4 Machine Learning Models

Machine learning models, a class of statistical/empirical models, offer a wide variety of emerging algorithms to find patterns or relationships in observed data or model output. Unlike most statistical/empirical models, machine learning models may contain large numbers of fitting parameters that result in complex statistical and/or mathematical relationships, which may not be evident to the end user, and would typically require special expertise to discern and interpret.

1.1.2.5 Agent Models

Agent models represent a system with agents (e.g. organisms, individuals, or households) that have individual behavior and respond to external drivers or to each other. Fish behavior models are a typical class of models that use the agent-based formulation.

1.2 Typical Uses of Models

Commonly used models for water and environmental applications in California support a variety of applications that can be broadly classified as follows:

- Planning, management and decision support – including, but not necessarily limited to, support for the development of new environmental regulations (e.g., changes to water quality standards, or water supply regulation), feasibility analysis of facility or operational modifications (e.g., changes to reservoir operating rules), assessment for design of new infrastructure (e.g., new alternatives for Delta conveyance or evaluation of new dam sites, or evaluation of new groundwater facilities), and evaluation of effects of climate change on the performance of the system.
- Science support and research – including the generation and testing of hypotheses to better understand a particular system, comprising natural and/or human elements. Science support activities include understanding the population behavior of key species, food web interactions, and changes in landscape over the long-term due to human pressures, climatic change, and extreme events.
- Real-time operations support – including reservoir outflows for flood management and water supply, water exports from the Delta, barrier operations used to manage salinity at various locations, or drawdown effects at a well field or movement of contamination plumes as a result of specific groundwater operations.
- Dispute settlement support – including legal proceedings in the context of water rights adjudication or allocation of water among different types of beneficial uses.

Each of the applications described above are associated with decisions wherein modeling provides fundamental information as a basis for potential decision and/or action. Since many of these decisions have large consequences to human communities and ecosystems, it is important to ensure that the models used are credible.

1.3 Individual and Institutional Roles in Modeling

These modeling protocols were developed to serve the needs of a variety of participants who are involved in directing, executing, and evaluating the outcomes of a modeling study, including the model sponsor, model specialist, domain experts, decision makers, and stakeholders. **Figure 2** shows the relationship between these participants. Usually, these participants belong to different institutions or organizations with different areas of interest and expertise. A model **sponsor** is an organization or group of organizations that have an interest in the outcome and provide the resources for the modeling work. The sponsor will likely define the scope of the model study, including question(s) to explore, scenarios of interest, schedule, funding, etc.

The actual development, testing, and reporting of a model study will likely be performed by **model specialists** with knowledge of the specific domain and with relevant software development skills. In some cases, the model specialist may rely on a **code developer** to actually write the computer code to meet the needs of modeling. One or more **domain experts**, who are often the same as the model specialists, may help with interpreting and communicating model results to the sponsor and stakeholders. In some cases, the sponsor and other decision makers will not work directly with model specialists or model results.

Finally, **stakeholders** with an interest in the outcome of a study may guide the process through the model sponsor or the decision makers. Indeed, with the growing application of models in many areas of decision-making, it is desirable to engage and enable stakeholders to play a larger role in modeling studies in a collaborative framework, both in California and more widely (Voinov and Bousquet, 2010; Voinov et al., 2016). See also the discussion of shared-vision modeling in Section 6.3.3 below.

Modeling studies that are focused on scientific advancement are often led by research teams and involve fewer participant roles than shown in **Figure 2**. Although such modeling studies may not be used directly by stakeholders and decision makers, they serve two roles: (i) the models mature over time and drive larger scale policy decisions, as described in the next chapter, or (ii) the individual expertise gained through model development gradually diffuses into the broader modeling community.

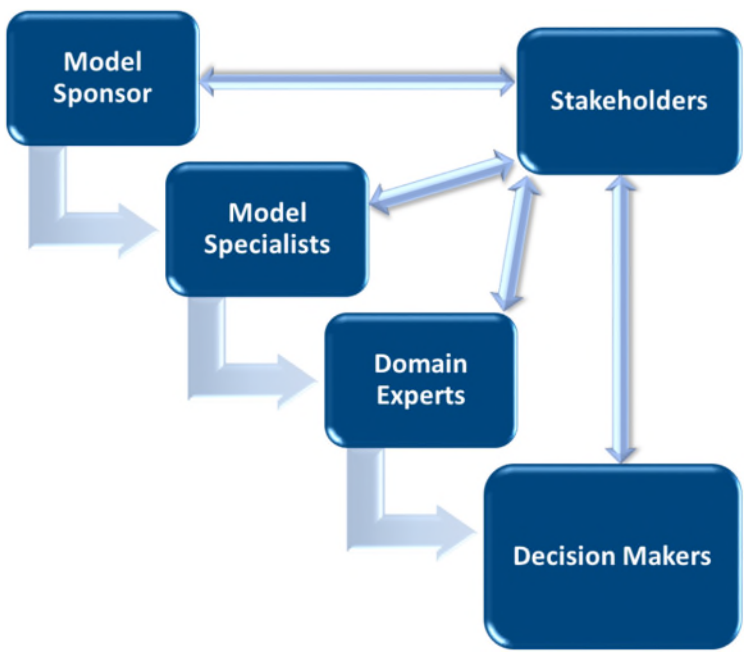

Figure 2. Key roles in modeling studies.

1.4 Other Published Modeling Protocols

As the science of water related modeling has matured and become more widely used, best modeling practices that emphasize particular modeling best practices for model development and application have been proposed. General guidance on water and environmental modeling that was published over the past two decades is presented in **Table 2**. Other domain-specific guidance has been developed and embedded within reviews of modeling studies in different disciplines, e.g., coastal and estuarine models (Ganju et al., 2016; Dawson et al., 2019); watershed models (Daniel et al., 2011); models for total maximum daily load (TMDL) development for water quality constituents (Shoemaker et al., 1997); models of nutrient behavior in aquatic systems (Trowbridge et al., 2016); and groundwater models developed for the Sustainable Groundwater Management Act (SGMA) (DWR, 2016). This document is informed by these published guidelines. Elements from these prior guidelines are cited, as appropriate, throughout the following chapters. Some agencies, such as the U.S. Army Corps of Engineers (USACE), have formal processes for auditing and certifying model results performed under their direction.

Table 2. Examples of Prior General Guidance for Water and Environmental Modeling

Year of Publication	Title	Author(s)	Focus
2000	*Protocols for Water and Environmental Modeling*	California Water & Environmental Modeling Forum (formerly Bay-Delta Modeling Forum)	Guidance on modeling protocols for the Bay-Delta
2002	*Guidance for Quality Assurance Project Plans for Modeling*	U.S. Environmental Protection Agency	Recommendations on how to develop a Quality Assurance Project Plan (QAPP) for projects involving model development or application
2006	*Ten Iterative Steps in Development and Evaluation of Environmental Models*	A. J. Jakeman, R. A. Letcher, and J. P. Norton	Widely cited general guidance on good practices
2007	*Models in Environmental Regulatory Decision Making*	National Academy of Sciences	General guidance on best practices in model use in complex regulatory settings
2008	*Good Modeling Practice*	N. Crout et al.; Chapter in book on Environmental Modeling, Software, and Decision Support, Jakeman et al., Eds.	General guidance on model development, application, and testing
2009	*Guidance on the Development, Evaluation, and Application of Environmental Models*	Gaber et al, 2009 (U.S. Environmental Protection Agency, Council for Regulatory Environmental Modeling)	General guidance on environmental models, considering both technical and institutional aspects

Year of Publication	Title	Author(s)	Focus
2012	*Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification*	National Research Council	Report with a technical focus on analysis approaches for evaluating complex scientific and engineering models
2016	*Best Management Practices for the Sustainable Management of Groundwater: Modeling*	California Department of Water Resources	Guidance for the development and use of groundwater and surface water models for the Sustainable Groundwater Management Act (SGMA)

1.5 Workshops to Elicit Insight from Practitioners

To elicit information and guidance in the current practice of modeling, five targeted meetings of modeling experts in different disciplines were held in February and March 2020. These meetings covered the following areas: (1) Hydraulics, Hydrodynamics and Water Quality, (2) Groundwater and Integrated Surface Water/Groundwater, (3) Biology and Ecosystems, (4) Surface Watershed Hydrology and Reservoir Operations, and (5) Hydro-Economic Modeling and Economic Analysis. An early outline of this protocols document was provided to the meeting participants to gather input on their approach to modeling tasks in their particular domains. Participants in these workshops are identified in Appendix A. This work is informed by the experience of practitioners captured through these workshops, the general literature cited in **Table 2**, as well as other specialized research studies.

1.6 Motivation for Current Work

The goal of this document is to provide guidance and best practices to water stakeholders, decision makers, and their technical staff as models are developed and used to solve California's water and environmental problems. This modeling protocols document is not a rigid methodology for modeling, but instead serves as a guideline to those who choose to use it. The protocols address new model development, structured applications of existing model frameworks, and activities pertaining to the management of a model over its life cycle, which can potentially span years to decades. The motivation for this work and its utility to the modeling community may be summarized as follows.

Establishing Credibility: There is an adage generally attributed to statistician George Box that "All models are wrong but some are useful."[2] This simple statement recognizes that natural and human systems are sufficiently complex that mathematical representations do not capture their full range of behaviors. But as a participant in a model study (such as those in **Figure 2**), how is one to know which models are useful? At a minimum, we interpret this to mean that a model should be credible and that it reproduces processes with a sufficient level of certainty that the model can be trusted by study

[2] Although this quote is generally attributed to the statistician George Box, no specific citation is available. See discussion at https://en.wikipedia.org/wiki/All_models_are_wrong.

participants as being appropriate for the problem at hand. These modeling protocols should assist model users and decision-makers in making informed judgments regarding model credibility.

Context for Non-Modelers: These protocols are intended to inform the larger modeling community (including model sponsors and stakeholders) so that study results can be reviewed in an inclusive and comprehensive manner. Users of model results, many of whom are not model specialists, are confronted with model outputs that may not be in an audience-appropriate format. The protocols can provide context regarding modeling approaches, strengths and limitations, thereby enhancing users' experiences with the model and promote an informed and positive vision about the results they observe from the model. Additional investments of resources and time are needed to implement best modeling practices, which needs to be communicated adequately to model sponsors and others within the larger modeling community.

Investment Protection: With the increasing complexity of water and environmental problems being addressed, model development and related analyses represent a large and growing investment of resources. Unlike databases of field observations, model results have limited shelf lives unless supported by adequate documentation, including source codes and input files. Absent this information, model studies can rarely be reused to guide future work. These protocols provide guidance on developing and maintaining such supporting material, so that the original investments in model development can be preserved and model components can be reused as much as possible.

Identification of Emerging Technologies: New technological developments have changed the practice of modeling since many of the guidance documents in **Table 2** were published and are anticipated to continue to do so. Some of these technical changes include advances in geospatial data availability, online data and code sharing, cloud-based computing, the use of smartphones as computer platforms, machine learning-based modeling frameworks, sensor-based data sources and telemetry, and data visualization. The role of these technological drivers in changing the practice of modeling is described in these protocols.

Providing Information in a Local Context: These protocols, as well as the earlier version from the year 2000, are tailored to the specific modeling issues in California, which are unique because of the state's climatic variability and interconnection of water resources infrastructure. These protocols may also be related to the evolving regulatory context in California, such as the adoption of Assembly Bill 1755 (The Open and Transparent Water Data Act), which provides specific requirements on the management of water-related data, and the adoption of the Sustainable Groundwater Management Act (SGMA).

Encouraging Adoption of these Protocols: Many of the practices described in the following chapters are often acknowledged by the modeling community as useful, but are not sufficiently implemented because of institutional or resource constraints. Our goal in recommending such practices is to highlight a range of realistic options that can be incorporated within ongoing studies without placing an unreasonable burden on model developers. To encourage adoption, we provide a set of three checklists in the Executive Summary that distill key concepts presented in this work.

1.7 Proposed Use of these Protocols

Models in general, and environmental models in particular, should strike a balance between the competing needs of accessibility and comprehensiveness. A model formulation that is more readily understandable or accessible may focus on key processes and provide a more simplified system representation while omitting more complex relevant drivers. A more comprehensive model formulation may represent many drivers and capture system complexity at the expense of greater challenges to implement, test, and explain. Model developers have flexibility in how they choose to represent a system but are usually limited by one or more of the following constraints: availability of observed data, availability of financial or human resources, availability of time, and computational resource requirements. Model development is a creative process that seeks to find the "right" or "best" course of action given the above constraints. However, given that an *a priori* "right" system representation rarely (if ever) exists, considerable subjectivity in the selection of a modeling approach exists. For these reasons, to determine if the modeled representation of a problem is correct and credible, additional testing and/or monitoring and field verification may need to be performed.

Given this context, this document provides guidance that can be used by modelers to support the execution of a model study through its various steps, such as: selecting a model, fulfilling the requirements of credibility, engaging stakeholders, performing peer review, creating suitable documentation and communicating with interested parties. In providing this guidance, we also note practical observations and constraints—typically time, personnel or other resource constraints—that prevent these practices from being applied in all circumstances. Thus, these protocols are not intended to be specific or prescriptive requirements, but to describe best practices that are expected to benefit the broader community of users shown in **Figure 2**. Some modeling domains may have customary practices or standards that can be used in tandem with these protocols, or they may take precedence over the protocols presented here. It is hoped that these protocols will provide clear direction on planning a modeling study from initial question to completion, although decisions on what specific elements will be included, will always be determined on a case-by-case basis. More importantly, these steps should be as transparent as possible, so that non-modelers are also able to actively engage in the process, and help them understand and/or decide what is most appropriate for their particular study. These protocols refer occasionally to specific model frameworks, but are intended to apply broadly across frameworks and disciplines relevant to California's water and environmental resources. For reference, a link to a specific online inventory of model frameworks is represented in Appendix B.

The modeling protocols proceed in a sequential manner—similar to how a study would be conducted—in the chapters that follow. Chapter 2 provides a classification of the different types of modeling activities, where different components of these protocols may be more suited to certain categories. Chapter 3 describes a variety of preliminary analyses that typically precede the actual exercise of modeling. Chapter 4 lists the steps involved in conducting a modeling study because this is where most of the computations and interpretations are performed. Chapter 5 presents the steps associated with typical model applications. Chapter 6 focuses on the communication and documentation of model results. Chapter 7 discusses various emerging technologies that are expected to influence the practice of modeling in coming years. Chapter 8 highlights issues that are related to long-term management of models that go beyond individual studies to the management of frameworks over long time horizons. Chapter 9 describes the need for enhanced collaboration mechanisms for the modeling community, again spanning broad topics that go beyond individual studies. Chapter 10 discusses next steps in the use and communication of these protocols, focusing on the California water and environmental modeling community.

The key ideas in the above chapters are summarized in a set of three checklists described in the Executive Summary. These checklists include questions, with either yes/no or narrative answers that can be used to characterize a modeling effort at three points in time:

1. At the inception of a modeling study,
2. Following completion of a modeling study, and
3. Over the long-term for managing a model framework that is utilized for multiple studies.

We believe that these checklists will aid communication among the model user community and will highlight important features that are described in these protocols.

2 Classification of Modeling Study Tools

Depending on the problem being addressed, modeling study tools can be classified into three categories. In the first category, models are developed and used to address problems where the science is reasonably mature and where general agreement exists on the mathematical formulations of the key modeling processes. In the second category, the study is characterized by an evolving science, and model development and application are part of the scientific investigation. In the third category, the study requires development and use of multiple models – thereby highlighting additional challenges related to model integration, data consistency and exchange, project scheduling, and interpretation of results. These modeling study tools are described below, because they have a bearing on the appropriateness of certain protocol components presented in the subsequent chapters.

2.1 Modeling with Established Frameworks

For modeling study with well-defined basic theoretical principles, mathematical representations, and computer implementations in place, the approach to performing a model study can be represented by a linear progression of steps (**Figure 3**). The main steps involve using observed or collected data from the field or reported data from other observations or methods (e.g. remote sensing) to configure and calibrate the model; apply to various scenarios; and report the results. Model results are compared against field observations or reported data and can be subjected to a variety of tests to evaluate performance. To provide additional specificity for these modeling best practices, we separate the evaluation step into two phases: an initial evaluation that is expected to be applied for all model applications and additional evaluation such as sensitivity analysis and uncertainty analysis. The latter phase requires more resources and time that are better suited for larger and more consequential exercises.

Many modeling studies follow the approach shown in **Figure 3**, where a modeling framework (various examples of which are presented in Appendix B) is customized for a specific geography. Despite the linear progression of steps illustrated in **Figure 3**, an opportunity for revision should be considered where appropriate as new insights appear along the progression of steps illustrated in the figure. Although the basic theory for this class of models is well-established, there are nonetheless many areas that are the focus of improved performance and research. These include collection of more spatially and temporally resolved field data to better configure the model; improving the calibration of the model to better fit observations; more efficient model run times; improved visualization and interpretation of results; and more sophisticated evaluation of performance. Over time, due to the level of scrutiny and the detailed questions posed, models in this category, while using the same theoretical equations to represent the underlying processes, are becoming more spatially and temporally detailed, resulting in greater computational requirements.

2.2 Modeling where Science is Evolving

The scientific understanding associated with many problems is evolving and modeling is a component of the analysis. **Figure 4** diagrams the sequence of steps that may be taken to analyze such a problem. The primary difference between an evolving problem and a well-defined problem is that model structure, data needs, or even outputs are less certain at study inception. Therefore, study focus is on collecting more data (typically new types of indicators to improve scientific understanding) and developing conceptual frameworks to explain relevant processes and drivers for a variable of interest. A conceptual framework may be thought of as a compact graphical representation of the key processes of interest in a modeling study. A conceptual framework may be converted to a quantitative model structure, thereby formally describing how inputs and outputs are related and then implemented in computer code. Such models may then be calibrated and evaluated in a manner consistent with more mature models. The modeling protocols proposed in this work apply to both newly-defined and well-established modeling processes. The distinction between **Figure 3** and **Figure 4** is made not to downplay the role of evaluating and testing practices in models with evolving science, but rather to point out that the primary attention may often be focused on improving the basic understanding and representation of the processes of interest.

Figure 3. Modeling steps for a topic with well-developed theoretical frameworks and computer implementation.

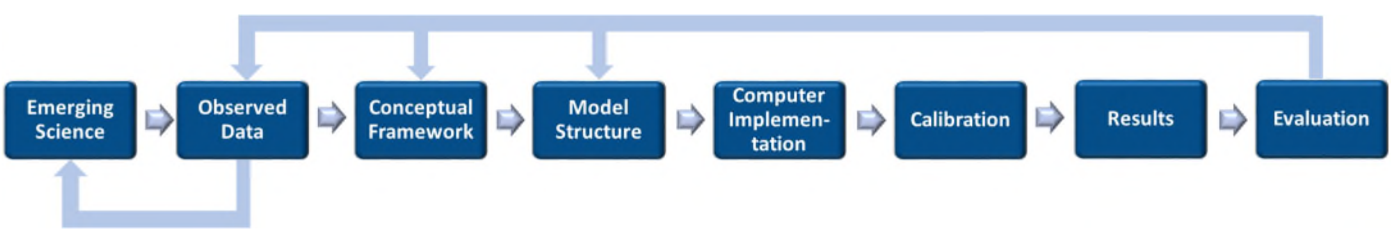

Figure 4. Modeling steps for a topic where the scientific understanding is still evolving.

2.3 Modeling with Multiple Frameworks or Across Disciplinary Boundaries

A growing need in the field of environmental modeling is to utilize multiple models to perform an analysis where impacts may span the domain of more than one model (Delta Stewardship Council, 2020). This is termed integrated modeling. Typically, this may be done with two or more models in a sequential manner, or through a more iterative exchange between models. The former approach is common and has been utilized in several major statewide efforts (Delta Stewardship Council, 2020). The use of multiple models raises additional questions related to modeling protocols, including the consideration of both technical and institutional challenges. Technical issues include computational and scientific challenges related to integration and are associated with model compatibility, data exchange and management, accessibility of models, overall complexity of integrated models, propagation of uncertainty across integrated models, and the overall limitations in model testing. Institutional challenges are primarily concerned with the human side of modeling and relate to the overall setting in which modeling occurs, the expertise needed to develop integrated models, the funding needs, and the engagement of stakeholders.

3 Preliminary Analyses

The early steps of a modeling study set the stage for a scientifically robust and defensible project and should not be rushed or curtailed. After identifying the problem or question, the next step is to confirm whether a model is, in fact, a useful tool to address the problem. Subsequently, the study participants determine what type of model would be most useful, and then whether an existing model can be used for this purpose or whether a new model should be developed. The level of model complexity chosen will be informed by budget and schedule considerations, the type and amount of available data, and the available expertise. These preliminary steps are described below.

3.1 Define the Question(s)

At the inception of a study, the specific purpose of a modeling exercise should be clearly defined. While this practice appears obvious, the modeling purpose is often not explicitly addressed up-front among modelers and stakeholders. A clear specification of the purpose in the form of goals and objectives is especially needed for modeling efforts that are for planning, management, and decision support, and not those focused on open-ended research. Goals of model development are intended to establish the long-term vision of model development to most efficiently and best serve the needs of the project. It is imperative to establish the long-term goals, as many models are developed for long shelf lives, and can span multiple studies and projects over time. Objectives of the model development can be more specific for the project or study at hand, and need to be aligned with the long-term goals. It is recommended that the stated goals and objectives should be developed in a transparent, open, and collaborative environment with stakeholders, including their technical, policy, and legal representatives. This would lead to a broad scope of work, including what processes will and will not be modeled, what data will be needed, what form the results will take, and what the expected accuracy and uncertainty will be. Importantly, a modeler needs to understand the stakeholders' viewpoint of how the model results will be used. In some instances, where a problem is addressable with an existing model framework, the model study purpose can be defined with greater clarity than when a completely new model needs to be created. The more specifics that are outlined early, the more efficiently the modeling exercise will progress. When elements of the modeling scope are not well-defined up-front and are later selected by decision-makers based on the results obtained, the result can be a less-than-optimal use of the modeling effort.

A National Research Council (NRC) evaluation on modeling practices for regulatory application (NRC, 2007) proposed the following relevant and valuable suggestions to help define the model purpose. Not all of these questions may apply to all modeling efforts, but a reasonable subset can be selected for most modeling studies:

- At what temporal and spatial scales is the model to be applied? This question involves the resolution in time and space and the spatial extent and temporal period that the model should represent.
- Who will review and benefit by model output, and what constraints does that imply for model application once developed? What is the level of expertise of these proposed users of model output?

Practical Observations. Questions often arise from stakeholders or model sponsors in the later phases of a modeling exercise that were not intended to be part of the modeling study. To the extent possible, transparency in the goals of the study will reduce such queries to the benefit of all participants. This is especially true for studies that apply established frameworks, which are also likely to be time and resource constrained efforts.

- What type of input data must the model users provide? How can these data be obtained (from other models and measurements)?
- What sources of data are available to support model calibration and evaluation?
- What are the basic outputs needed to support the decision made by policy makers and/or support the regulatory questions? What additional outputs might be useful to enhance model transparency and flexibility?
- What level of reliability (defined as the degree to which the model result can be depended on to be accurate) is required?
- What evaluation criteria should be applied to determine the applicability of the model?

The exercise to define the model's purpose should be formally and clearly documented, a common engineering practice across many other disciplines (e.g., civil, environmental, mechanical) in their design endeavors. That documentation can then be revisited and revised as the modeling effort evolves.

3.2 Translate Question(s) into Modeling Analysis

Once a question has been defined, the next step is to lay out an appropriate modeling approach. For questions where the underlying mechanistic processes are well understood (Section 2.1), one or more established frameworks will be evaluated for use, based on cost, availability, and familiarity of the model users with the framework. Such evaluation will also need to consider the type and amount of data that will be needed to conduct a reasonable analysis. The types of models available represent the range of study domains in environmental and water resource applications.

3.2.1 Conceptual Frameworks

For questions where the science is evolving or highly site-specific (Section 2.2), several additional steps need to be addressed. No established modeling framework may exist, or additional modification/customization is required to allow an existing framework to address the relevant question. In the former case, a new model may need to be developed, while in the latter case, additional characterization should be performed before an existing model is adapted. Biological and economic problems often fall in the former category, while water quality and ecosystem-level problems often fall in the latter category (Delta Stewardship Council, 2020). A key step in many such models is the development of a conceptual framework that is a precursor to a more detailed quantitative representation.

Conceptual frameworks, as used in ecological and biological analyses, are abstractions of reality, ranging from a schematic representation of processes (**Figure 5**) to a more detailed description of the state of the science related to a specific environmental concern. When developing new models, creating a conceptual framework, even a simple schematic representation, is recommended as a first step. A conceptual framework is a useful tool that helps explain the system to be modeled and further guides model development, experimentation, and evaluation. Moreover, conceptual frameworks are effective for communicating model information with stakeholders, particularly when the conceptual framework represents processes graphically and highlights key quantitative information. Finally, a conceptual framework creates a point of reference for model developers to revisit when considering changes to the model.

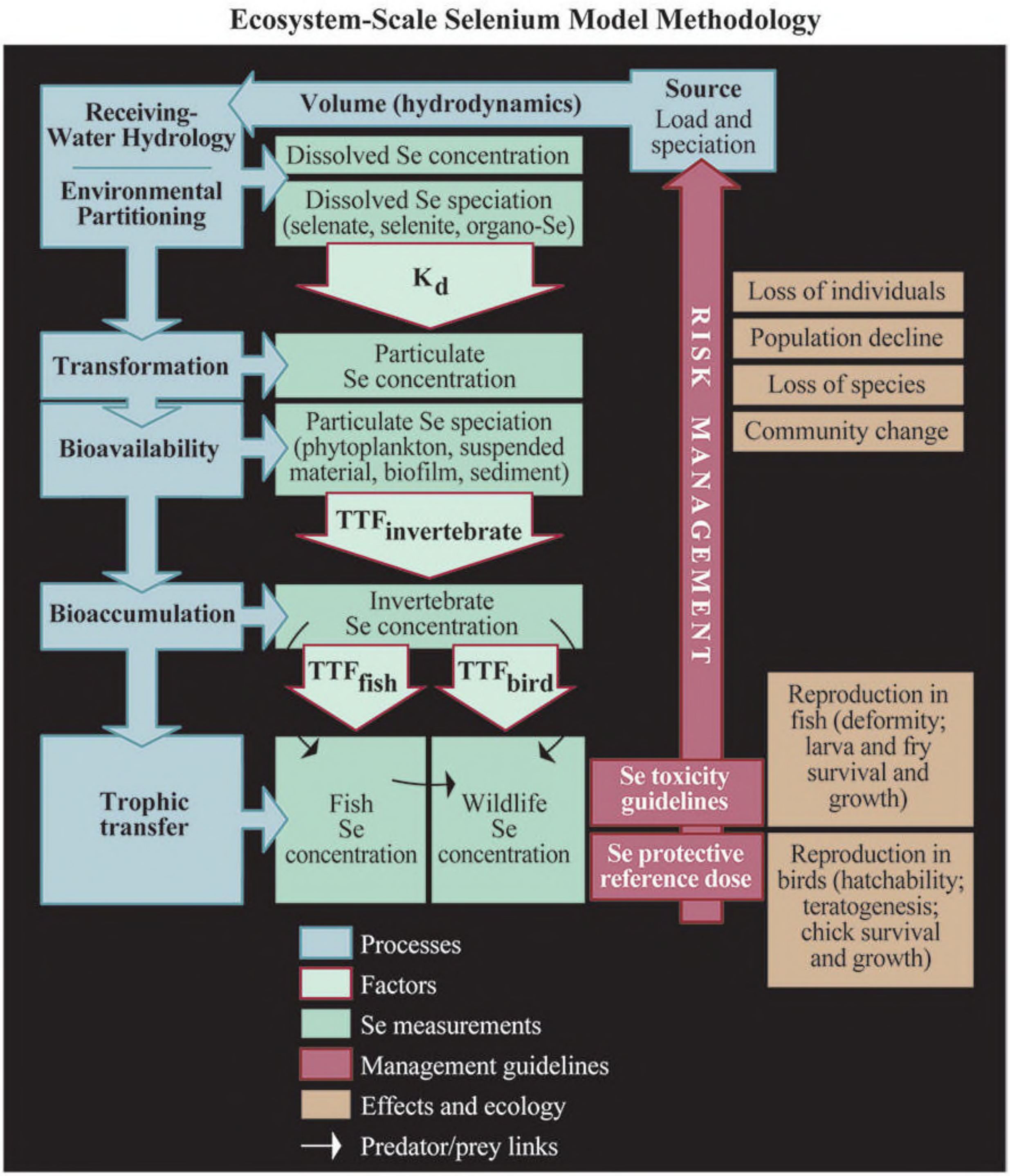

Figure 5. Conceptual representation of the uptake of selenium, a bioaccumulative element, in ecosystems from water to biota. K_d represents the partition coefficient between dissolved and particulate phases and TTF represents the Trophic Transfer Factor (Source: Project on *Linking Selenium Sources to Ecosystems: Modeling; https://www.usgs.gov/mission-areas/water-resources/science/linking-selenium-sources-ecosystems-modeling*).

In addition to enhancing communication, under certain circumstances, a well-designed conceptual framework may accommodate formal hypothesis-testing. A good conceptual framework may lead to the early realization that the development of a quantitative system model would be premature due to data and knowledge gaps.

In some instances, conceptual frameworks play a role following data synthesis and after completion of a modeling study. Typically, the initial conceptual framework would be refined over the course of model application, and more quantitative information provided in the revised conceptual framework. Such a model can serve as a basis for further communication with stakeholders. Graphical representation of modeled processes, with key quantitative information being highlighted when available, is a significant aid to communicating with stakeholders.

Conceptual frameworks have been developed and documented as stand-alone tools, combining graphical representations and narrative syntheses of available information. Examples of conceptual framework documentation include a model for Delta Smelt (Interagency Ecological Program, 2015) and a framework for nutrients in the Central Valley and Delta (Tetra Tech, 2006). These detailed conceptual frameworks are applicable when a large amount of information exists for a problem of interest and can provide a strong foundation for model development.

Practical Observations. Conceptual frameworks are useful to engage a variety of participants early in the development process, before significant investments have been made in numerical model development. Their use is widespread in areas, typically related to water quality or ecosystem processes, where the scientific understanding is still evolving.

A potential pitfall associated with conceptual frameworks is that they may be overly abstract to sufficiently guide the implementation of quantitative models. This problem is particularly relevant for large, complex processes and/or in areas when an integrated model is being considered. As a countermeasure, a conceptual framework should be re-evaluated and revised as the quantitative model is being developed.

Other potential pitfalls associated with conceptual frameworks are that they may be too complicated to allow prioritization of key processes, are overly difficult to understand, or are overly difficult to communicate to the modeling team and to stakeholders. In developing conceptual frameworks, comprehensive understanding of the system to be modeled and professional judgment are needed to identify and prioritize key processes and mechanisms, and assumptions should be clearly documented.

3.2.2 Types of model representation

A variety of modeling frameworks and associated tools for water resources problems are in common use in California. A set of these models is listed in Appendix B, with details on each model provided in a supporting online inventory. The primary modeling focus areas are listed below.

3.2.2.1 Surface water hydrology models

Surface water hydrology models simulate flow conditions in natural and developed streams and rivers, associated floodplains, and in man-made channels. Hydrology models can focus on peak flows and simulate the flood flow conditions, or may focus on low flow conditions and runoff and percolation during low flow regime. Some hydrologic models can also be applied to include reservoirs, estuaries, and coastal waters. Watershed models simulate the hydrologic processes involving simulation of runoff and percolation from rainfall and/or snowmelt, which is a function of land slope, land use and possibly crops cover, and soil conditions.

3.2.2.2 Water Quality Models

Water quality models quantify the movement and concentration of contaminants in lakes, streams, estuaries, and marine environments[3]. Water quality models may be used to assess water quality conditions and causes of impairment, predict how surface waters will respond to changes in their watersheds and the environment (i.e., from climate change), develop regulatory guidance (Total Maximum Daily Loads or TMDLs), or simulate benefits of new surface water protection policies. Water quality constituents can be categorized as physical, organic, inorganic, chemical, and biological. Pollutants may be classified by specific forms, such as biochemical oxygen demand, nitrogen, phosphorus, bacteria, or specific toxic substances. Unstable pollutants, which increase or decay with time, are termed nonconservative.

3.2.2.3 Groundwater Models

Groundwater flow models are used to simulate the rate and direction of water movement through the subsurface environment. Some groundwater models also simulate water movement through the integrated land surface, groundwater, and stream and river systems. In this case, the modeling environment provides for flow of water through all processes, and mathematical equations govern the flow process through each system and among the systems in an integrated framework. Additionally, groundwater models incorporate mathematical representations of some or all of the following processes: movement of water and other fluids through saturated or unsaturated porous media or fractured rock, transport of water-soluble constituents, transport of constituents that partition between water, air, and soil, transport of constituents subject to retardation in their movement due to sorption and desorption with soil particles (clay and organic material), and transformation of contaminants by chemical, biological, and physical processes. Examples of water quality processes simulated are the movement of point source of contaminants such as from an oil spill or underground storage tank, or from non-point sources of contaminants, such as nitrate or total dissolved solids from long-term application of fertilizers on agricultural fields.

3.2.2.4 Operations Models

Operations models involve the flow and water quantity aspects of reservoir system operations and water resources planning. The two types of operations models are: (i) simulation models, where the operation of the reservoirs and river systems are simulated with fixed rules and in a sequential manner over the single or multi-reservoir systems and the river system, and (ii) optimization models, which use stochastic and optimization rules to operate the reservoirs and river system to optimize an objective function (e.g., maximize downstream deliveries), subject to a set of constraints (e.g. meet minimum flow requirements during drought conditions).

[3] While groundwater models may simulate water quality processes, groundwater quality models (i.e., transport models) are described within the groundwater model category in this chapter and in Appendix B.

3.2.2.5 Biological Models

Biological models include representation of key ecosystem processes, such as cycling of major elements like carbon and nutrients, as well as representation of organism behavior in response to environmental drivers such as invasive species or endangered species. Key model types are described below, and specific models in use in California are presented in Appendix B.

Food web models, one category of biological models, link organisms by their feeding relationships. For simplicity, species are often placed in functional groups. Increasingly, tropic relationships are being inferred through stable-isotope analysis. In the absence of data from field studies, food webs are constructed from lab studies and literature reviews. Because large-scale food webs are inherently complex, there is a high degree of uncertainty in food web data. A dynamic food web model requires an understanding of how changes in the physical properties of the system affect the topology and magnitude of energy flows through the food web.

Another category of biological models are fish models, where statistical relationships between operations and fish parameters are often incorporated rather than differentiating the different mechanistic pathways that water operations may have on fish. One key challenge in using physical data in fish models is finding the right level of abstraction. In nature, fish react to instantaneous changes in velocity, salinity, temperature, and turbidity. For many models, though, it is not practical to run the model at a sub-hourly timestep even if physical data are available at that timescale. A necessary and useful simplification is to treat flow as a master variable that affects the underlying mechanisms influencing fish behavior and survival.

3.2.2.6 Economic Models

Models focused on economics use a variety of datasets to link economic outcomes to agricultural practices, availability of land and water, and changes therein. The major components are the amount of production, the economic value of production, and cost of resources. Key inputs to economic agricultural models are water use data and crop irrigation method data. Cost and return studies provide a breakdown of costs associated with labor, materials, equipment, and contract services to a high level of detail. The economic synopsis provides critical information on the associated cost of production for a typical acre of commodity as well as expected returns on sales for the same acre. Typically, these studies are used by operators to guide decisions, estimate potential returns, and prepare budgets; however, they also provide key data in the development of economic models.

3.3 Identify Available Information

Observed data, a fundamental part of sound modeling development and application, are critical in defining initial conditions, boundary conditions, and model input parameters. In most instances, environmental models contain parameters that are defined independently or are adjusted as part of the model setup. Parameters that cannot be determined independent of the model (e.g., a reaction rate for a chemical process within a water body or a roughness coefficient for a stream bed) are estimated through the process of model calibration (described below), which involves tuning the parameters to obtain a good fit between the model and observed data. Thus, data that are credibly measured, have good quality, have been reviewed for potential erroneous values, and are well documented (metadata, including any identified limitations) are an essential part of the modeling process. California Assembly Bill 1755 (Open and Transparent Water Data Act, AB 1755) is a major step toward standardizing data resources available to modelers. The bill requires California state agencies to make data publicly available and to develop protocols for data sharing, documentation, quality control, and promotion of open-source platforms and decision support tools related to water data. Once fully implemented, AB 1755 may provide observed data in a form that is suitable for model studies (i.e., for calibration and testing). Additionally, the California Water Quality Monitoring Council has requirements for quality assurance program plans that should be used in collecting water quality data in a monitoring program.[4]

Practical Observations. In many real-world situations where environmental models are being developed, data availability is less than ideal in type and/or spatial and temporal detail. In these situations, the time and budget necessary to collect additional data far exceeds available resources. As a result, modelers should strive to ensure that model complexity is consistent with data availability. Overly complex models supported with inadequate data will be difficult to test and evaluate.

Modelers generally prepare data sets for model calibration and testing, pulling from a variety of available data sources. Data preparation typically involves some form of compilation across different sources, conversion to common units, and quality control to remove known outliers, all of which can be time consuming, and more importantly, have a bearing on the model calibration. Modelers may obtain different model calibration results depending on the quantity and quality of data used and the specific process steps used to prepare model input data, even when utilizing the same model framework. In some cases, models may use processed values (e.g., loads derived from pollutant concentrations, or salinity isohalines from point-based values) rather than directly observed data. The creation of standardized input datasets — whether using directly observed data or some processed form — is recommended, especially when many different users are expected to be involved in parallel studies.

Standardized datasets should be thoroughly described via metadata, corroborated with explicit references, and prepared for analysis with a reproducible and documented workflow. Use of standardized datasets limits the time lost due to errors in model runs arising from incorrect input data. Preparation of standardized datasets involves decisions about missing values and specifications of data types (e.g., date, integer, or string). One of the central challenges of preparing standardized datasets is anticipating the possible ways that a dataset could contain inconsistencies. For example, does the data provider use letter codes, dashes, or numerical codes (e.g., 999) in place of missing data causing potential type mismatches? Does the test dataset include all the possible permutations of codes produced by a data provider? Does the data quality control process check for the distribution of the data for early recognition of possible changes in data structure? Ideally, these decisions should be documented for future users.

Typically, such standardized datasets are not easily available and may not be part of raw data sources. Therefore, a related recommendation is the creation of a searchable database or similar data repository of such standardized datasets where a user can identify appropriate information for use in a modeling study.

3.4 Model Selection

After the consideration of type of model representation needed (as discussed above in Section 3.2.2), model selection requires consideration of the modeling objective and model complexity. Other

[4] Available at https://www.mywaterquality.ca.gov/monitoring_council/index2.html

considerations include whether an off-the-shelf model can be used or if a new model needs to be constructed, if a public domain or proprietary model is appropriate, the need for code verification, and if multiple models are to be used how they will be integrated. Each of these topics is addressed below.

3.4.1 Understanding the Role of Accuracy Versus Precision

> *"Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise"* (Tukey, 1962).

Accuracy in the context of modeling refers to how close modeled values are to the true or observed values. Precision refers to how close model predictions of the same item are to each other. Precision is independent of accuracy as shown in **Figure 6**. While the ideal in modeling is to be both accurate and precise, this is rarely achieved. A more practical goal is to be accurate within a reasonable range, i.e., to be approximately correct. A precise model that is not accurate is not particularly useful. To the extent possible, model selection should carefully consider the level of accuracy and precision that can be achieved.

3.4.2 What Level of Complexity is Needed?

Model complexity describes the spatial and temporal resolution of the model as well as which processes are incorporated into the model. An appropriate level of model complexity is a function of multiple factors including, but not limited to, objectives of the modeling exercise, knowledge of the system, and data availability. A useful rule of thumb for deciding on the level of complexity a priori is that the model outcomes should be testable by observed data spatially and temporally. For example, when choosing between a simple lumped parameter model and a more sophisticated distributed parameter model, a model developer should be able to 1) support the added complexity by more detailed input data available for the distributed model (e.g. distributed measurement of related properties), and 2) test whether this added complexity is providing an additional benefit by comparing the simulations with available observed data.

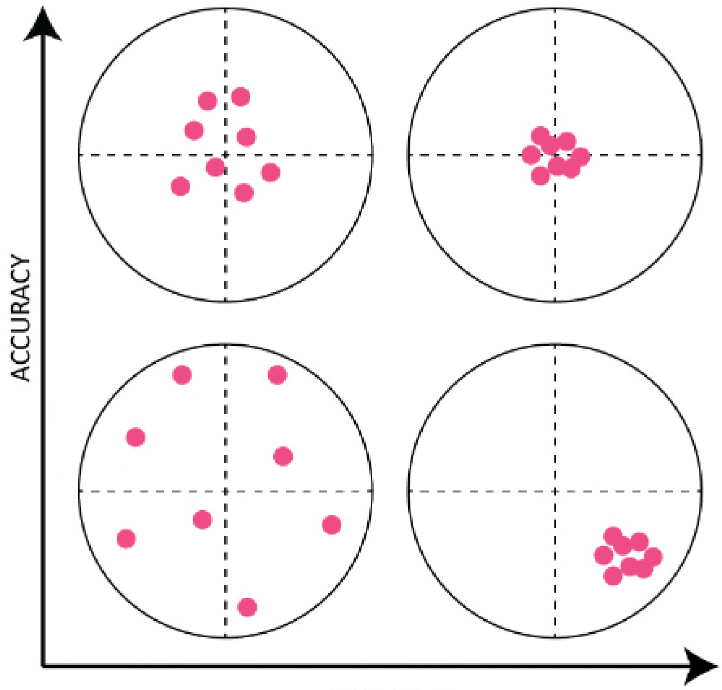

Figure 6. A schematic representation of the concepts of accuracy and precision. The center of each circle is the desired target.

> *Practical Observations.* Often, model study stakeholders push model specialists to consider using more complex or more spatially detailed models than is warranted by available data. The stakeholders make this push because they believe that more complex models are more accurate. It is important to clearly communicate to stakeholders the essential role of observed data in creating credible models, and, where data availability is modest, to use models of lower complexity.

The inability to identify a single representative model is a clear cost of excessive model complexity. Complex models often have a larger number of parameters, and under these conditions, different combinations of parameter values can lead to similar model results when compared to observed data (e.g., runoff at the catchment outlet or water level in groundwater bores). Such a result implies that the observations are insufficient to properly test the model structure or parameter values. Furthermore, even if a model appears to accurately simulate a particular response, this result does not necessarily indicate that other model predictions are correct. For example, although a rainfall-runoff model may provide good fits to streamflow at a catchment outlet, it may not necessarily provide accurate streamflow estimates at internal gauging stations or correct spatial patterns of saturation deficit. This issue has been clearly identified by many researchers (Grayson and Blöschl, 2001; Tasdighi et al., 2018), yet it is commonly ignored by model users. This issue is often referred to as "equifinality" or "non-uniqueness" in the literature and is a subject of continuing discussion (Beven, 2001).

Figure 7 illustrates the conceptual relationship between model complexity, data availability, and predictive performance. The term "data availability" refers to both the amount and quality of the data in terms of its use for model testing. Within the context of hydrology, access to spatial patterns of surface runoff data is considered "high" availability while scarce streamflow measurements as aggregated runoff implies "low" availability. The term "model complexity" means detail of process representation and spatial/temporal detail. Complex models include more processes and report values at greater spatial and temporal density. As illustrated in **Figure 7**, for a given data availability, there is an optimum level of model complexity giving the highest predictive performance; additional complexity leads to concerns with a larger number of uncertain inputs. Another way of looking at this is through the concept of overfitting, shown schematically in **Figure 8**. A model may be too simple for the data or may overfit the data, the challenge of model selection is to find the right level of complexity.

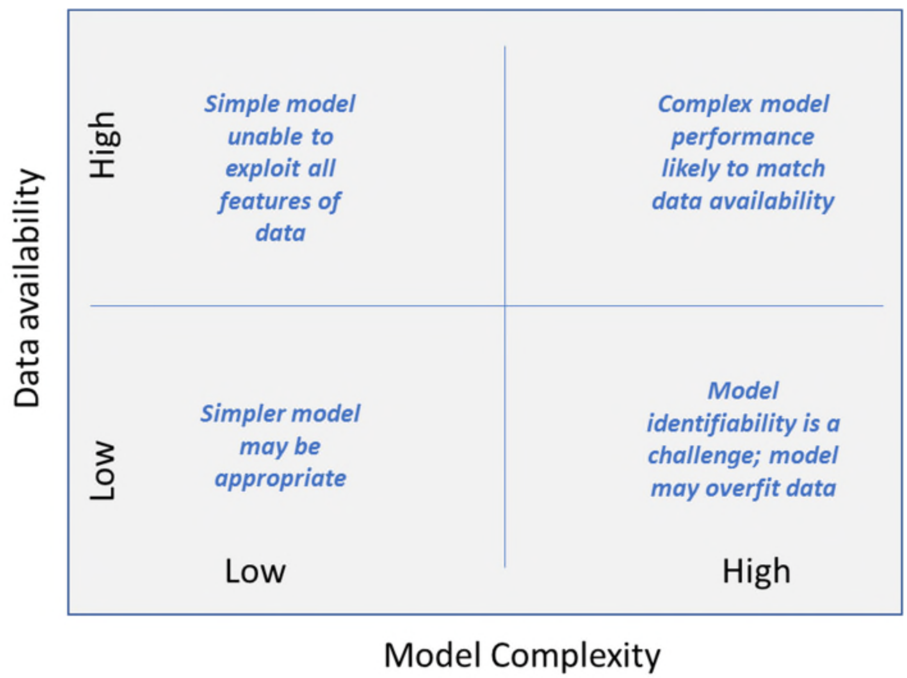

Figure 7. The conceptual relationship between model complexity, data availability, and performance (modified from concepts in Grayson and Blöschl, 2001).

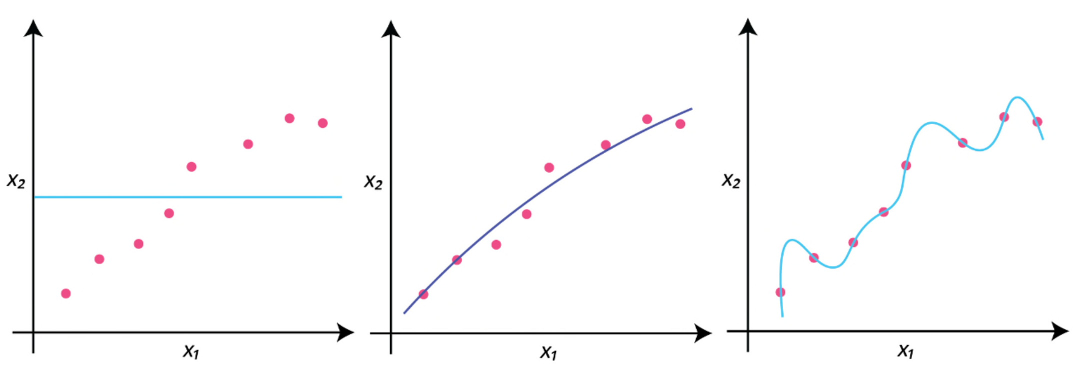

Figure 8. A schematic representation of model fitting (blue line) for observed data (red points). The left plot shows a model that is too simple for the data, and the right plot suggests a model overfitting the data. The middle plot is a conceptual representation of the "right" level of complexity.

For a given model complexity, greater data availability usually results in better predictive performance up to a point, beyond which the data does not provide more useful information to improve the model with that level of complexity. Under these conditions, a user may wish to consider a more complex model to better exploit the information from the available data.

3.4.3 Off the Shelf vs. Building a New Model

The following questions should be considered when assessing the need for a new model versus using an existing model (USEPA, 2002):

- For what specific tasks will the model be used in the given application?
- What data will be collected or obtained to characterize the application site and to develop a site conceptual framework that will be compared with existing models? What is the needed spatial and temporal scale of the model inputs?
- What model outputs are needed? What is the spatial and temporal scale needed for the outputs?
- What levels of uncertainty are acceptable in model outputs?
- What are the strengths and weaknesses of existing models?
- If an existing model is available, are its parameter default values, input data, boundary conditions, and underlying assumptions acceptable?
- Is the existing model software compatible with the modeler's hardware/software configuration requirements for the new application?
- Are any improvements in the existing model's computer code operating characteristics (e.g., run time) needed?
- Do the quality and documentation associated with the existing model meet the project-specific requirements?

As noted in the context of schedule and resource considerations, new model development is generally expected to be more expensive than the use of a prior framework. However, based on the above questions, there will always be specific project needs that require the creation of new models.

Where the task requires use of an existing model, more than one model may be available for use. In this case, the decision process should include consideration of the model life cycle (see Chapter 8). Where the task requires the development of a new model, the modeler has some discretion on the level of process complexity to be used. In both cases (existing or new model), the modeler has the flexibility to determine the level of spatial and temporal detail incorporated. As an example, a dynamic model may compute and report values at timesteps of minutes, days, or longer. A spatially detailed model may contain a grid with sizes ranging from square meters to hundreds of square kilometers.

3.4.4 Open Source, Public Domain and Proprietary Models

Environmental models may be open source, public domain, or proprietary. Open source models are those where the underlying source code of the model is available for anyone to examine and modify, potentially creating a new executable version of the model. Public-domain models are those where the executable version of a model is freely available, although the source code may not necessarily be available. Finally, proprietary models are owned by a non-public entity and there is usually a cost for leasing and applying the model.

Each approach has strengths and weaknesses as outlined below:

- Open-source models: These models are free to use and their source codes can be modified by anyone. In many cases, well maintained and documented open-source models may be the basis for major modeling studies, as is the case with three water and environmental models frequently applied in California: DSM2, IWFM, and CalSim. Open source models are also suitable for new scientific applications, where there may be a need to add new process information to an existing model by making changes at the computer code level. In most cases, considerable user expertise is needed to make meaningful changes to complex environmental models. Where a community of modeling experts exists, open-source models are an effective means for continued development. In general, however, the ability of any user to change the model can create a concern with version control, in that specific outcomes may be a consequence of the particular variant of the model being used. Furthermore, for open source models to be sustained, there is a requirement for funding of staff for development; often this is done through government or academic organizations.
- Public-domain models: These are free to use, although there may be limits to what can be changed in a published form of the model framework. The costs of development are borne by the sponsoring organization. In some cases, sponsoring agencies have provided resources to make their public-domain models easy to use in a manner similar to some proprietary models. Model credibility is generally a function of the credibility of the sponsoring organization. These models are suitable for studies with large teams of modelers, and as educational tools.
- Proprietary models: Fees for use may be significant, and thus limit who can directly use the model. Fees provide continuing resources for the developing organization to protect the intellectual properties and improve the code and the user-friendliness of the model. Where a model has uses in many geographic domains, the development costs may be spread over a larger user base. These models are suitable for studies where available model features adequately represent the modeling purpose, and an off-the-shelf product can be used. This general description applies to proprietary model frameworks and proprietary model applications.

In California, a mix of open source, public domain and proprietary models has evolved in response to several factors, including: the history of development in different domains, sponsoring agency involvement, and resources for new models. As with other elements described in this chapter, the ownership of the model can in some cases influence the best practices actions that can be applied.

3.4.5 Code Verification

Code verification is the process of determining how accurately a computer program correctly solves the equations of a mathematical model. It is assumed that most established model frameworks in common use will have undergone this test and, thus, this task is appropriate when a new code or module is being developed for a specific application. Code verification also provides an opportunity to evaluate or reevaluate the efficiency of the code, which may enable its use for situations that require multiple model runs, such as for sensitivity analysis. Typically, computer codes are verified with well-documented data sets and the results of published and documented analytical or semi-analytical models. Within many large-scale computational models, opportunities exist to perform verification studies that reflect the hierarchy or collection of these models. For example, code verification can successfully employ "unit tests" that assess whether the fundamental software building blocks of a given code correctly execute their intended algorithms. Documentation of code verification, especially for newer models or for models where modifications are being made to established codes, is an important part of establishing model robustness.

3.4.6 Integration Between Models

Integrated modeling is an approach where two or more models, typically with different areas of focus, are used together in an analysis. Integration of more than one physical process can also occur internally under the umbrella of a single model. Integrated modeling can be applied to support analyses that cross different disciplinary domains, and are often needed to understand the effects of new regulations or major new infrastructure. Integrated modeling is widely used in the physical, chemical, and biological domains, with growing and emerging opportunities in the economic and social science domains, respectively (Delta Stewardship Council, 2020). Typical examples where such modeling may be used include: long-term planning, short-term forecasting, regulatory decision-making, planning for changes to or developing new infrastructure, and even for developing a scientific understanding of a complex system. Different approaches are used for integration, ranging from simple file exchange across pre-existing models (with minimal code modification required) to the development of entirely new model codes.

3.5 Schedule and Resource Considerations

While the discussion of schedule and resource considerations is presented here as one step of the preliminary analyses, in reality, schedule and resource considerations must be considered throughout the entirety of the modeling process. Often, the model study sponsor provides the modeling team with a budget and an overall schedule and asks the team to come up with a more detailed schedule that factors in the intricacies of the modeling work. Although it is challenging to provide guidance on the actual time needed for a particular modeling study, key elements can be highlighted for consideration.

As noted in the Introduction, time allocations are important components in the development of a credible modeling study and may not be sufficiently performed without sufficient time being allotted. For model applications with established frameworks (Section 2.1), it is important to utilize not more than 50 percent of the time available in performing the first complete set of model runs. Of the remaining 50 percent, approximately half should be allocated for reviews, re-running the model, and stakeholder interaction, and the remaining half should be allocated for communication, documentation and wrap up of the study. While the exact percentages will vary, of course, in the experience of the authors, too much time is spent in preparing the first set of model runs while other aspects are shortchanged. Sufficient time is needed for modelers to adequately respond to initial review by others and to the stakeholders' issues, questions, and concerns. Additional details on these protocol components are provided in Chapter 4.

For model studies where the science is evolving (Section 2.2), the time allocated for modeling may be on the order of a few years even when using less-complex models. Many of these models are developed in research and academic settings, where multi-year time frames are typical. This allows time for development and refinement of conceptual frameworks, and even supporting data collection, creation and testing of new codes, and developing and implementing applications. These modeling studies should also consider adding review processes as described above, which adds further time to the effort.

> *Practical Observations.* Modeling, as with any complex intellectual endeavor, almost always takes more resources and time than originally envisioned. Advances in computing speed over the past two decades appear to have had minimal effect on the time needed for conducting credible model studies.

Resource requirements vary depending on the model study setting. Although computational and software resources should be included in the budget, in most cases the primary resource requirement is for staff support, whether staff in an agency or outside consultants. In the experience of the authors, modeling studies often cost more and take more time than envisioned, with more challenges in a consultant setting with defined goals and fixed budgets than for in-house agency staff, due to the level of effort being unknown ahead of time. Schedule and possibly cost over-runs are related to the cycles of revision and review that are commonplace in model studies, and are likely caused by an over-optimistic estimate of the time needed for these phases. However, problem complexity is also a reason for development times and resources exceeding initial estimates.

If an established modeling framework exists and is applicable to the modeling question to be addressed, its use will likely be more cost- and time-efficient than developing a new model framework. This is usually true even when the model framework is proprietary and involves significant license fees for its use. This is because the development of any new model is generally an uncertain endeavor and can end up costing far more than initial optimistic estimates. New model development should be considered as an option, but only when existing models cannot fully address the questions being asked.

4 Framing the Modeling Study

After the problem or question is defined and the modeling framework has been selected, the modeling study can be initiated. Model setup steps such as selection of boundary and initial conditions, time step, and physical geographic layout should be carefully considered to accommodate possible future scenarios. Input data should be selected to ensure that model results are not later called into question due to improper data quality assurance. Steps in the modeling analysis such as the model calibration, model validation, and model sensitivity should be clearly documented as they are conducted, along with record-keeping of assumptions and limitations, so that these elements of the modeling study will later become part of the model application documentation.

4.1 Frame the Analysis

Framing of the analysis refers to the setting of the geographic extent, spatial scale, and dimensionality of the model, configuring the model to represent the background conditions, and defining the boundary conditions to focus on the problem at hand. Additional details on these aspects of the model setup are provided below.

4.1.1 Boundary and Initial Conditions

For most of the model types described above, the system is composed of the elements shown in **Figure 9**. The model is driven by initial and boundary conditions of the variables of interest, where the initial conditions represent the values at the beginning of the model run and the boundary conditions represent values at the edge of the spatial extent to be modeled. Specification of initial and boundary values influence the time evolution and spatial scale of model calculations. The model configuration is used to define the background setting over which the calculation is being performed, such as the bathymetry of a water body or the depth of an aquifer. Within the model, there are usually some pre-defined or adjustable parameters. Pre-defined parameters refer to values that are independently measured or known, such as the properties of water density as a function of temperature. Adjustable parameters are typically those that cannot be measured directly and are derived by fitting the model to observed data (a process called calibration). The model may calculate values (over time or space) based on the equations, configuration, and boundary conditions, termed the internal state variables. A subset of or an interpreted summary of the state variables may be presented as outputs; outputs may be presented in tabular form or in various graphical forms. Best practices for modeling are related to each of these elements.

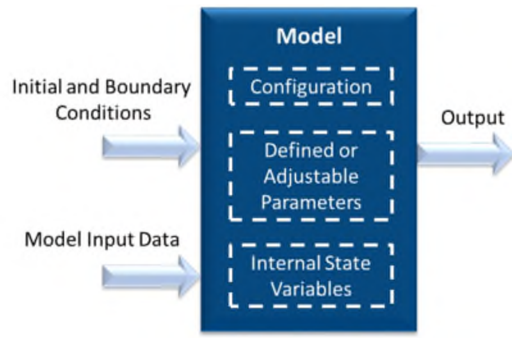

Figure 9. Major elements in model systems.

4.1.2 Geographic Scope

The geographic focus of a model will usually be determined by the question being asked, although there will be a need to consider upstream effects that can influence outcomes in the focus area. For example, a model for a receiving water body will need to consider inputs from its contributing watershed. If no data are available at the boundary between the receiving water and its watershed, the model extent may need to be extended to include the upstream area as well.

4.1.3 Spatial Scale

Spatial scale refers to the discretization of the model geographic scope or model domain to smaller units for analysis and solution of the equations representing the physical system. The size of model units may be determined by the amount of model parameter or input data or the scale at which model output is needed.

4.1.4 Time Window

The time window refers to a period, real or synthetic, over which model results are presented. This depends on the data availability, the time frame of interest in the model question, and also the model complexity. For complex, numerically detailed models, computational demands limit the length of the time window over which model results are reported.

4.1.5 Time Step

The time step refers to discrete time points associated with model calculations and can range from seconds for process-oriented physically-based models to years for economic planning models. The time step is often a feature of the model formulation.

4.2 Model Preparation and Evaluation

4.2.1 Model Configuration

Model configuration refers to the basic setup of a model representing the geographic extent, dimensionality, and the physical properties of interest, such as the bathymetry of a water body or the depth and layers in an aquifer. Spatial data are needed to adequately configure a model, and the granularity of the configuration should be consistent with the quantity of data available.

4.2.2 Data Quality Assurance

Models are dependent on data, and thus the quality of the data used for various steps in the modeling (i.e., configuration, calibration, and validation) has a bearing on the study results. USEPA (2002) identifies quality assurance measures for data that have been collected under previous efforts outside of the modeling project. A quality assurance project plan document for this data should be prepared, following guidance in USEPA (2002). This will address four issues regarding how measurements are acquired and used for the project:

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

Examination of these four issues will inform a determination of whether the given data source is an acceptable input to the modeling project, which in turn may reduce the chance of making decision errors based on the model results.

If suitable or insufficient data is not available from previous efforts outside of the modeling project, then it is recommended that data be collected in a monitoring program that follows the USEPA Guidance on Systemic Planning Using the Data Quality Objectives Process (USEPA, 2006) and the USEPA Guidance for Quality Assurance Project Plans (USEPA, 2002). A quality assurance project plan should also be prepared for the monitoring program, following the 24 elements provided in USEPA (2002).

4.2.3 Model Calibration

As previously discussed, environmental models often use parameters that are known within a range and the most appropriate values are derived on a site-specific basis from the observed data. Models use parameters within equations to relate various influences and responses (e.g., rainfall to runoff). Some of these parameters may be readily determined based on field measurements or other observations. Often, however, many model parameters are either too difficult to measure (specifically with proper spatial resolution) or practically impossible to measure (non-measurable parameters). An example of a parameter that is too difficult to measure with adequate spatial resolution includes the hydraulic conductivity in aquifers (used for groundwater modeling); or the parameter Manning's n coefficient for roughness in surface water bodies (used for streamflow modeling). Furthermore, some domains, notably in the biological, economic, and social sciences, inherently use parameters that are lumped and location specific, and not known *a priori*.

Depending on the level of complexity, models can be posed with a small number of parameters or can be posed with a very large number of parameters – in extreme cases numbering in the thousands. The task of calibration—also termed training—is to find the set of best-fit parameters that describe the observed data with a given model. Formally, calibration is the mathematical process of searching for a solution that minimizes or maximizes an objective function (i.e., a function quantifying a measure of error based on model simulations and observed data), by adjusting the values of *n* unknown parameters, which is a search in *n*-dimensional space. The general goal is to find a global best-fit, but in complex models this is often difficult, and it is not uncommon to find model calibration codes settling in local minima (see **Figure 10**). Superficially, local minima have some features of a global minimum, but formally, they do not represent the best parameter fit.

There are a wide range of common performance metrics for model calibration and testing used in the literature of environmental modeling as presented in **Table 3**. Since all model performance metrics have strengths and weaknesses, it is recommended that more than one metric (i.e., multi-objective optimization) be considered for calibration/testing of models. However, care should be taken as these metrics have different units and ranges. There are numerous published algorithms to help perform this search that are used in conjunction with environmental models, of which the Parameter Estimation and Uncertainty Analysis (PEST) tool is widely used for environmental models (theory in Doherty and Hunt, 2010; Doherty, 2015; example application in Doherty and Johnston, 2003).

The search process of finding best-fit parameters in calibration requires the model to be run multiple times, each run using a new combination of parameter values. As the number of parameters in a model grows, and as the model run-time increases, the computational burden of automated calibration grows exponentially. In many cases where complex, computationally intensive models are being used (with single run times over hours to days), calibration is often a more manual process, with expert users interacting with the model and applying knowledge of the parameter space to tune the overall performance. In a manual calibration process, model parameters are essentially tuned to minimize the difference between the model simulation and observed data. This is an iterative procedure and usually several rounds of model runs are performed to locate parameters that mimic the observed data with reasonable accuracy. Alternatively, additional computer resources are deployed during the calibration period, running the model on supercomputers or on the cloud to circumvent the computational burden.

A good practice in all types of model calibration is to set aside a fraction of the observed data to independently evaluate the performance of a calibrated model.[5] This is an essential step for statistical/empirical/machine learning models that have no underlying theory and the model credibility is based entirely on application to additional data. Indeed, in many machine learning platforms the separation of the available data into subsets for additional testing is a standard feature. This practice of setting aside some fraction of the total data for evaluation is also widely used for mechanistic simulation models. Although such models are based on theoretical foundations, it is advantageous in most settings to demonstrate good performance across a range of conditions represented by independent data.

[5] In some modeling literature, this step has been referred to as validation. For the purpose of this work, validation is defined more broadly and presented in the next chapter.

Table 3. Common Model Performance Evaluation Metrics

General category	Performance metric	Description	Issues	Reference
Standard Regression	*Slope and y-intercept*	The slope indicates the relative relationship between simulated and measured values. The y-intercept indicates the presence of a lag between simulated and measured data, or that the data sets are not perfectly aligned. A slope of 1 and y-intercept of 0 indicate that the model perfectly reproduces the measured data.	Most often the underlying assumptions of linear regression (normality, randomness, etc.) are overlooked which can undermine the credibility of the inference from a regression model	Willmott, 1981
	Pearson's correlation coefficient (r) and coefficient of determination (R^2)	r and R^2 indicate the degree of collinearity between simulated and measured data. r is an index of the degree of linear relationship between observed and simulated data and ranges from −1 to 1. If r = 0, no linear relationship exists. If r = 1 or −1, a perfect positive or negative linear relationship exists. Similarly, R^2 describes the proportion of the variance in measured data explained by the model. R^2 ranges from 0 to 1, with higher values indicating less error variance, and typically values greater than 0.5 are considered acceptable.	r and R^2 are very sensitive to high extreme values (outliers) and insensitive to additive and proportional differences between model predictions and measured data.	Santhi et al., 2001
Dimensionless	*Index of agreement (d)*	Standardized measure of the degree of model prediction error and varies between 0 and 1. A computed value of 1 indicates a perfect agreement between the simulated and measured values, and 0 indicates no agreement at all.	d is overly sensitive to extreme values due to the squared differences.	Willmott, 1981

General category	Performance metric	Description	Issues	Reference
	Nash-Sutcliffe efficiency (NSE)	The Nash-Sutcliffe efficiency (NSE) is a normalized statistic that determines the relative magnitude of the residual variance ("noise") compared to the measured data variance ("information"). NSE ranges between −∞ and 1.0 (1 inclusive), with NSE = 1 being the optimal value. Values between 0 and 1.0 are generally viewed as acceptable levels of performance. Values less than zero (<0) indicate that the mean observed value is a better predictor than the simulated value, indicating unacceptable performance.	NSE is sensitive to high extreme values.	Nash and Sutcliffe, 1970
	Persistence model efficiency (PME)	PME is a normalized model evaluation statistic that quantifies the relative magnitude of the residual variance ("noise") to the variance of the errors obtained by the use of a simple persistence model. PME ranges from 0 to 1, with PME = 1 being the optimal value. PME values should be larger than 0.0 to indicate "minimally acceptable" model performance.	Explicit assumption that variance increases linearly with time which should be revisited depending on the problem	Gupta et al., 1999
	Prediction efficiency (Pe)	P_e is the coefficient of determination (R^2) calculated by regressing the rank (descending) of observed versus simulated constituent values for a given time step. P_e determines how well the probability distributions of simulated and observed data fit each other. A prediction efficiency of 1 is perfect agreement at all times. Prediction efficiencies less than or equal to 0 do not provide useful predictions of the time variation of the observations.	Sensitive to high extreme values	Santhi et al., 2001

General category	Performance metric	Description	Issues	Reference
	Performance virtue statistic (PVk)	The performance virtue statistic (PV_k) is the weighted average of the Nash-Sutcliffe coefficients, deviations of volume, and error functions across all flow gauging stations within the watershed of interest. PV_k can range from $-\infty$ to 1.0, with a PV_k value of 1.0 indicating that the model exactly simulates all three aspects of observed flow for all gauging stations within the watershed.	Since the main criteria used is NSE, this metric can also be prone to biases from large error residuals	Wang and Melesse, 2005
	Logarithmic transformation variable (e)	The logarithmic transformation variable (e) is the logarithm of the predicted/observed data ratio. The value of e is centered on zero, symmetrical in under- or overprediction, and approximately normally distributed.	Not widely used and may not add much value considering the underlying distribution	Willmott, 1981
Error Index	*Mean absolute error (MAE), Mean square error (MSE), and Root mean square error (RMSE)*	RMSE, MAE, and MSE values of 0 indicate a perfect fit. RMSE and MAE values less than half the standard deviation of the measured data may be considered low and that either is appropriate for model evaluation.	Since these metrics use averaging on error residuals, they may not be suitable as an objective function for calibration. However, they can be used as additional performance validity metrics once the model is calibrated.	Moriasi et al., 2007
	Percent Bias (PBIAS)	Percent bias (PBIAS) measures the average tendency of the simulated data to be larger or smaller than their observed corresponding values. The optimal value of PBIAS is 0.0, with low-magnitude values indicating accurate model simulation. Positive values indicate model underestimation bias, and negative values indicate model overestimation bias.	The effects of individual error residuals may smooth out due to averaging	Gupta et al., 1999

General category	Performance metric	Description	Issues	Reference
	RMSE-observations standard deviation ratio (RSR)	RSR standardizes RMSE using the observations standard deviation, and it combines both an error index and the additional information. RSR is calculated as the ratio of the RMSE and standard deviation of measured data. RSR varies from the optimal value of 0, which indicates zero RMSE or residual variation and therefore perfect model simulation, to a large positive value. The lower RSR, the lower the RMSE, and the better the model simulation performance.	Same issues with RMSE	Gupta et al., 1999

Another fundamental challenge associated with model calibration is that the relationship between model error and fitting parameters (termed the error surface) may be complex and fitting procedures may produce locally-optimum rather than globally-optimum parameter values. Local and global minima for a single variable are shown conceptually in **Figure 10**. Parameter identifiability is the possibility of learning the true values of underlying parameters with a large experimental dataset (Raue et al., 2009). Parameter identification for complex models is very challenging and true parameters values are often not obtained because of the increased computation burden. The topic of parameter estimation in environmental models is an active area of research, focusing on improving efficiency in search strategies and on finding global best fits (Solomatine et al., 1999; Thiemann et al., 2001; Madsen, 2003; Zhang et al., 2011; van Vliet et al., 2016). Regardless of the approach used for calibration, model documentation should describe the approach and explain why the approach is credible for a specific model.

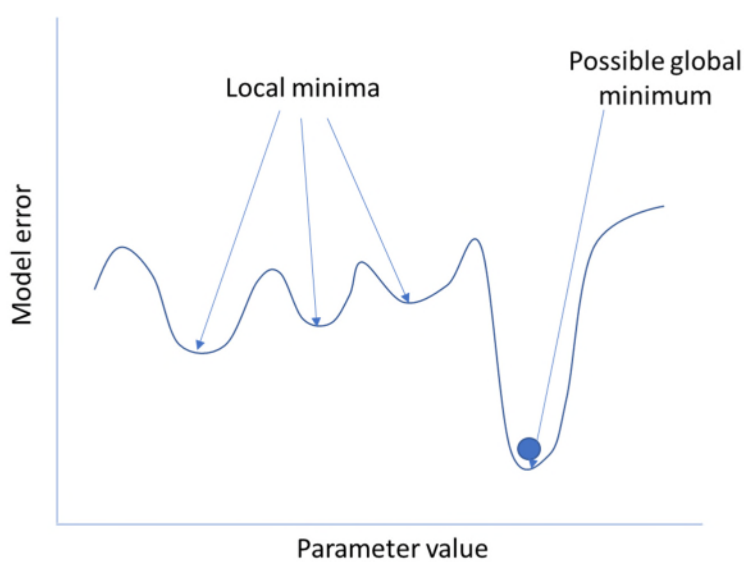

Figure 10. Schematic representation of a complex error surface with multiple local minima.

4.2.4 Reporting Model Performance

A common framework may be used for evaluating model results in a systematic manner. A range of visualization approaches (one or more of the combinations shown in **Figure 11**) is considered suitable for evaluating quantitative results of the performance of a previously calibrated model. A model's target performance may be defined as part of the stated modeling purpose or based on the best professional judgment of the modelers, given the uncertainties in input data, model parameters, and model structure.

4.2.5 Model Validation

"No matter how many times the results of experiments agree with some theory, you can never be sure that the next time the result will not contradict the theory. On the other hand, you can disprove a theory by finding even a single observation that disagrees with the predictions of the theory" (Hawking, 1988).

In this work, validation is referred to as a set of steps, beyond calibration, that are used to test the correctness of a model. When considering the process of modeling and validation, it is important to note, with a few exceptions, the theory underlying most environmental models is often provisional. Site-specific water resources and ecosystem models are elements of applied science -- in effect, an agglomeration of multiple physical, chemical and biological theories. As such, they are subject to improvement via invalidation, but cannot be proven valid under all conditions.

For mechanistic models, validation can be thought of as a broad and continuing process, comprising varied analysis of the underlying biophysical representations under different conditions and exploring the role of uncertainty and parameter sensitivity, as described in the following chapters. Although the model calibration may include a diverse set of conditions, (e.g., wet and dry hydrologic cycles), the goal of such validation is to further demonstrate that the fundamental processes in the model perform reasonably under as wide a range of conditions as it may possibly be applied as known *at the time the model is being developed.* It is entirely possible, that in future, a model is exposed to conditions outside the bounds of model calibration, and that it may not perform as accurately as conditions it was calibrated for, i.e., model predictions do not match real-world observations as well as for the period of calibration. In such an instance, the calibration and validation processes need to be repeated over this broader time or geographic space. Implicit in the above statements is the acknowledgement that, despite calibration and validation, a model is not proven to be generally true for all conditions, only that it is demonstrated to work well over a range of conditions considered adequate by the model developers. These conditions include those for which data are available, but also extrapolation to a wider range based on the theoretical underpinnings of the model. However, this does not mean that a validated mechanistic model will necessarily apply for all conditions. In the specific context of California water resources models, future extreme conditions related to drought, high temperatures, or sea level rise, all influenced by climate change, can be thought of as conditions that are outside the domain of mechanistic models being calibrated and validated today. The application of models to these future extremes is often required for planning, but in the best case, model calibrations should be updated as new data become available.

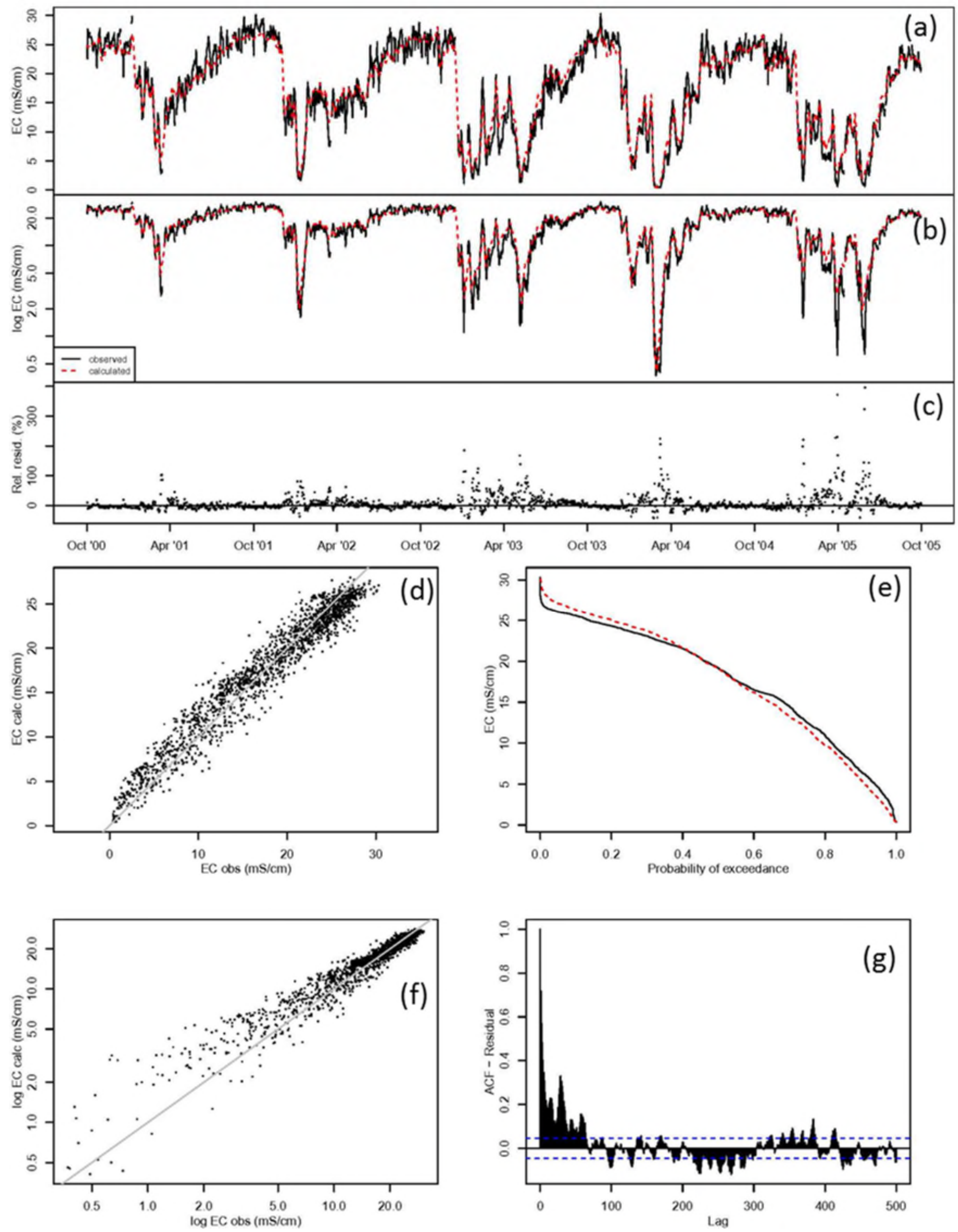

Figure 11. Visualization of adequacy of model performance. Following Crout et al. (2008), but applied to salinity at Martinez in the western Delta, using observed data (Hutton et al., 2015) and a published model of salinity (Rath et al., 2017). (a) Linear time series plots of data and observations (solid line: model; dashed line: observations), (b) log-scale time series plots, (c) plot of residuals (difference between modeled and observed values), (d) observed versus modeled data on a linear scale, (e) cumulative distribution function of observed and modeled values, (f) observed versus modeled data on a log scale, and (g) autocorrelation function of residuals.

The term validation is also used in the context of empirical/ statistical/machine learning models (or empirical models for short), but here the meaning is more limited, and generally refers to the evaluation of model performance using additional data, beyond what was used for model formulation and fitting. And, to be done correctly, the validation data must be strictly separated from the data used for model development. Multiple splits of the same data set can be performed, in multiple cycles of fitting and validation, but in each cycle some data must be set aside for validation. Unlike a mechanistic model, however, a validated empirical model has no theoretical underpinning, and is of limited value outside the data space in which it was developed. For this reason, such models need extensive data sets to be calibrated and validated, and should be routinely updated as new data become available.

The above discussion highlights that validation is an important part of modeling, but to be true to its meaning, especially in the context of mechanistic models, it needs to be performed in a more exhaustive manner than simply testing with an alternative set of data.

4.2.6 Modeling Uncertainty

Models, as simplifications of reality, are subject to various forms of uncertainty. In water and environmental models specifically, these sources of uncertainty include:

i) Structural uncertainties associated with the model conceptualization and formulation, i.e., uncertainties in the basic representation of the natural world in mathematical form,

ii) Calibration uncertainties related to the fitting approach and the values of the estimated parameters,

iii) Data uncertainties including initial state variables, configuration and input variables,

iv) Data uncertainties related to observed data used for training and testing the model,

v) Projection uncertainties resulting from the need to make forecasts for developing future scenarios, incorporating variables such as climate, landuse, population, economic growth, etc.

Further, the nature of uncertainty can be categorized into epistemic uncertainty and aleatory uncertainty or stochastic uncertainty (Walker et al., 2003). Epistemic uncertainties stem from our lack of knowledge and they can be reduced with additional collection of data. In contrast, aleatory uncertainties originate from inherent variability and stochasticity of natural phenomena (e.g., climatic variability). Aleatory uncertainties cannot be reduced by collection of more data. For certain natural phenomena, this means that there is no direct way of getting perfect knowledge, given current understanding of the science. Climate predictions over different time scales are perhaps the most common example of aleatory uncertainty in environmental models. Modeling applications typically include both epistemic and aleatoric uncertainties.

The lack of accounting for uncertainties when applying models may result in biased and unreliable results which will directly affect the decisions made based on the modeling results (Beven and Binley, 1992; Refsgard et al., 2007; Bastin et al., 2013). Various methods have been proposed to address the uncertainties from model parameters (Moradkhani et al., 2005), input data (Kavetski et al., 2003), monitoring data (Harmel and Smith, 2007), and model structure (Ajami et al., 2007) in hydrologic and water quality models.

Uncertainty assessment methods fall under one of two classifications: forward uncertainty propagation and inverse uncertainty quantification. In forward propagation methods, uncertainties in model inputs are propagated to the model outputs. In inverse uncertainty quantification methods, posterior distributions of model parameters are derived based on discrepancies between model simulations and observations and values of likelihood function. Inverse quantification of uncertainty is much more complex than forward propagation of uncertainty, as the modeler is essentially solving the problem in reverse (similar to calibration). However, the method provides essential benefits when modeling as in most cases the uncertainties associated with various model elements (parameters, inputs, etc.) are initially unknown and using an inverse approach, the modeler can estimate the most consequential uncertainties, and select them for further evaluation. Thus, these uncertainties can be propagated to simulations through a forward approach. In most inverse uncertainty quantification applications, the overall modeling uncertainties are quantified as a lumped value as quantifying the uncertainties associated with each model components is very time-consuming and in some cases impossible. Specifically, in highly complex integrated environmental models, decomposition of uncertainty and attributing portions of total uncertainty (total error) to various sources of uncertainty is an extremely challenging task which still is a subject of extensive ongoing research (Bastin et al., 2013).

Bayesian-based methods are among the most commonly used assessment techniques for conducting uncertainty analysis for complex environmental models (Jia et al., 2018). Bayesian uncertainty analysis methods, rooted in Bayes' Theorem, quantify parameter uncertainty by deriving the posterior parameter distribution from a combination of prior parameter distribution and a likelihood function. In most environmental models, specifically more complex models, the analytical solution to derive the explicit functional form of the posterior distribution is infeasible. Hence, sampling is often used to derive the posterior distribution. The Markov Chain Monte Carlo (MCMC) sampling schemes provide efficient algorithms to derive the posterior parameter distribution (Rath et al., 2017; Tasdighi et al., 2018). In this regard, multi-chain MCMC methods have proven superior performance and efficiency in sampling the parameter space and deriving the posterior distributions. Application of multiple Markov chains enhances the efficiency of the search algorithm and reduces the chance of being trapped in local optima (Ter Braak, 2006). Two common multi-chain MCMC algorithms frequently used for environmental models are the DiffeRential Evolution Adaptive Metropolis (DREAM) algorithm (Vrugt, 2016) and the Shuffled Complex Evolution Metropolis (SCEM) algorithm (Duan et al., 1992; Vrugt et al., 2003). While multichain MCMC algorithms have been employed in conducting uncertainty analysis for various environmental models, their application to integrated model frameworks remain very limited due to computational burden (Tscheickner-Gratl et al., 2019).

Another source of model uncertainty, discussed much less frequently, is related to human decisions regarding process representation in a model, including the underlying assumptions and prioritization. Although this human-imposed bias cannot be evaluated quantitively, it should be considered as part of a broader validation and peer-review exercise.

4.2.7 Model Sensitivity

Sensitivity analysis explores how changes in model inputs—most generally, boundary conditions, parameters, or configuration (as shown in **Figure 9**)—affect the variation in model outputs. Sensitivity analysis can illustrate which parameters have the least effect on results of interest, and in some cases, may allow for reduction of model complexity, by streamlining process representation. A related concept is uncertainty analysis, where model inputs are presented in a probabilistic form (i.e., as a distribution of values based on current information) to a calibrated model and the effects on model output are evaluated as shown in **Figure 12**. Sensitivity analysis also complements model calibration, which involves selecting parameter values based on the fit between model output and actual observations. Performing sensitivity analysis after model calibration helps to identify which fitted parameters are close to an optimal estimate because low sensitivity indicates high uncertainty in the fitted parameter estimate. Both sensitivity and uncertainty analysis require the running of a model multiple times with a range of inputs. Specific steps for uncertainty analysis are described in the following chapter.

Sensitivity analysis is often used prior to conducting uncertainty analysis to increase the efficiency of uncertainty analysis by reducing the dimensionality of the model. Using sensitivity analysis, the modeler determines which model parameters have the highest impact on simulations (Saltelli et al., 2008). This will help the modeler to decide which model parameters should be included in the uncertainty analysis procedure, thereby increasing the efficiency of uncertainty analysis. Because sensitivity analysis of complex models can be highly computationally demanding, it is a focus of current research to help improve efficiency and applicability (Spear et al., 2020).

Typically, sensitivity methods are categorized into local (LSA) and global sensitivity analysis (GSA) techniques. Basically, LSA methods analyze sensitivity of model responses around some point in input parameter space (ideally around optimal locations), while GSA methods analyze the variability of model responses across the full parameter space. **Figure 13** Illustrates the concept of local and global sensitivity analysis for a model with two parameters. For a model with larger number of parameters, the 2D response surface will change to a more than 2-dimensional (dimension dependent on the number of parameters) response space. Each black dot represents a combination of parameters used to quantify model response and ultimately determine the sensitivity of model response to each parameter.

LSA is a partial derivative-based method to investigate the response of a small disturbance of each parameter around a specific location in parameter space on model output (Baroni and Tarantola, 2014). A common approach for conducting LSA is the one-factor-at-a-time (OAT) method (Yang, 2011). In OAT, one parameter is changed at each iteration. LSA techniques are appropriate for relatively simple models that show linear responses. Although LSA is computationally efficient and popular, it is not suitable for reducing the dimensionality of complex non-linear environmental models as it disregards the correlation between model parameters, and its results are dependent on location and often there is a lack of knowledge on the suitable location, i.e., the parameter true value (Saltelli et al., 2008).

GSA investigates the effect of variations over the entire prior parameter space on model output (Saltelli et al., 2008; Pianosi et al., 2016). A sensitivity analysis approach that is commonly used with GSA is the "All-at-a-time" (AAT) approach. GSA does not have the limitations associated with LSA, as it does not rely on a pre-known optimal location for parameters. A common approach for GSA is rooted in relating the variance of the model responses to the change in input parameters (variance-based techniques). Variance-based sensitivity methods have shown very promising results. However, the sample size required to achieve reasonably accurate approximations can be rather large, which compromises their applicability to highly complex models. Several methods have been proposed to reduce the required number of model evaluations for approximating the variance-based indices. These include: (i) methods using the Fourier series expansion of the model outputs, such as Fourier Amplitude Sensitivity Test (FAST) for the approximation of the first-order indices, and the extended FAST for the total-order indices; and (ii) methods rooted in application of a model **emulator** which will be discussed further in proceeding chapters.

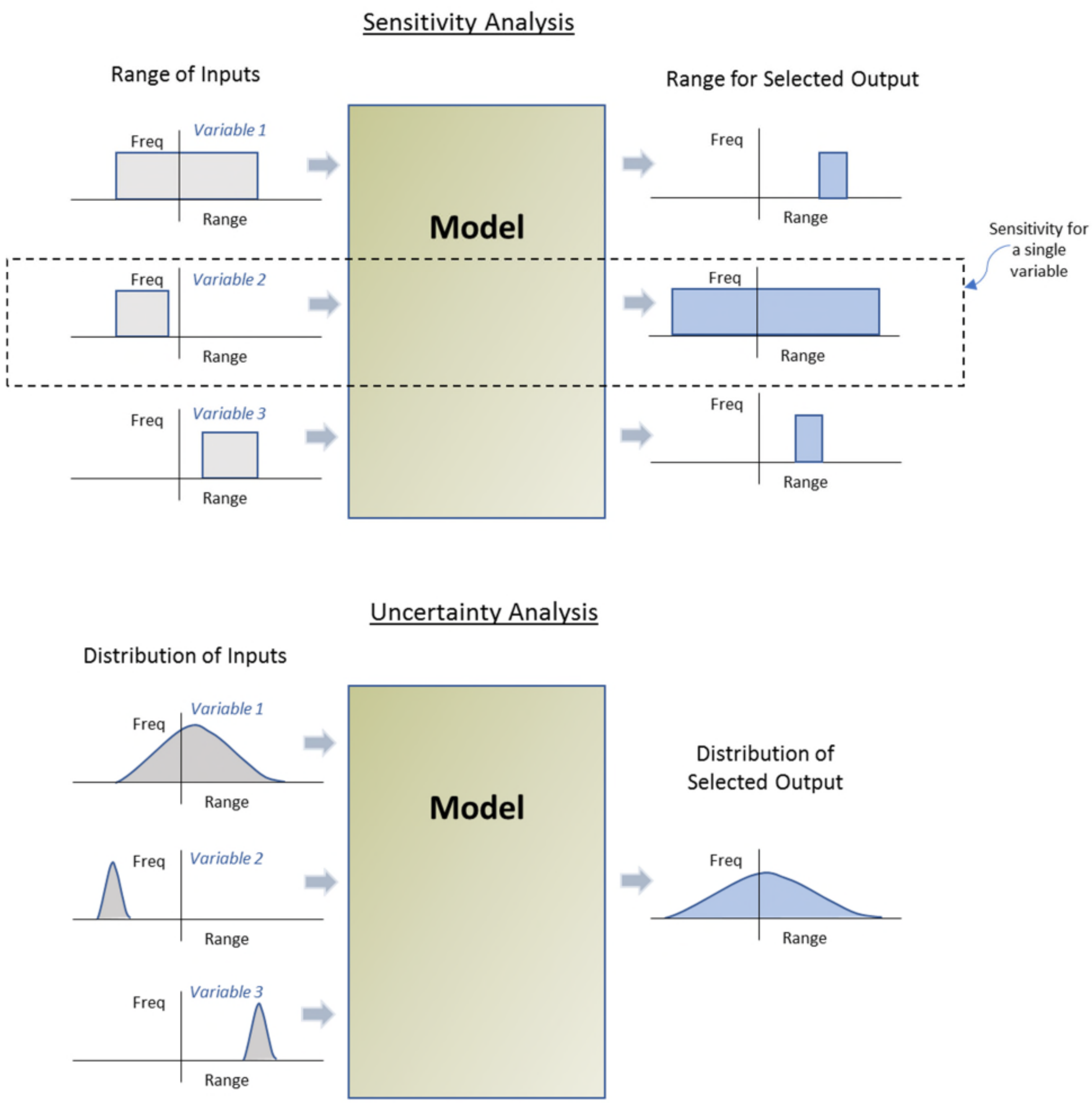

Figure 12. Simplified representation of sensitivity and uncertainty analyses. Inputs in this context may include parameter values, initial conditions and boundary conditions that are used for a single model run. During sensitivity analysis a model is run with a range of values for key inputs and the corresponding range in one or more outputs is evaluated. As part of uncertainty analysis, inputs are assigned ranges in values based on known estimates.

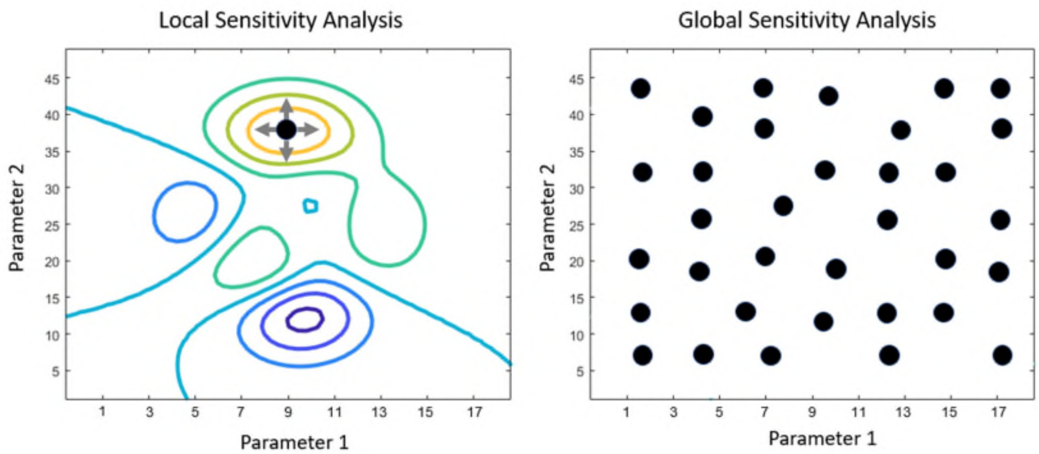

Figure 13. Illustration of the concept of local (left panel) and global sensitivity analysis (right panel) for a model with two parameters. For a model with larger number of parameters the 2D response surface will change to more than 2-dimension (dimension dependent on the number of parameters) response space. Each black dot represents a combination of parameters used to quantify model response and ultimately quantify sensitivity to each model parameter.

4.3 Summary

A model is ready to be applied to explore alternative scenarios once the analysis has been framed and the model has been prepared and evaluated. With regard to model preparation, while most environmental models should be configured and calibrated, the need for data quality assurance may be limited to when processed, clean data are unavailable. With regard to model evaluation, best practices require some form of performance reporting. Model validation, as defined here, is generally desirable. But depending on the model formulation and domain (e.g. biological or economic models), validation often cannot be performed. Analysis of model uncertainty and sensitivity, although providing great insight into model performance, is not often performed (often due to computational constraints). The steps outlined in this chapter are guidelines toward developing credible model studies; however, they are not meant to be applicable to all model studies performed in support of water resources and environmental problems.

5 Application of the Model

The steps in evaluating a model, as laid out in the preceding chapter, are followed to compare model results against recent or past observed data/conditions and thus establish the credibility of the model. Once this evaluation is successfully completed, a model may be applied to answer the questions it was designed for, which often deal with conditions that are different from what is directly observable today. As noted in Chapter 3, it is important to clearly communicate the desired goals of a model study as early as possible and to align model sponsor and stakeholder expectations with the model features being implemented. Assuming that this has been done, there are still some decisions to be made during the application phase of a study, as described here.

At the inception stage of a modeling study, the modeler will typically define the study goals and determine how the study results will be applied. Environmental model applications were previously identified in Section 1.2 under four unique classifications: i) planning and decision support, ii) science support and research, iii) real-time operations support, and iv) dispute settlement support. Special considerations are needed for each of these model applications as presented below.

5.1 Consideration of Generalization During Application of Calibrated and Validated Model

The parameter space of calibration and validation is usually finite, and a function of the data available. This is shown schematically in **Figure 14** and referred to as P1. Once the calibration and validation steps are performed, it is possible that a model can be applied across a somewhat larger range, shown as P2. For mechanistic models, based on underlying physical/chemical principles, it is possible that a model can be generalized with some credibility from P1 to P2. In most cases, empirical models are less credible outside the P1 space. Finally, there is almost always a global space, denoted as P3, over which most models, previously calibrated in P1, may not be correctly applied. In considering model applications to new conditions, it is important to make sure that the model is not falling into the P3 space. This is difficult to know *a priori* but in scenarios where it is apparent that conditions outside P1 are possible, i.e., unusually extreme boundary conditions are present, additional testing of reasonableness of results should be performed. This situation is most likely to occur where the calibration/validation data set P1 is small, and conditions outside it are probable. Similarly, where a large P1 data set is available for calibration/validation, falling into the P3 space is less likely.

5.2 Modeling in Support of Planning and Decision-Making

Applications in this area include, but are not necessarily limited to, support for the development of new environmental regulations (e.g., changes to water quality standards or water supply regulation), support for facility or operational modifications (e.g., changes to reservoir operating rules), support for the design of new infrastructure (e.g., evaluation of new conveyance facilities, dam sites, or groundwater storage and extraction facilities), and support for potential future demand and supply conditions (e.g., changes in future land use and cropping patterns, water supply conditions and sources or climate conditions). For such applications, the model configuration will generally need to be modified for the new conditions being represented. Examples of these reconfigurations may include changing channel properties to reflect a restored section or a new element of infrastructure, such as a flow barrier or a pump-station on a stream, or a new aquifer recharge and pumping facility, or change to an existing landuse, such as changing agricultural land to a suburban division. In each of these examples, the model setup would need to be changed to understand how a response of interest may vary. Thus, following the above examples, the imposed changes may be used to understand the impact on stream flows, groundwater storage, or water quality. Typically, a model may be used to explore a large number of scenarios to compare outcomes. Indeed, this ability to test different scenarios is perhaps the reason for which the modeling may have been undertaken in the first place.

Another aspect of planning scenario development is the consideration of changing conditions (often temporal) over which stakeholders have limited control. Examples include future climatic conditions and extremes, future levels of development and growth, and future regulatory requirements. Scenarios will often be formulated with modified boundary conditions, representing external factors driving a model. Scenarios usually involve assumptions about the future and are thus associated with some uncertainty, and a wide range of alternatives may need to be evaluated.

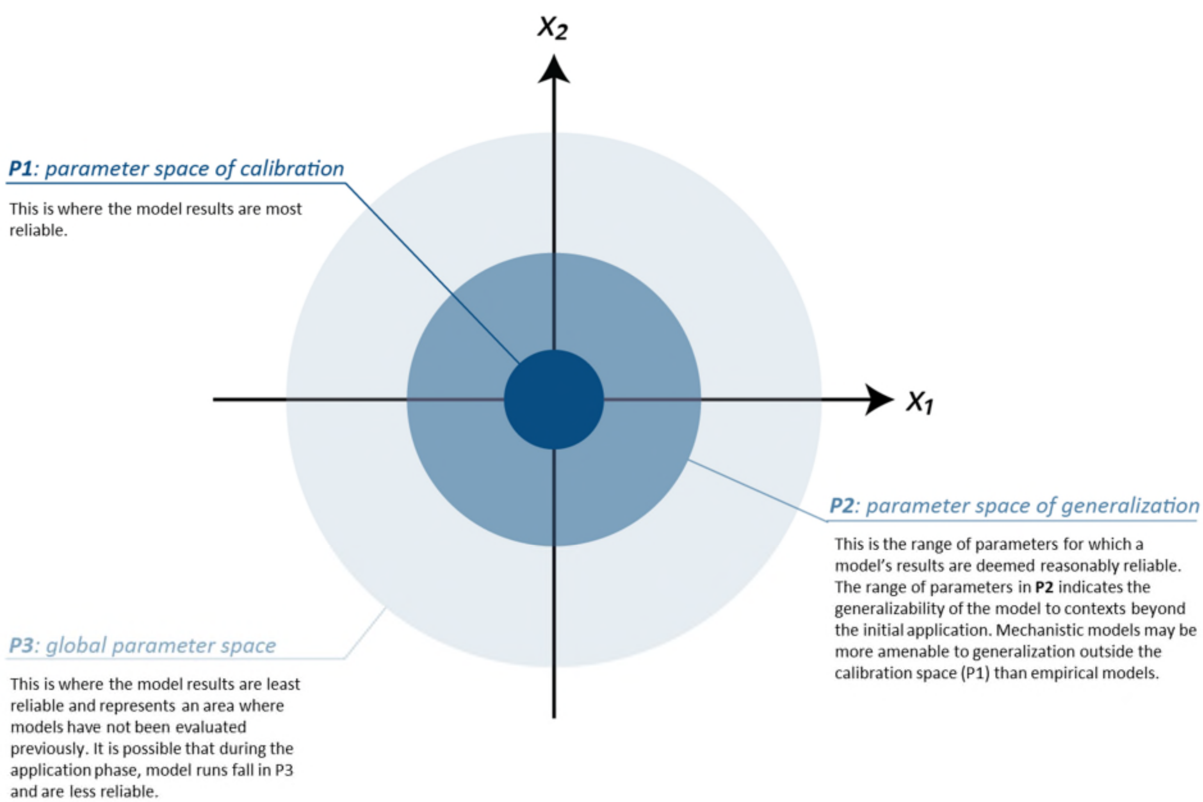

Figure 14. Illustration of concept of parameter generalization and the global parameter space.

In most cases, the results of future scenario runs from a model are interpreted as a change from a baseline condition. A baseline condition may be observed conditions that occurred previously, or a baseline condition may itself be a modeled sequence using different historical boundary values as input. The change interpreted in either case, whether from real past values or a from modeled values, mean different things, and need to be explained adequately to stakeholders.

Scenarios may not always be forward-looking. In many instances, there is a need to understand how a natural system may have behaved in the past, perhaps in the absence of some human modifications that have been made. This could include, for example, the behavior of streamflows prior to the construction of dams on a river system. This may help define reference conditions for future restoration.

Stakeholders will usually have an interest in identifying appropriate scenarios for a study and will be highly engaged in this phase of the work.

5.3 Modeling in Support of Science and Research

Outputs from spatially and temporally resolved models, particularly when the models are developed from basic physical and chemical principles, can be used for gaining a better understanding of a complex system. In such cases the model is run using well-defined boundary and initial conditions, and the model is used to interpolate in time and space using fundamental mass balance and transport equations. Such a model can provide more detailed insight than might be possible using direct observations alone. These models have a benefit in advancing scientific understanding, going beyond the typical use of models in exploring alternative "what-if" conditions. This includes the generation and testing of hypotheses to better understand a particular system, comprising natural and/or human elements. Science support activities include understanding the population behavior of key species, food web interactions, and changes in landscape over the long-term due to human pressures, climatic change, and extreme events. In these applications, the theoretical construction of the model is a tool for improved interpretation of the real world, to be used to provide explanations for real-world observations.

The distinction between planning and decision-support models and scientific models is not a sharp one, and scientific models can also be used for decision support and planning. In many cases, however, detailed models for scientific research are computationally demanding, thus limiting their use in many planning-type studies.

5.4 Modeling in Support of Real-Time Operations

Modeling for real-time support includes calculations to support decisions on the time scales of hours or days. In California, this may include activities such as planning reservoir outflows for flood management and water supply, water exports from the Delta, barrier operations used to manage salinity at various locations, or drawdown effects at a well field or movement of contamination plumes as a result of specific groundwater operations. A key feature of real-time support models is the ability to be conducted rapidly and repeatedly, often many times a day, as new observed input data become available. Although such models must go through the same rigorous framing process to be credible as described in Chapter 4, once adopted, there is a greater focus on model efficiency. For a model to be useful in a real-time setting, it needs to provide responses quickly and be set up in a way that input and output data can be transmitted efficiently and often times with automated processes, such as Supervisory Control And Data Acquisition (SCADA) systems for rainfall, streamflow, groundwater levels, or operations data.

Also, models used for operations have a need to develop forecasts looking forward in time (on the time scale of hours to months), and in such instances there is a need to forecast boundary condition values at future times. For evaluating the results of such modeling, it may be useful to discern the source of the uncertainty in the forecast, which may arise from boundary condition uncertainty and model uncertainty.

5.5 Modeling in Support of Dispute Settlement

Model studies for dispute settlement are called out separately because of their common use in California water resources applications, such as for water rights adjudication or allocation of water among different types of beneficial uses. Models used in this application are not fundamentally different from those used in other applications; therefore, they should be developed and tested with the same scientific rigor. Because these models are specifically developed to address a legal issue, it is critical that the modeling question (as described in Chapter 3) be carefully addressed at the study inception. Given the additional scrutiny on modeling results during a legal process, it is perhaps more important to stress all of the framing issues that are presented in Chapter 4. Similarly, these model studies call for a high standard of documentation to fully and transparently explain the approach and assumptions used.

5.6 Post Audit after Application: Compare Model Results to Future Data Being Collected

For model applications that are used to make near-term forecasts or longer-term predictions, it is important to revisit model outcomes and to compare field observations with previously made model predictions. This process is termed a post-audit, and its importance has been highlighted in other modeling guidance as well, notably, the *Guidance on the Development, Evaluation, and Application of Environmental Models* (Gaber et al, 2009). For major models that are typically in use for a decade or longer, and where supporting observed data continue to be collected, a post-audit is not very difficult to implement. A post-audit can provide insight on conditions under which model performance was acceptable and in line with prior calibration history, thus providing credibility to the model and related modeling studies. A post-audit may also result in the opposite outcome. Under conditions where model performance was poorer than expected, the post-audit provides an excellent opportunity to revisit the fundamental conceptual framework and/or the model calibration. Indeed, a post-audit can provide an excellent basis for future model improvements.

6 Communicating and Documenting Results

Effectively communicating the results of a modeling study is a critical step in ensuring the success of a project. It is crucial that sufficient resources (time and money) are set aside at the outset of a modeling study so that proper documentation is not neglected as a study is nearing completion. Communication of model findings should satisfy the information needs of different audiences, from technical specialists to members of the general public. Different types of review beyond public participation, such as technical advisory committees, shared-vision modeling, and peer review, should be considered where appropriate.

6.1 Presenting Results

Model findings will be used by and will need to satisfy the information needs of different audiences, from technical specialists to members of the general public. Therefore, it is important that modelers are also engaged at different levels of this process such that the right information is transferred to each audience. Furthermore, audiences may weigh in on a modeling study during various phases of the project. Considerations for different audiences at project inception and completion are described below:

- At project inception:
  - Technical specialists. Such audiences will need to understand why the modeling is needed and the approach to be used. Some of the items in the project inception checklist may serve to aid this goal (see **Checklist 1**). It is also important to convey the uniqueness of a modeling exercise, and how it extends current thinking.
  - Stakeholders. Such audiences will need to know the specific results and answers to be obtained through modeling and whether similar answers can be developed without modeling. They will need to know the costs, time frames, and major unknowns. Modelers should use this opportunity to highlight known and potential uncertainties, and how this might affect the outcome of the findings. Conceptual frameworks can be used as a tool to highlight the areas of focus of the modeling exercise. Stakeholders need to receive clear definitions and as little jargon as possible.
- Near completion:
  - Technical specialists. Such audiences will expect to see many of the technical steps, such as the basis of the model, specific assumptions used, the results of testing and evaluation of uncertainty. It is also important to convey the novelty of a modeling exercise, and how it extends current thinking. The peer-review and publishing of key model studies provides additional credibility and also provides archival benefits for a modeling exercise. Finally, audiences may want to understand next steps or the long-term plan for the study, such as additional modeling or data collection.
  - Stakeholders. Such audiences need a high-level overview of key findings that can be quickly understood across a broad range of people, expertise and experience. Good results are simple and memorable and tell the key elements of a story in a compact manner. A new or updated

conceptual framework is a good summary of the overall exercise. Additional graphical resources, beyond the conceptual framework, may also be developed to help readers understand important findings. It is helpful to describe what was achieved through modeling and what remains unknown.

6.2 Documentation

The preparation of model documentation is an essential step in model development and use. Often, model frameworks are developed and maintained over years, sometimes by different individuals or teams with changing member composition. Good model documentation should serve the needs of developers and users, and may be accomplished by using the same set of documents. Ideally, documentation should be prepared in a manner that contains enough information to allow for the long-term evolution of a model, both within the organization and external to it. From the perspective of external users in particular, documentation should explain the basis of the model and its use, including how key input variables are selected. Such documentation should include representative input files and result files to allow a user to reproduce a basic set of scenarios.

In the case of a specific application, documentation needs to explain the best practice elements that are outlined in this chapter, including, model purpose, input data used, calibration approach, model evaluation, and model results in the context of the intended purpose.

Writing model documentation is an essential step in model development. However, under short timelines and tight budgets, preparing documentation may become a low priority, particularly for models developed for a specific application with no expectation of re-use. Missing, inadequate, or out-of-date documentation is a barrier to model acceptance and may result in duplication of effort because a potentially suitable model may be overlooked for use in a later modeling study.

Documentation can be broadly classified as internal and external. Internal documentation is generally embedded within the code in the form of function descriptions, code comments, etc. Internal documentation is important and should follow the conventions of the programming language(s) used to build to the model. External documentation is generally written for three audiences: (1) the developer(s) building, maintaining, and updating the model, (2) other modelers interested in the details of the model, including those interested in integration, and (3) users of the model with lesser need to understand the inner details of the model.

Writing documentation that is easily understood across a variety of audiences is one of the challenges for model developers. The following elements are recommended to produce good documentation:

- A general description of the model that includes the modeling goals and the scope of the model.
- A point of contact for the model and information about how to get started using the model, including download links, installation instructions, hardware requirements, and licensing costs (if applicable).
- Assumptions and limitations of the model.
- Model relationships and mathematical methods used, including a graphical conceptual model.
- Data used to inform model relationships and input data requirements to run the model, including example input files.
- Model output format(s), including example output files.

- Representation of uncertainty.
- Availability of tools for conducting sensitivity analysis, post-processing results, etc.
- Table(s) with all model parameters and their default values.

A recent set of documents prepared for the California Central Valley Simulation Model (C2VSIM; provided online at https://water.ca.gov/Library/Modeling-and-Analysis/Central-Valley-models-and-tools/C2VSim) is an excellent example of documentation addressing most of the questions above, and serving a range of audiences.

6.3 Review of Model and Study

Most models undergo some form of internal review within an agency, firm or workgroup. For model studies that are used to make major decisions, or where general-purpose frameworks are being developed, a rigorous, and ideally, public review process should also be adopted. This review provides for an independent evaluation of a study and provides confidence in its findings and should be done in a manner that allows time for the review and for the modelers to adequately respond to comments. For this reason, it is important to allocate time for this phase of the study early in the scheduling process.

6.3.1 Public Participation

Interested and affected stakeholders, agencies, organizations, and individuals should be provided opportunities to participate throughout the modeling process. A wide variety of ways exist for agencies and consultants who conduct modeling studies to effectively communicate their modeling results and incorporate the ideas and comments of others into their work. Efforts to secure public participation should be pursued through public workshops, meetings, scientific/technical conferences, and technical advisory and citizens committees. A project website and list of interested parties should be maintained and used to communicate steps in the planning process. Where appropriate, specialized outreach such as language translation services should be conducted to engage members of disadvantaged communities (DACs). The lack of stakeholder and decision-maker communication as well as insufficient public participation in the water planning process can cause water projects to fail, illustrating the importance of investing resources in a robust public participation process from the early stages of a water resources planning project.

6.3.2 Technical Advisory Committees

Technical advisory committees (TACs) are often formed to provide ongoing review for modeling and planning studies. TACs typically come in two forms, as a committee of technical people representing stakeholders or as a committee of recognized independent technical experts. A TAC consisting of technical stakeholder representatives is usually formed to do the following (CWEMF, 2000):

- Ensure that local and diverse expertise is used to address the problem. Often, the entities involved in a problem have different special expertise relevant for a modeling study. Having technical representatives from each knowledgeable entity helps to make this expertise available for the development and application of models.
- Enhance communication. Enhanced communication allows TAC members to become familiar with the details of a modeling study, which should reduce stakeholder misunderstandings of the model

and model results. Ultimately, this should help build stakeholder confidence in model and planning study results.

- Help model results be relevant for a wider range of interests and problems. A major model or modeling study will have implications and applications for many entities in a region. Thus, many entities will seek to use or modify the model to enhance their own understanding or for their own purposes. If a single model development exercise can support these broader interests, the regional interest is served.
- Provide local experts a structured opportunity to contribute ideas and concerns. This is a very local form of "peer-review," occurring early in the modeling process.

The second type of TAC, a committee of recognized independent technical experts, can also have several uses. Technical experts independent of stakeholder interests can provide a form of technical arbitration on any controversial issues and may suggest additional approaches to address such problems. In addition, the credibility of model and planning study results can be enhanced by the involvement of recognized technical experts.

6.3.3 Shared-vision Modeling

A form of stakeholder involvement in which stakeholders and decision makers are involved from the outset is termed 'shared-vision modeling' or 'collaborative modeling'. The fundamental concept is that those affected by water resource modeling should be provided the opportunity to participate in model design, development, evaluation, enhancement and use. In shared-vision modeling, the model is typically developed by a single neutral entity with very close coordination by technical representatives from each stakeholder group.

Shared-vision modeling, like other consensus building processes, requires that strong motivation exists among the stakeholders to develop a consensus (Walters, 1997). Model development will progress much more slowly than if performed by a single group (Lund and Palmer, 1998). However, if the modeling and negotiation steps are considered as one extended process, shared-vision modeling usually saves time in the long run. If participants can arrive at agreement on what is contained in the model, then later efforts can focus on meaningful discussions among stakeholders and interpretation of the results, rather than negotiations about model content.

6.3.4 Peer Review of Model

Peer review is the process of soliciting input from experts who are not involved in a particular study but are familiar with the general topic. Peer review should provide timely, open, fair and helpful input and should ideally occur at various stages of the modeling life cycle, including conceptual framework development, model implementation in code, and model application to specific geographic area or problem. Engagement of the peer reviewers early rather than solely near the end of the project can allow for adaptive corrections of the modeling study.

Peer review is most helpful when the following conditions are met: i) peer review is conducted in an atmosphere of transparency, collaboration and shared sense of purpose; ii) the review team reviews the source material and modelers' responses to their comments; iii) adequate time and funding is budgeted for review; and iv) the review team contains some interdisciplinary membership to allow for a broader evaluation of basic assumptions and utility of the exercise. When conducting peer reviews, it is important to screen for potential conflicts of interest, and to select reviewers who will be independent and have prior minimal connection with the study. For a complex project, a review team may need to include some persons who are intimately familiar with the project who can provide needed information to the independent reviewers. If a sincere commitment to obtaining constructive feedback is not made through the above steps, there is a risk that peer review becomes more of a rubber-stamp than a positive contribution to a modeling study. Usually, adequate peer reviews of the study models can increase acceptance of a project.

We recommend that most complex and consequential model studies be subject to peer review. Usually such reviews are conducted by the organization sponsoring a model study. For newly developed model frameworks, the process of anonymous peer review required by scientific journals serves as the touchstone for validation of a modeling study and is also recommended. Some major studies, especially those with major societal and ecological implications should also be subject to a "deep" peer review. This is not a standard term in the literature, but we use it here to refer to a peer review where reviewers are given the time and potentially funding resources to delve deeply into the concepts and application of a modeling study. Such deep review is unlikely to occur when reviewers are limited to basing their judgment on published documentation only.

CWEMF has developed a process for peer reviewing models[6]. These peer reviews are not intended to be "stamps-of-approval" for particular models or to disapprove of models. Instead, they are intended to inform stakeholders and decision-makers of (1) whether or not a given model is suitable for intended applications, and (2) the temporal, geographic, or other limits on the use of the model. In general, CWEMF's peer review process follows the eight steps outlined below:

- Model specification
- Obtain funding
- Select reviewers
- Define scope
- Assemble model, data, and documentation
- Conduct initial review
- Prepare draft report
- Prepare final report

6.3.5 Reproducibility

Ensuring the transparency and reproducibility of a model can be considered an extension of the peer review process. One way to ensure model reproducibility is to provide all modeling input files so that the model can be re-run by another user. Or, the input data itself along with other model configuration data could be provided so that the user can re-create the model application using a different model framework. The objectives and phase of the modeling study or the needs of the model sponsor may dictate whether model reproducibility is appropriate. Even for some public agencies, there is no need or intention for the model to become public if the study is part of litigation. If the model is part of a

[6] More information is available at: http://www.cwemf.org/Pubs/CWEMFPeerReviews.pdf

preliminary study, it may not be appropriate to release the model for review until later stages of the study.

Transparency and reproducibility can be limited by several considerations. Information on a particular model can be scattered over several sources, some of which may not be readily available to model users. In some cases, model documentation may not provide a sufficient description of the model or the calculation of model results. Furthermore, results presented in scientific publications may not correspond with parameter values or equations in the model source code. The shifting change in attitude and growing awareness towards the need for greater transparency is evidenced in part by the increasing publication of data and supporting materials on scientific journal websites (De Vos et al., 2011).

As for many phases of the modeling study, adequate model reproducibility can be ensured if the project allocates adequate resources for this phase of the modeling study. Allocating enough resources for reproducibility and other stages of the model study can increase acceptance and utility of a modeling effort.

7 Encouraging Collaboration in the Modeling Community

There are distinct individual and institutional roles in a larger modeling community: individuals/teams who develop and maintain specific models; individuals/teams that apply existing models to specific situations; individuals/agencies who direct and use model results and drive the need for integration across disciplines, but are not directly involved in running models; and other stakeholders who are affected by model outputs in some form. Modelers in different domains interact with one another and are typically aware of each other's needs, even though there is not one top-down model structure that everyone adheres to. Engaging this community's shared focus around important challenges can be accomplished with various approaches listed below.

7.1 User Groups

Model user groups typically focus on problem solving and development issues related to specific high-use models. The formation of additional user groups to support high-use models or domains would benefit model development and user training in much the same way as existing user groups have. Some currently active user groups are identified below. Except for WEAP, these communities are for models specific to California.

Delta Modeling – The Delta Modeling User Group (DMUG) was created by and receives ongoing support from the California Department of Water Resources (DWR) to facilitate the exchange of ideas and problem solving around the use of Delta hydrodynamic models. The user group is open to any interested parties and holds meetings three times a year. The website archives meeting presentations, notes, and annual newsletters. https://water.ca.gov/Library/Modeling-and-Analysis/Bay-Delta-Region-models-and-tools/Delta-Modeling-User-Group

Integrated Water Flow Model (IWFM) – This user group, hosted by DWR and the United States Bureau of Reclamation (USBR), focuses on the development and understanding of the IWFM and IDC models. The group holds quarterly meetings with its records archived on the California and Environmental Modeling Forum (CWEMF) website. https://water.ca.gov/Library/Modeling-and-Analysis/Modeling-Platforms/Integrated-Water-Flow-Model

Water Evaluation and Planning (WEAP) – This online user group was created to support WEAP model implementation. With over 30,000 members and many thousands active on forums, the website provides a virtual community for the model in addition to tutorials and user manuals. https://www.weap21.org

Groundwater Exchange —This online resource is a community/information site to share information related to the implementation of the Sustainable Groundwater Management Act (SGMA), including planning documents, data, and models. https://groundwaterexchange.org/

7.2 Virtual Community of Practice

In California, the “virtual” or online community provides a vast network of development and support for modelers. Online forums and user groups have filled the local gaps in technical support and many regional models have roots in the broader modeling literature and community. Additionally, the virtual community has benefited from online resources such as code repositories and cloud storage and computing. Cloud storage and sharing, such as **Box**, **Dropbox**, **SharePoint**, and **Google Drive**, have also allowed for more efficient transfer files and data and collaboration. Transparency and open communication about models have been enhanced through the use of these storage and sharing tools and continue to be utilized by regional modelers. In the future, integrating existing virtual infrastructure utilized by the modeling community will facilitate efficient engagement.

Modern software development in general—even outside of scientific modeling—has taken on an increasingly distributed, collaborative paradigm. The use of online development platforms like GitHub has become a very common way to distribute open-source software for projects of all sizes, ranging from small specialized packages of just a few files and dozens of lines of code to entire programming languages central to modern scientific computing like Python.

These types of development platforms incorporate features like version control for project source code and data files, integrated issue/bug tracking, automated build and testing processes, and project wiki pages. Scientific models that want to follow trends in best practices from the broader field of software development will consider using these types of structured, collaborative coding tools. Even for projects that cannot be open-sourced either temporarily or permanently, similar enterprise-focused workflows exist to facilitate a collaborative software development process among a closed group of users.

An important challenge that should be addressed for all virtual communities is continuity and retention of information. In most cases, information is lost upon completion of a project or when an immediate need is met. Implementing processes for managing and archiving information for future use will require dedicated staff time in an organized framework.

8 Emerging Technologies Supporting Model Development

Model developments are dependent on computer technologies whose rapid evolution has society-wide impacts. The purpose of this chapter is to provide a snapshot of technologies and open-source and proprietary tools that appear promising in the domain of water and environmental modeling. Emerging technologies are described across six key areas: (1) innovations in data capture, (2) data analysis frameworks, (3) machine learning frameworks, (4) data visualization and communication technologies, (5) tools for organizing workflows, and (6) new methods in software engineering and architecture. Even though these technologies may provide future modelers with significant new capabilities, the general sequence of steps for performing robust modeling as outlined in Chapters 3 through 6 are still expected to apply.

8.1 Innovations in Data Capture

With breakthroughs in data gathering frequency and data storage capacity, the future of modeling is increasingly shaped by how the modeling community wishes to deal with large volumes of different types of data. In general, the data that are commonly used in water and environmental modeling are observed and gathered through remote sensing, earth monitoring systems, field campaigns, citizen science initiatives, digitized historical data, and even other model outputs (Blair, et al., 2019).

Several concerns arise when there is a plethora of data sources that are available to be applied to a single modeling question. First, the heterogeneity of the sources of data makes bridging the various data types a complex task as there may not be a one-to-one pairing of data points between collection methods. Additionally, each method offers its own measure of accuracy and precision. While the combination of multiple sources of data helps to bridge the gaps in observations and to address the shortcomings of a particular method, one should be mindful of the uncertainty in and degree of quality control as this is manifested in the credibility of the model results. There is both need for and motivation to traverse through many forms of data and to improve interoperability across data sets. To do this, it can be helpful to first identify the various features of data sets and the methods with which we can better interpret them.

8.1.1 Geospatial Data

As studies of environmental phenomena are to be applied in space and time, it is important to synthesize the varying spatiotemporal resolutions of each of the various sources of data to develop an understanding of patterns and dependencies within the area of interest. Advancements in remote sensing technologies, the ubiquity of mobile platforms, and the expansion of *in situ* sensor networks have contributed to a data pool of growing space and time complexity.

In the field, mobile platforms provide new interfaces for models, especially where location-specific data, model inputs, or model output information can be provided with geospatially stamped information. Thus, instead of imagining model outputs as static documents, they may be served online to support activities in the field and may encourage a closer integration of desktop-based modeling and field activities. This may be especially relevant in complex urban settings, related to stormwater management for example, where data on ground conditions can be input directly via mobile devices or key outputs can be viewed while in the field, more efficiently than is currently done. Observations can then be collocated with other data types by matching them in space and time.

Geospatial data not only allow patterns to be visualized but also provide the structure needed to identify and analyze covariate effects between variables and to develop a more physically realistic model. By anchoring data in space and time, geo-referenced models develop an understanding of distance, proximity, contiguity, affiliation, co-occurrence, dependence, and segmentation – patterns which were otherwise masked in the rows and columns of traditional spreadsheets and databases. **Figure 15** illustrates some of the relationships that could be hidden in geospatial data. The search for these patterns in geospatial data is supported by the development and improvement of spatial analysis techniques including overlay analyses, spatial interpolation techniques, geostatistical Gaussian processes, and multivariate statistics, all of which are available on open-source and proprietary geographic information system (GIS) software. The key features of some notable GIS software are summarized below.

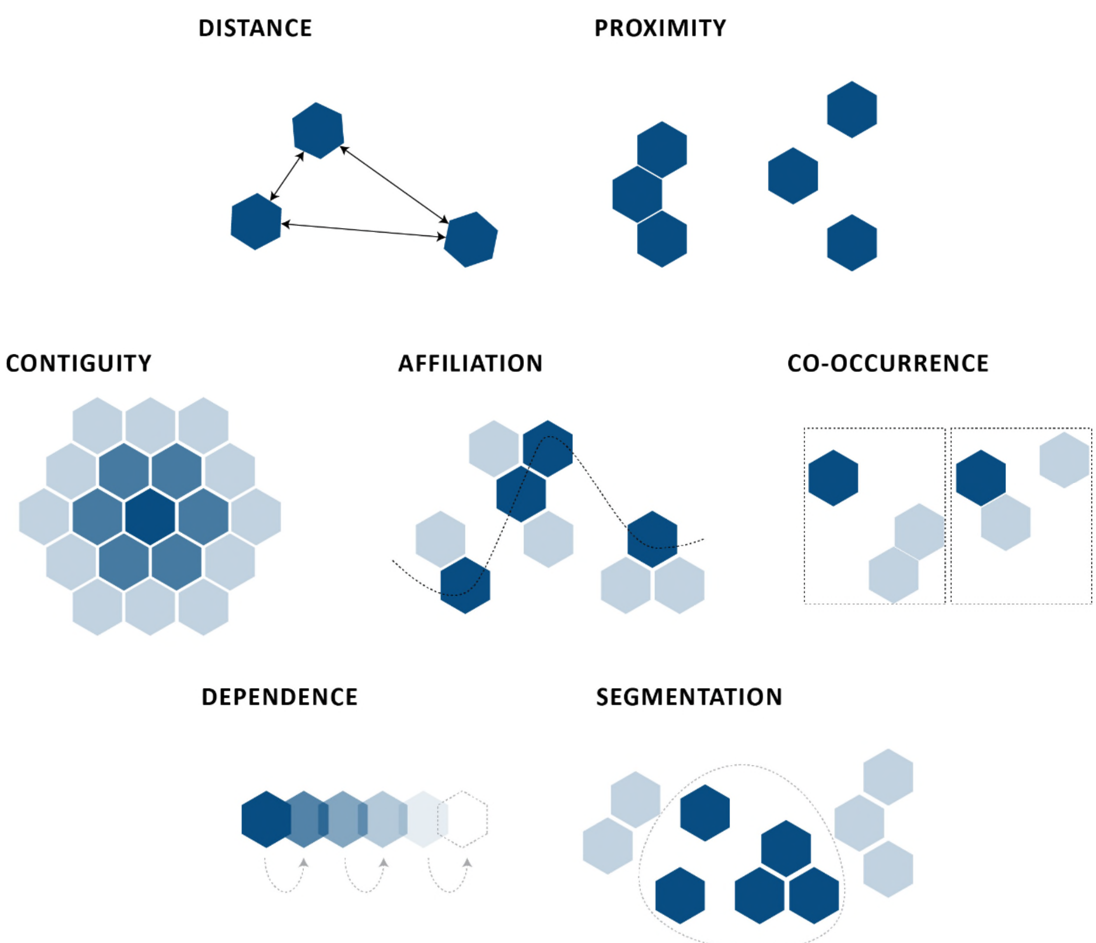

Figure 15. Illustration of spatial relationships such as distance, proximity, contiguity, affiliation, co-occurrence, dependence, and segmentation.

- ArcGIS is a mapping and spatial analytics platform that offers imagery tools to create smart 3D models, developer tools to create custom web, mobile, and desktop applications, APIs for Python and SQL queries, and a database of basemaps, imagery, and authoritative maps on a multitude of topics. ArcGIS can be deployed on its own highly scalable cloud infrastructure, on cloud platforms such as AWS, Microsoft Azure, or on a local distributed system.
- QGIS offers many of the basic capabilities that are available on ArcGIS. For more experienced GIS analysts, the limited functionality of QGIS can be improved upon through the extensive plugin architecture and libraries. Modelers can even code new applications in C++ and Python.
- Google Earth Engine is a platform which integrates custom algorithms with satellite imagery and geospatial datasets. It offers APIs for coding in Python and JavaScript, code editor tools for interactive algorithm development, and integration with Google's cloud services such as Google Compute Engine and Google Cloud Machine Learning. In addition, Google Earth Studio, an animation tool, and Google Earth VR provide the visualization tools needed for a multi-perspective survey of a model environment which can then be published on web and mobile platforms.
- Golden Software is a scientific graphics development kit with tools to deal with the many dimensions of geospatial data, including: Surfer, to explore and analyze geospatial data; Grapher, a data presentation and plotting toolkit; Strater, to integrate subsurface, cross-sectional information; Voxler, which offers 3D modeling and spatial analysis capabilities' MapViewer, a mapping and spatial analytics package' and Didger, for geoprocessing static data formats to create dynamic data with adjustable projections.
- Earth Genome serves public and private geospatial research, providing extensive datasets from a myriad of sources, microservices that improve data traversal, decision making tools and applications to analyze and interpret the data, and satellite imagery to encourage monitoring and reporting. GRAT, the Groundwater Recharge Assessment Tool, is one such cloud-based application that has informed planning and decision making in California and beyond.

Geospatial data and modeling offer an opportunity to more effectively communicate model results with stakeholders and sponsors. They can be seamlessly integrated with augmented reality and 3D imagery platforms to encourage engagement and to communicate model structure, organization, and results in a more intuitive manner. The growing development of multivariate spatiotemporal methods of analytics beyond clustering, propagation, interpolation, and extrapolation suggests that geospatial data will continue to play a key role in model development.

8.1.2 Open Data

Citizen science initiatives increase the availability of geospatial data and inform natural resource management, environmental protection, and conservation efforts all while fostering public engagement and contribution (McKinley et al., 2017). Currently, real-time air quality monitoring using "Internet of Things" sensors is setting the precedent for large scale open data collection, transmission, and analysis. In California, where wildfire-induced air quality changes have a severe impact on human health, home based air sensors have gained popularity. These personal laser particle counters measure particulate matter concentrations in real-time and use WiFi to project this data on a publicly accessible map.

When effectively designed, standardized, and thoroughly evaluated, the open data that is gathered by citizen scientists can be of high quality and resolution and can circumvent the deployment of expensive and vulnerable sensors. However, the reliability of such data is often questioned as it can be difficult to impose quality control and ensure the proper data collection protocol is being followed without being present at the time of data collection.

Yet, data collected by citizen scientists have been utilized by the EPA, the Smithsonian Environmental Research Center, and the Earthwatch Institute. According to the EPA, by circumventing time, geographic, and resource constraints, crowdsourced data can leverage a large, untapped network of people while demanding minimal resources such as technical support and equipment training to support data collection. The Virginia Department of Environmental Quality, for instance, earned more than 275% return on investment on resources spent on volunteer water monitoring (EPA, 2019). The EPA also supports Georgia's Adopt-A-Stream program, a volunteer-based initiative to sample local waterways, the results of which have been published in a handful of scientific journals.

A belief in the use of crowdsourced, open data is that by increasing the availability and accessibility of tools, the sampling frequency and sample size can increase. Once a sufficiently large sample size is obtained, statistical methods can be used to isolate and remove outliers and gain a more accurate distribution of the data without necessarily engaging in rigorous quality control measures for each data point. There are several mobile applications that have been created to automate data collection and to deviate from the use of traditional equipment and sensors which are prone to malfunctioning and calibration errors. Some examples include:

- mPing, which was created as part of the Precipitation Identification Near the Ground project and enables citizen scientists to report on precipitation such as rain, freezing rain, drizzle, snow, ice pellets, and mixed rain. This information is used by the National Severe Storms Laboratory to corroborate observations from radar detections, calibrate and correct discrepancies in satellite retrievals, and inform the future development of precipitation prediction technologies. The scope of this study was recently broadened to incorporate reports of wind damage, tornadoes, flooding, landslides, and visibility changes.

- Marine Debris Tracker and Creek Watch, which allow users to take a picture of and report on waterways. These platforms are used to monitor the health of waterbodies by categorizing trash and debris sightings according to a preloaded list. This information can then be used by watershed groups and water management agencies to allocate resources towards rehabilitation and clean-up.

- NOAA's Water Level Reporter, which allows citizen scientists to submit water level reports that are accompanied by geo-referenced pictures. Here, the photographs serve as evidence of the visual conditions and can be used to standardize qualitative field observations using simple cross-referencing methods or more advanced machine learning tools. This information is used by NOAA to study flooding and to communicate its impacts on a more precise scale.

In California, examples of water projects which have benefited from citizen monitoring include:

- SWAMP, Surface Water Ambient Monitoring Program, has a Clean Water Team volunteer water quality monitoring program. Volunteers are trained to be technical assistants, to support quality assessments, and to contribute to guidance documents that are curated by watershed stewardship organizations.

- Since 1988, the Elkhorn Slough National Estuarine Research Reserve, along with other foundations and agencies, has sponsored a water monitoring program that consists of twenty-six stations at which volunteer citizen scientists collect data on temperature, salinity, dissolved oxygen, pH, turbidity, nitrate, ammonium, and dissolved inorganic phosphate. The data is presented on an interactive report card and has supported a number of publications.

8.1.3 Non-numerical Data

From the discussion of open data, it is apparent that pictures and videos constitute data that can be collected to support more traditional numerical data. For instance, precipitation measurements that are obtained from a network of *in situ* sensors can be supplemented by satellite imagery and even videos taken by drones to better understand the spatial extent of riverine flooding and the resulting flow of water. In this example, *in situ* sensors offer a discrete, numerical snapshot of the event while satellite images can be stitched together with precision for a more continuous view of the spatial domain. Additionally, drone videography can provide continuous data in the temporal dimension for further analysis.

In general, media comprising digital images, videos, and audio files, are unconventional yet useful sources of data that can be utilized in modeling. Novel image analysis and change detection techniques can be used to extract information from digital images and from videos containing radar and sonar data. Currently, innovation in this area is largely driven by the need to precisely extract information from biological microscopy image sets; however, the tools that have been developed can still be applied to environmental and water modeling. Some relevant examples of image analysis software and their environmental applications are listed below.

- Resonon, which offers hyperspectral image analysis capabilities, offers image analysis software for both static (mounted or benchtop) and dynamic (drone-based) images. Coupled with Machine Vision software and a C++ software development kit, Resonon's toolkit supports data processing, visualization, and custom plugins and has been used in hyperspectral imaging of algae blooms in Lake Erie and detection of nitrates, phosphates, and sediments in the Mississippi River. Additionally, Resonon provides cameras and supporting apparatus to optimize data collection and transmission in the field.

- VIAME, Video Image Analytics for Marine Environments, is an open-source framework that contains workflows for object detection, classification, and size estimation models. It was developed in cooperation with the National Oceanic and Atmospheric Administration (NOAA) and is commonly used for fisheries stock assessment.

- Malvern Panalytical, is another company that has developed a spectral image analysis tool. Their static image analyzers can determine particle size (length, width), shape (circularity, convexity), and transparency and characterize both spherical and irregularly shaped particles in either dry or wet suspension. The findings are then statistically presented in particle size distributions that can be cross validated with alternative particle sizing approaches. Raman spectroscopy can also be applied to the samples to chemically identify particles and foreign substances. Typical applications of Malvern Panalytical's image analyzers and other tools within the environmental realm include water treatment, microplastics characterization, and soil analysis.

Examples of open source computer vision software libraries and tools include: OpenCV, GPUImage, and ImageJ. They are suited to extract information from both images and videos. Alternatively, more established computer vision services and APIs provide baseline functions and tunable algorithms. Some examples include:

- Clarifai, an image and video recognition service with features such as image and video processing, tagging, similarity searching, and a customizable detection and categorization algorithm.
- Google Cloud Vision, an image analysis API with label and landmark detection.
- Amazon Rekognition, a deep learning-based image recognition and analysis platform with object and scene detection that can be integrated with Amazon Web Services (AWS) to add tunable functionalities.
- Microsoft Cognitive Service, a Computer Vision API with motion tracking, image tagging and categorization, line drawing detection, and thumbnail generation capabilities.
- IBM Watson Vision Recognition Service with image detection, image class taxonomy and description, and image matching with a confidence interval.

Further technological advancements in digital imaging, videography, and audio recording can be harnessed to supplement traditional, numerical data and provide a more holistic observation.

8.2 Data Analysis Frameworks

Data analysis is the process of systematically applying statistical and/or logical techniques to describe, condense, illustrate, and evaluate data. Data analysis and integration frameworks can be used as comprehensive tools to manage model input and output and display results. Commercial tools for data analysis and integration include Tableau, Qlik, Palantir, and Matlab. The programming languages R and Python are the most widely used non-commercial, or open source, programming environments for data analysis and graphics. These frameworks allow for integration of technical code and provide a means for managing the flow of input and output files. Data or model results can be tabulated or visualized by model stakeholders through the use of "data dashboards", some of which can be freely published on the internet.

According to Shamoo and Resnik (2003), data analysis procedures "provide a way of drawing inductive inferences from data and distinguishing the signal (the phenomenon of interest) from the noise (statistical fluctuations) present in the data." Technological advances are driving exponential growth in volume and speed of data generation, giving rise to the concept of "Big Data". Big data, although informal in origin, has come to serve as a term to describe data that are high in volume, velocity, and variety, requiring new technologies and techniques to capture, store, and analyze.

In the environmental modeling realm, the big data concept primarily pertains to techniques to capture, process, analyze, and visualize large datasets in a reasonable amount of time. When analyzed properly, big data can enhance decision making, provide insight and discovery, and support integrated model applications.

Another approach that has potential is the use of data-driven (i.e. black box) models with process-based models, building on the strengths of each modeling methodology. Big data analysis tools can be used to reconcile the strengths of black box and process-based modeling approaches and may allow mixing of models with different levels of information (Figure 16). The inter/multi-disciplinary nature of the integration problem necessitates the merging of large, disparate datasets (model inputs/outputs) which eventually should be analyzed to make inferences about the system being modeled. These inferences can be guided by process-specific knowledge in such hybrid models.

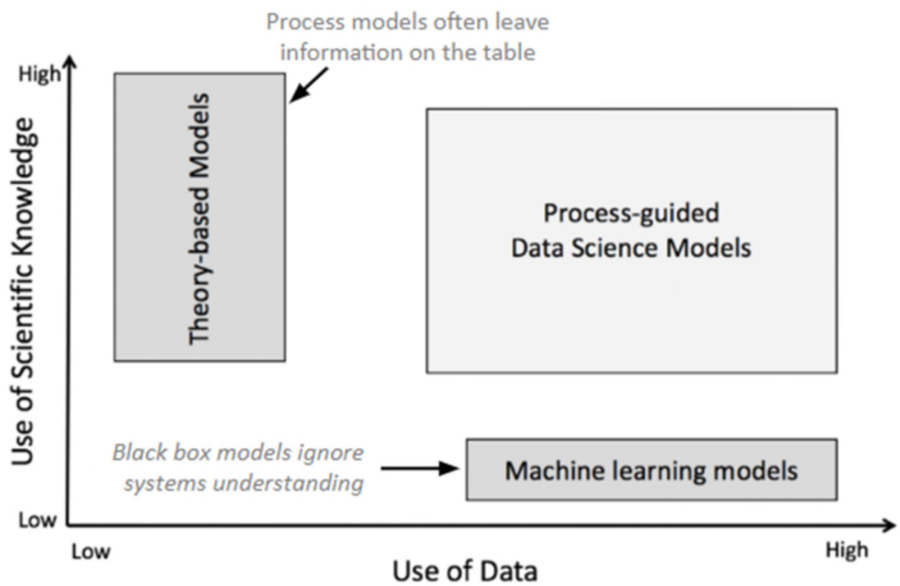

Figure 16. Big data analysis can help benefit both black box and process-based modeling approaches. Modified from Karpatne et al. 2017.

A common method of creating hybrid, process-guided data science models is to incorporate prior knowledge using a Bayesian framework. Here, the model inferences are predictive and conditioned on a prior distribution of beliefs before any data is observed. Another alternative is to use cost functions that heavily penalize an inference that violates certain rules or laws such as conservation of mass, energy, and momentum. Conversely, if the goal is to preserve the internal structure of the model, then the model can be hybridized by simply creating artificial training data that "learns" the theories that we wish to impose on an otherwise purely statistical model.

8.2.1 Model Emulators to Represent Complex Models

When computationally intensive models are used in analyses containing multiple models, the combined model run time can be time-prohibitive on desktop machines and alternative approaches such as cloud computing may need to be considered. Another alternative that has gained some currency in the literature is to use an emulator for one or more models within an integrated modeling framework. Emulators need some resources to develop, but once created, they may allow certain types of model integration that may not be possible with the original models. Emulation approaches, summarized in Table 4, range from simple linear regression to sophisticated deep learning artificial neural networks.

Emulators represent the input/output relationships in a model with a statistical surrogate to reduce the computational cost of model exploration. In this approach, the computer model is viewed as a “black box” and constructing the emulator can be thought of as a type of response-surface modeling exercise (Box and Draper, 2007). The approach establishes an approximation to the input-output map of the model using a limited number of complex model runs. Of course, as with any approximation, emulators produce less accurate estimates. Therefore, model developers must consider this trade-off between accuracy and computational cost.

Table 4. Model Emulation Approaches

Algorithm	Description
Linear Regression	Linear regression is a ubiquitous technique that estimates one numerical variable as a linear function of one or more other variables. It is conceptually simple and computationally efficient for datasets of almost any size. Assumptions on data structure are quite restrictive compared to some of the other more complicated algorithms listed below; thus, the ability to make full use of the theoretical results about a linear model is generally unlikely on real-world data. Nevertheless, linear regression models can serve as useful building blocks in more complex models. In principle, approaches such as regression should be limited to the range of data used to develop the regression, and not extrapolated beyond.
Logistic Regression	Logistic regression is a type of regression for binary (yes/no) variables. The estimated parameters of the model are still linear with the input variables, but a sigmoidal function maps the underlying linear predictor to fall within the range of 0–1. The value that a given combination of input variables outputs is the probability that the corresponding output variable has value 1 (e.g., yes/true). The use cases of logistic regression for binary variables are similar to those of linear regression for continuous variables: it is a conceptually simple and computationally efficient model that has restrictive assumptions compared to other more complex algorithms. Logistic regression is often a building block in artificial neural networks (ANNs) discussed below. Both linear and logistic regression fall in a family of techniques called “Generalized Linear Models,” but these two are the most common.
Bayesian Inference	Bayesian inference isn’t a specific model but rather a method for estimating model parameters that can be specified by probability distributions. In practice, many of the models that practitioners in water resources might be interested in using fall into this category, the main exceptions being “nonparametric” procedures like the Mann-Kendall rank-based trend tests. The main strengths of Bayesian inference are that uncertainties for the estimated parameters are automatically generated in a straightforward manner and that it is possible to incorporate prior information (e.g. expert knowledge, results of previous studies) as a regularizing effect to improve estimates on parameters in more complex models where the data alone might be insufficient. Bayesian inference is also one of the best ways to fit structured *multilevel* models, where the data is organized in a hierarchical fashion: e.g., a model of water samples from several lakes in a region might be organized so that the samples from the same lake are in the same group and share information with each other.

Markov Chain Monte Carlo (MCMC) Techniques	A simple algebraic expression for the properties of a probability distribution generally only exists for the simplest examples. In other cases, including many of the Bayesian models that one would like to use in practice, alternative methods must be used to estimate the necessary calculations. MCMC refers to a state-of-the-art family of methods that explore probability spaces with a sequential (Markov) chain. These methods are particularly good at evaluating high-dimensional spaces that come up in real-world multivariate problems. However, they tend to be computationally intensive and can require some fine-tuning on the part of the analyst to ensure that they have converged.
Spline Methods	There is often a need to estimate the relationship between variables with unknown but nonlinear functional form. Splines are one way to do this—they are unknown smooth functions evaluated at a limited number of points (knots) that have some constraints on their degree of smoothness, often expressed as a penalty on the second derivative of the function. Splines can be computationally less expensive than other techniques discussed below, but the determination of where to place the knots can be difficult or arbitrary. Generalized Additive Models (GAMs) often use spline functions as a basis for expressing unknown smooth functions.
Gaussian Processes	Gaussian Processes is another method to estimate smooth functions. In contrast to being evaluated at a discrete set of points like splines, Gaussian Processes are parameterized in terms of a known (or assumed) covariance function between pairs of observed data points. This is often conceptually more elegant and sidesteps that question of knot placement, but it is computationally expensive in the general case and approximations often must be made on all but the smallest of datasets. Kriging techniques, often used by GIS practitioners, are a type of Gaussian Process.

8.2.2 Big Data Analysis Technologies and Applications

There is a variety of available big data analysis tools and frameworks that can be used for integrated models. Considering the large data requirements and computational power demand of integrated models, application of big data analysis tools is expected to create new efficiencies and new opportunities, such as the hybrid modeling approach described above. This chapter provides a list of the most popular big data analysis frameworks in use that have potential applicability in the environmental domain. There are some published environmental applications of specific tools (as noted below), although for many of these tools, their use in environmental applications has not been documented in the scientific literature.

- **Apache Hadoop:** The Apache Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Hadoop is open source and many large organizations are already implementing its capabilities. Hu et al. (2015a) coupled a multi-agent system model with an environmental model for watershed modeling with Hadoop-based cloud computing. They reported an 80% reduction in runtime for the coupled model. The practice showed a good potential for scalable execution of the coupled model through application of Hadoop. Hu et al. (2015b) also used Hadoop-based cloud computing for global sensitivity analysis of a large-scale socio-hydrological model. They were able to reduce the computation time of 1000 simulations from 42 days to two hours.
- **Apache Spark:** Apache Spark is an open-source distributed general-purpose cluster-computing framework. Spark provides an interface for programming entire clusters with implicit data parallelism and fault tolerance. Spark facilitates the implementation of both iterative algorithms (which visit their data set multiple times in a loop) and interactive/exploratory data analysis, i.e.,

the repeated database-style querying of data. Omrani et al. (2019) implemented the Apache Spark framework to reduce the high computational burden of land change simulation model across a large region and span of time. Their results showed significant computational performance improvements compared to running the model out of the Spark framework.

- **Apache SAMOA:** Apache SAMOA (Scalable Advanced Massive Online Analysis) is an open-source platform for mining big data streams. SAMOA provides a collection of distributed streaming algorithms for the most common data mining and machine learning tasks such as classification, clustering, and regression, as well as programming abstractions to develop new algorithms.
- **Microsoft Azure HDInsight:** Azure HDInsight is a Spark and Hadoop service in the cloud. It provides an enterprise-scale cluster for the organization to run their big data workloads.
- **Teradata Database:** Teradata database allows analytic queries across multiple systems, including bi-direction data import and export from Hadoop. It also has three-dimensional representation and processing of geospatial data, along with enhanced workload management and system availability. A cloud-based version is called Teradata Everywhere, featuring massive parallel processing analytics between public cloud-based data and on-premises data.
- **IBM Watson:** Watson Analytics is IBM's cloud-based data analysis service. When data are uploaded to Watson, it asks questions it can help answer based on its analysis of the data and provide key data visualizations immediately. It also does simple analysis, predictive analytics, smart data discovery, and offers a variety of self-service dashboards. IBM has another analytics product, SPSS, which can be used to uncover patterns from data and find associations between data points.
- **Skytree:** Skytree is a big data analytics tool that allows the development of data-driven models using machine learning approaches. The tool provides capabilities for data scientists to visualize and understand the logic behind machine learning decisions. Skytree provides model interoperability capabilities and allows access through a GUI or programming in Java.
- **Talend:** Talend is a big data tool that simplifies and automates big data integration. It also allows big data integration, master data management and checks data quality. Talend is open source and provides various software and services for data integration, data management, enterprise application integration, data quality, cloud storage and Big Data.
- **R:** R is a language and environment for statistical computing and graphics. It is also used for big data analysis and provides a wide variety of statistical tests. R provides effective data handling and storage facility, a range of matrix operations, several big data tools, and great visualization capabilities. Many R packages for machine learning are also available off the shelf.
- **MATLAB:** Matlab is a multi-paradigm numerical computing environment and proprietary programming language developed by MathWorks. Matlab has numerous designated data analysis toolsets. Statistics and Machine Learning Toolbox provides functions and apps to describe, analyze, and model data. Regression and classification algorithms provide the capability to draw inferences from data and build predictive models. The toolbox provides supervised and unsupervised machine learning algorithms for big data, including support vector machines (SVMs), boosted and bagged decision trees, k-nearest neighbor, k-means, k-medoids, hierarchical clustering, Gaussian mixture models, and hidden Markov models. Matlab also has strong visualization capabilities which are essential for big data analysis.
- **Python:** Python is an interpreted, high-level, general-purpose programming language. Similar to R and Matlab, Python has numerous data analysis toolsets including NumPy, pandas, and Scikit-Learn. Scikit-Learn implements a wide-range of machine-learning algorithms and allows them to

be plugged into actual applications. A range of functions are available through Scikit-Learn such as regression, clustering, model selection, preprocessing, classification and more. Scikit-Learn in in widespread use today for big data analysis.

8.3 Machine Learning Methodologies and Frameworks

Machine learning is the process of studying data to detect patterns by applying known rules to categorize data, predict outcomes, and detect anomalies. Machine learning is driven by algorithms which perform more accurately as new data is provided. According to Nevala, 2007, these algorithms are suited to four main categories of problems:

- where associations between data are qualitative and/or intuitively understood but not easily described by programmable logic rules;
- where potential outputs are defined but non-uniquely depend on a diverse set of conditions;
- where accuracy is more important than interpretability; and,
- where a dataset is large and highly correlated such that traditional analytical techniques may develop biases based on how frequently correlated features appear in the input data, rather than their physical importance in the environmental system at hand.

Machine learning can be supervised, reinforced, or unsupervised. In supervised machine learning, the algorithm is provided with sample inputs and corresponding outputs and learns by example. In reinforced learning, the algorithm is provided with rules and potential outcomes which govern its decision-making. In unsupervised machine learning, the algorithm identifies patterns and draws its own inferences. In all three cases, the machine learning algorithm is trained on a set of data and then tested on another, unseen set to gauge its performance.

Oftentimes, machine learning is criticized for being a black box. This is explained in further detail in the subsections below. There are, however, techniques that can be used to assess the soundness of a model's machine learning to ensure that it is capturing a theory-guided phenomenon rather than employing unfounded correlation. Ablation studies involve running a series of iterations in which the model capabilities are reduced one at a time to experimentally determine which functionalities embedded within the model are of greatest importance in the model's performance. Created to unmask machine learning mechanisms, Layer-wise Relevance Propagation (LRP) is a technique that decomposes the output of a neural network into a relevance heat map, showing the importance of input variables and values in determining the output. This is an effective visualization of how the model learns and can be used to advise steps that can be taken to ensure that the learning process makes physical sense. Alternative methods that illuminate machine learning and its physical interpretations are explored in McGovern et al., 2019.

Machine Learning Methods in Environmental Science (Hsieh, 2009) is a useful textbook to further explore applications of machine learning in modeling environmental systems, including water. Alternatively, *Machine Learning for Spatial Environmental Data* (Kanevski et al., 2009) explores machine learning applications in problems that specifically use geospatial data. For further reading on the statistical theory and the conceptual backbone of machine learning, *Pattern Recognition and Machine Learning* (Bishop, 2006).

8.3.1 Neural Networks

Artificial Neural Networks (ANNs) encompass a broad class of models that represent relationships among data in a fashion that has some similarities to biological neurons: variables correspond to nodes and the parameters of the model correspond to connections between the different nodes, usually between intermediate nodes that represent internal model state. The relationships that ANNs can represent are very general—they are often described as "black box" models—and the complexity of those relationships is determined by the structure of the connections between the nodes in the network.

Recurrent Neural Networks (RNNs) are best suited for time series data. Here, the sequential order of input data points is captured so that time-dependency can be incorporated into the model's method of learning. The model parameters that are embedded in RNNs are shared across time steps. Once decided, these parameters remain unchanged and can significantly increase the speed at which neural network models are trained.

A Convolutional Neural Network (CNN) is another variation of neural network architecture and is commonly used in image recognition and classification. A CNN contains one or more layers of convolutional nodes which act as filters. Their goal is to pinpoint the main features of a dataset by considering smaller subsets of the input data and figuring out their relative importance. In environmental modeling, this is used to distinguish signals of interest from potentially confounding noise.

ANNs, CNNs, and RNNs are just some examples of different neural network model set-ups. These three are depicted in **Figure 17**. There are many more that have been developed in recent years, inspired by specific problems that data scientists have encountered. Neural Networks, in general, are very flexible models that can pick out unknown relationships among multiple variables, but they are computationally expensive to train. Non-deep networks (deep networks are described below) can require expert knowledge and pre-processing of data to get accurate, structurally valid, and generalizable models. As such, they provide the foundation for artificial intelligence and machine learning.

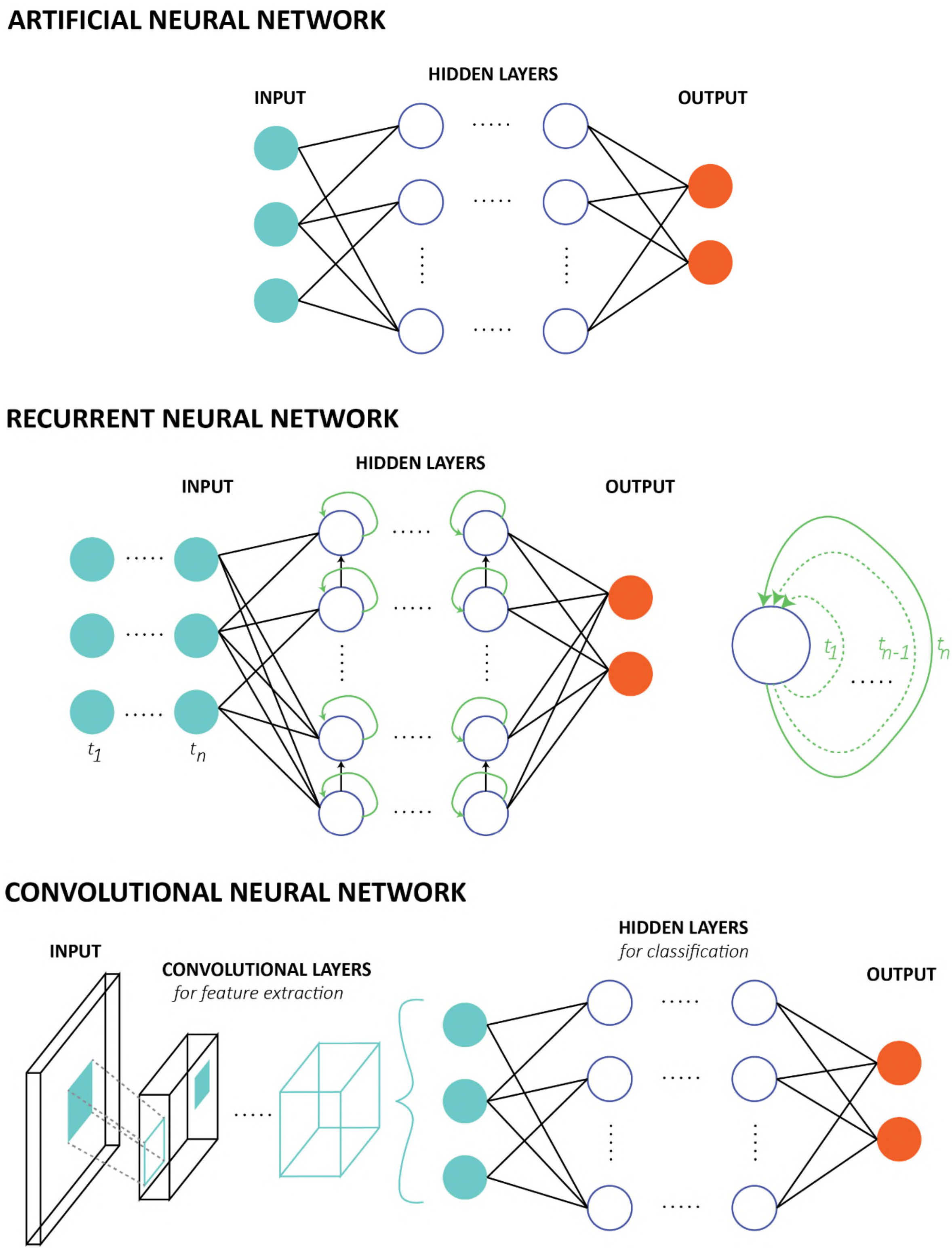

Figure 17. Graphical illustration of an artificial neural network, a recurrent neural network (RNN), and a convolutional neural network (CNN). The structure of a RNN and a CNN explain why they are better suited for time-series and classification problems respectively. However, the neural network structures above can be applied to a variety of machine learning problems.

8.3.2 Deep Learning

Deep learning is a term that refers to the direction taken by a large part of the research effort in the field of machine learning over the past decade or so. It refers to specialized data analysis algorithms making use of a particular type of neural network with multiple layers (hence the "deep" terminology). Deep learning approaches have been able to handle machine learning challenges that were previously intractable.

Among the best-known successes of deep learning algorithms are examples such as computer vision, natural language processing, or strategy games. However, these approaches have the potential to be uniquely applicable to any domain with a rich, complicated dataset where representation learning could be useful.

The algorithms and software behind training and implementing a deep learning model can be complex, but there has been a proliferation of software environments, both open-source and enterprise-minded, for helping developers to create and successfully apply deep learning models. Many of these environments form an ecosystem centered around the Python programming language, but interfaces to other languages such as R or Julia exist as well.

- TensorFlow is one of the best-known general-purpose deep learning frameworks. It originated at Google and has interfaces to various programming languages, including Python and C++.
- Keras is another popular deep learning framework that is built atop TensorFlow 2.0. It is scalable and so can run on large clusters of GPUs for distributed model training and has the full deployment capabilities of TensorFlow.
- Sagemaker is Amazon's all-in-one machine learning platform, designed for integration with their cloud computing resources to train and deploy models that require that type of high-performance computing.
- Azure Machine Learning is a similar competitor from Microsoft that runs on their Azure cloud computing platform.
- Intel oneAPI Deep Learning Framework Developer Toolkit is a software development kit that is optimized for large datasets and offers high performance using Intel's CPUs and GPUs. It supports programming in C and C++.
- NVIDIA CUDA-X AI is a unified deep learning software stack with libraries, toolkits, and pretrained models that GPU-accelerate deep learning in every framework and across applications including conversational AI, computer vision, and natural language processing.

8.3.3 Ensemble Learning

Ensemble modeling is widely used in the climate realm to obtain better predictive performance than can be achieved using any one model alone. The theory here is that a single model represents a single hypothesis or interpretation of the input data on which it was trained. Other models with their own parameters and structures represent alternative hypotheses that can be drawn from the same input data. The amalgamation of these models creates an ensemble which can counter the characteristic variance and biases that are introduced by each of the constituent models. In this way, older, off-the-shelf models that were deemed too narrow in scope can be revisited and improved upon in an ensemble set-up where their limitations are addressed by newer methods.

Ensemble learning allows modelers to take advantage of various machine learning algorithms that have been developed to tackle water modeling problems that are specific to sites or environmental characteristics. In theory, an ensemble model can result in overfitting of the input data; however, techniques such as bagging, boosting, and stacking address such issues.

In hydrological models, a single model may perform well under certain conditions but may produce drastically inaccurate results in other climatic regions or even seasons. Thus, the concept of multi-model combinations has gained popularity. Xu et al., 2020 tests the performance of an ensemble of watershed models and that of a single watershed hydrological model using a number of evaluation metrics such as root mean squared error (RMSE), and the coefficient of determination (R^2). The results show that regardless of the climatic conditions of the study area and the metric used for evaluation, the ensemble model consistently outperformed individual models. This observation is consistent across the literature in this field of research.

A variety of techniques can be employed to create a suitable ensemble, ranging from taking a simple ensemble mean to determining non-linear weighting schemes which prioritize performance at discrete intervals. Sometimes, the latter example can be accompanied by yet another machine learning algorithm that derives a new optimal weighting scheme each time the ensemble is trained. Alternatively, ensemble methods can be used to fill missing historical data that might be useful for training models in the future.

In all these instances, it is important to note that while in theory, an infinite ensemble size will produce the lowest error, in practice, ensemble construction follows a law of diminishing marginal returns: there is a great improvement in ensemble performance when a few, well-chosen models are used. As illustrated in **Figure 18**, following a certain point, the marginal improvement in ensemble performance associated with adding a new model to the ensemble may not be worth the additional computational cost and complexity.

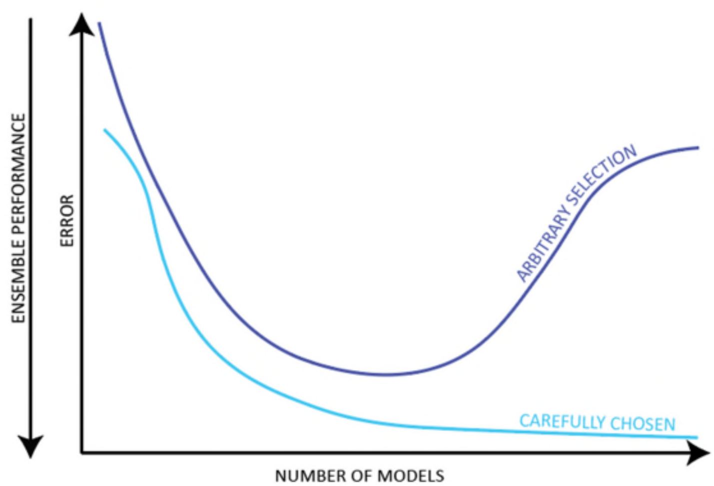

Figure 18. Diminishing returns in ensemble learning.

8.3.4 Evolutionary Algorithms

Water and environmental modeling are used to support decision-making processes. Oftentimes, the goal is to aid in environmental management and to evaluate the effects of certain activities. As such, models need to be designed to simulate real world dynamics as well as alternative realities, wherein other environmental conditions, opportunity cost scenarios, and strategic choices can be explored and evaluated. The process of formulating an environmental problem in terms of objectives, decision variables, and constraints allows the model to evaluate the best solutions through some form of quantitative or qualitative optimization process. In practice, this would involve an evaluation of many different scenarios in search of the best alternative. However, this may not always be possible as an algorithm could converge at a local optimal rather than a global one (previously discussed in **Figure 10**).

Evolutionary algorithms are a class of modern optimization techniques known as "metaheuristics". They are best suited to find globally optimal solutions. This is because an evolutionary algorithm first generates a population of solutions and evaluates each solution's fitness. This increases the probability of finding the global minima as the algorithm first explores a diverse range of points before choosing one to converge to. In addition, each new population of solutions is generated using a variety of heuristic operators such as crossover, recombination, and mutation, with an emphasis on the subset of solutions that have performed the best thus far. This allows the search to be governed by the properties that made past solutions most successful. This kind of search is well-suited for complex, non-linear problems with large solution spaces, multiple constraints, and multiple or competing definitions of cost functions. These issues regularly arise in environmental and water modeling.

The searching behavior of evolutionary algorithms can be customized based on the optimization problem at hand. For instance, gradient methods tend to converge immediately so the direction of convergence and subsequent optimal solution is heavily dependent on the first iteration. Random searches, on the other hand, may take longer to converge. An evolutionary search algorithm can take advantage of both methods and apply them appropriately; however, this can be computationally expensive. In addition, evolutionary algorithms utilize past computations and solutions and can even rank the fitness of solutions that have been searched. This is a good bookkeeping strategy which also communicates how the algorithm is assessing solutions in search for a global optimum and which solutions are similar in cost. This can often be very insightful as several suboptimal but appropriate solutions can also be found and evaluated. Constraints, which can be embedded in the form of cost functions, are easily handled by evolutionary algorithms and can provide physical relevance to the model and search behavior. Evolutionary algorithms are also very efficient at dealing with multiple cost functions as they can search for an optimum across a multi-dimensional fitness landscape.

In the realm of water resources, evolutionary algorithms have been employed in groundwater monitoring design, water resources planning and management, water distribution network design and optimization, transport modeling, water quality sampling optimization, lake and reservoir simulations, salinity estimation, and hydrological model calibration. Maier et al., 2014 provides a thorough review of evolutionary algorithms in water resource modeling and points to many leading researchers and their studies.

8.4 Data Visualization and Communication Techniques

An important facet of environmental and water modeling is the communication of model structure and results to diverse audiences that comprise both seasoned modelers and inexperienced but invested model stakeholders and sponsors. As the datasets that are used by environmental modelers are multidimensional and complex, a number of 2D and 3D experiential tools can be employed to both engage and illuminate the inner workings of water and environmental modeling.

8.4.1 Data-driven Storytelling

Developments in data-driven storytelling are largely propelled by the needs of business intelligence analysts which, in many ways, are similar to the needs of modelers: both need to make models and data appeal to those who may not be directly involved in the modeling process. Some examples of data visualization tools include:

- **Tableau:** Tableau is a widely-used data analysis and visualization tool. Tableau queries relational databases, online analytical processing cubes, cloud databases, and spreadsheets to generate graph-type data visualizations. The tool can also extract, store, and retrieve data from an in-memory data engine. Tableau also has a mapping functionality with the ability to plot latitude and longitude coordinates and connect to geospatial information such as ESRI Shapefiles, Google Earth KML files, and GeoJSON.

- **Power BI:** Power BI is a similar platform, created by Microsoft. Power BI provides the tools needed for users to create data dashboards and visualizations which integrate with Excel queries and data models. Supported by Microsoft's robust repertoire of big data resources, Power BI enables non-data enthusiasts to build machine learning models which can deal with non-numerical data including images and texts. Power BI's platform can tap into sensors and provide real-time analytics as well while Power BI Pro provides cloud service for augmented analytics and automated machine learning.

- **Plotly:** Plotly, or Plot.ly, is focused on data visualization without requiring programming or data science skills. Its GUI is designed for importing and analyzing data and uses the D3.js JavaScript library for all its graphics. Its dashboards can be generated in real-time as well as from existing data pools, and it supports exporting to a variety of visualization tools as well, including Excel, SQL databases, Python, R, and MATLAB.

- Domo: Domo is a big data analysis and visualization tool that automatically pulls in data from spreadsheets, on-premise storage, databases, cloud-based storage, and data warehouses and presents information on a customizable dashboard. It has been lauded for its ease of use and how it can be set up and used by a wide range of users, not just a data scientist. It comes with a number of preloaded designs for charts and data sources to get moving quickly.

A more exhaustive list of analytics platforms and their unique differentiators is compiled in Richardson et al., 2020, a report that was released by Gartner. While the above platforms were developed for business analytics, they are adequately furnished with tools to provide insights into many conventional environmental data sets as well.

8.4.2 Augmented Reality and Virtual Reality

Augmented reality (AR) and virtual reality (VR) have introduced new and powerful methods of data visualization. Such opportunities are only beginning to be exploited in the domain of water and environmental modeling. Mobile-based AR applications have been created to illustrate the flooding in immersive 3D experience as opposed to static top-down visualizations that can be made using traditional mapping software. For modelers, mobile AR is particularly compelling and easy way to communicate results with physical importance. All that is needed is a modern mobile device with 3D graphics capabilities and a stakeholder can now be shown field conditions first-hand. VR can take this a step further by creating an interactive experience.

Both desktop and web-based AR and VR tools can be employed in water and environmental modeling to aid many aspects of the modeling project from data collection to visualization. Google SketchUp's geo-location tool can simulate exact terrain imagery. Oxagile supports a 360º video viewing experience with advanced viewing analytics. Virtualitics offers a platform that harnesses research conducted by the California Institute of Technology and NASA Jet Propulsion Laboratory. It integrates machine learning to not only show how the effect of different variables based on the input data but create an immersive, multidimensional experience to demonstrate such findings in smarter ways. This is particularly useful in model calibration and parameter estimation.

Augmented Environments (ANTS) is a project which specifically uses AR technology to explore physical and natural structures for the purpose of environmental management. Using a combination of tracking devices, a video camera, and human interfaces such as headsets, the ANTS system has been applied in monitoring water quality using pollutant transport models, visualizing temporal changes in water bodies, and superimposing synthetic images on the ground to reveal the location of underground water supply networks and subsoil structure (Romao et al., 2004). Haynes et al., 2018 demonstrates the use of mobile AR in flood visualization which is linked to a network of sensors that provide live measurements from the field. A preliminary evaluation of their proposed mobile AR platform shows positive stakeholder feedback.

8.5 Workflow Organization Tools

Code sharing, data storage, and workflow organization tools are of importance to modelers who are dealing with large data sets, complex models, multiple iterations of code, and multiple contributors to the project. Some examples of workflow organization tools include:

- GitHub is a platform for organizing this team set-up. Code can be pulled, new features can be proposed, versions can be branched and then merged upon thorough review by teammates and upon clearing status checks. This creates and stores a repository where the source code is protected while new edits are evaluated. Github also integrates with other project management applications that specialize in debugging, quality control assessments, running test cases, and integrating annotations or notes.
- BitBucket is another such repository hosting platform owned by Atlassian. BitBucket uses Sourcetree, a Git graphical user interface that helps coders visualize and manage their repository and traverse through versions of code with ease.
- Google Cloud Source is a popular collaborative and scalable Git repository which offers the opportunity to connect the modeling workflow with other Google Cloud tools. This can help in efficient deployment of code when the goal is to create apps to enhance fieldwork or can allow coders to seamlessly build unique cloud functions to manage large databases on the cloud.

8.6 New Methods in Software Engineering and Architecture

As models grow in complexity and development schedules are compressed, management tools from the field of software engineering can identify practices that may be suitable for adaptation for the problems typically addressed in the water resources domain. Here we highlight methods in cloud and cognitive computing which have evolved to automate, scale up, and manage modeling tasks where real-time data updates and new ways of thinking need to be handled efficiently and incorporated effectively.

8.6.1 Cloud Computing

Innovations in data capture (Section 8.1) indicate that there are several large and diverse datasets that can be applied to water and environmental modeling. The rapid creation and deployment of Internet of Things-based environmental sensors requires appropriate application programming interfaces to communicate data capturing protocols, host the data streams, and parallelly process large amounts of input data. This is both a technical and a logistical challenge.

Cloud infrastructure is increasingly being deployed to support and manage these large data streams that are captured by smart objects and sensors. Cloud computing is a ready-to-use, scalable, on-demand computing environment that is hosted on the internet and uses a network of remote servers to provide computing power and storage beyond that which is provided by a local server. A modeler can create, launch, and terminate servers as needed to run a variety of simulations with varying computational requirements. Aside from the speed, flexibility, and cost effectiveness, cloud computing can be used to run routine model calibration and uncertainty analyses in more enlightening ways than traditional methods offer.

Cloud platforms such as Google Cloud and Amazon Web Services offer relational databases which are structured to detect relations among the stored information. Examples include databases Amazon Aurora, Google Cloud Databases, and Azure Databases. Column-store databases such as Amazon Redshift and Google Cloud BigTable have the advantage of compressing and partitioning data, aggregating information, improving functionality using composite columns, scaling up for parallel processing, and efficiently loading and querying the data. Alternatively, document store databases such as Google Cloud Firestore and Azure Cosmos DB are best suited for data of a variety of structures, from spreadsheets to imagery. Finally, graph databases store data in the form of nodes that are connected by edges, allowing for fast querying and traversal with explicit linkages between data based on relationships that are prioritized. This is supported by Amazon Neptune, for instance. Thus, regardless of the future data storage, traversal, or querying challenges, there are numerous cloud tools available to create optimized databases.

Many cloud providers and their computing services are available for modelers to use. According to Granell et al., 2016, these services can be broadly categorized as software, platforms, and infrastructures. Software such as Google Maps is already well-suited for collecting and organizing geospatial field data. With the help of Google App Engine as a platform, applications that use this data in modeling activities can be developed and run on the cloud. Hardware and system resources such as data centers comprise the infrastructural needs to support platforms and software that are traversing through large data sets. Amazon Web Services (AWS) is a cloud computing platform that is supported by Amazon EC2 and Amazon S3. GoGrid and The Rackspace Cloud are other examples of cloud infrastructure.

Fustes et al., 2014 delineates the utility of cloud resources in marine data processing applications such as detection and localization of marine spills using remote sensing methods and advanced algorithms deployed within a cloud platform. Wan et al., 2014 shows how cloud infrastructure can improve computing performance and can manage, query, and analyze a global flood database in real-time while simultaneously providing effective location-based visualizations for the public. CyberFlood, the cyber-infrastructure in this study, is built on Google Fusion Table. Microsoft AI for Earth, a bundle of cloud platforms and infrastructures, is an initiative to make cloud computing more accessible. AI for Earth hosts a variety of geospatial datasets on which Azure cloud can operate and accelerate the modeling workflow. Various applications can also be created and deployed to provide further insights in the field while the cloud platform eliminates limitations pertaining to data structure and storage capacity.

Cloud computing opens many opportunities in groundwater modeling. The increasing complexity and volume of environmental datasets make distributed parallel processing an asset in modeling. According to Hunt et al., 2010, parallel computing, enabled via the cloud, can “automate calibration and uncertainty analysis using parameter estimation techniques”, which is widely used for groundwater modeling. Beginning with user-specified model parameters, the parameter estimation process can be used to perturb parameters, observe the resultant effect on model performance, and use this knowledge to adjust the parameters optimally. In effect, parallel computing on the cloud performs a sensitivity analysis on the model with respect to changes in each parameter; however, the efficient delegation of adjusting parameters and running iterations gives parallel computing an edge over traditional sensitivity analyses. This method has been used to calibrate SWAT watershed models (Ercan et al., 2014). In the field of groundwater modeling, where the large number of parameters can introduce extremely high computational costs, cloud computing is a cost-effective and flexible approach for optimizing model calibration and running uncertainty analyses concurrently (Hunt et al., 2010).

8.6.2 DevOps

In the field of computer science there has been increasing focus on streamlining the process of development and delivery of increasing complex software systems. DevOps includes processes for version control, continuous integration, artifact management, automated testing, continuous delivery, and system monitoring that work together to both reduce the time to develop and deploy software and to improve the reliability of the software (Kim et al., 2016; Pipinellis, 2015). Version control has been in common use for decades to track and manage changes in sources code. Other components of DevOps are much newer and only gained wide use in the last ten to fifteen years. Continuous integration is a technique to automatically trigger new builds of a software package when new code is posted to the version control system. Artifact management is a technique for managing compiled software versions and resources (data files etc.) usually accessible through a web data service. Automated testing can be triggered with each new build or changes in dependent artifacts performing fast, low level unit tests or full system regression. With sufficient automated testing, it is possible to enforce quality assurance of the software and to automatically deploy software into a production setting. System monitoring and reporting is needed to administer the DevOps processes and head off any errors as rapidly as possible. These tools can be utilized with many different strategies for software development. One common strategy is where developers work in a copy of the code referred to as a feature branch. The developer performs both manual and automated testing on their local copy, and when the code is ready it is submitted to a staging branch. Further acceptance testing including deeper system regression is done on the staging branch, and only when the code is verified is it submitted to the mainline (production) branch.

In many ways, the development and application of numerical models is similar to the process of creating and deploying complex software. Certainly, the development of model executables is in fact a software development process. But production application of numerical models can also benefit from DevOps tools and concepts. Pre- and post-processing scripts can easily be treated as source code and managed through a version control system. Model software versions and prepared model data sets (e.g., alternative scenarios) can be managed and distributed through artifact management systems. Continuous integration tools together with automated testing can be used to verify new model versions and supporting data sets meet appropriate requirements, including automated validation by recomputing historical periods and comparing against previous results or historical data. Automated reporting, particularly of complex calibration/validation reports or QA/QC reports, can greatly reduce the time required to accept new versions of models and supporting data into the production modeling environment. The benefits to these ideas seem clear for real-time applications of models. But even for planning studies, streamlining the flow of the modeling process can be extremely useful. In particular, post processing of raw model output is almost always required to produce derived metrics which are the primary outcome of the modeling study, and complexity of the post processing may be similar to the complexity of the model itself. Treating the post processing steps as stream of operations through continuous integration/continuous delivery, any errors discovered can be corrected and results efficiently recreated, and, more importantly, the modeling team can be much more responsive to requests for revised outputs.

Establishing the computer infrastructure and software tools to support DevOps within an organization can require significant investment and may also require some cultural changes within the modeling team. However, the long-term benefits that can be gained may outweigh the upfront costs particularly for groups with sustained commitment to carrying out modeling studies.

8.6.3 Cognitive Computing

Cognitive computing describes a system that strives to bridge the interface between people and software. Commonly associated with artificial intelligence, cognitive computing often involves the deployment of platforms which allow machines to interpret a user's "native tongue" without requiring them to write or understand code or algorithms. Cognitive computing can be harnessed in environmental modeling where there is a need for context-based understanding of the underlying problem at hand and a method by which evidence can be gathered and qualitative reasoning can be organized.

Cognitive mapping can be used to organize theory-based knowledge and statistical relationships based on causal relationships. This is important in a Bayesian sense where there is need to create hybrid, process-guided data science models to take advantage of both theory and statistics while also being able to account for uncertainties. An example of cognitive mapping in lake eutrophication modeling and resource management is explored in Kouwen et al. (2008). A simple causal loop diagram can be used to show that nutrients promote the growth of algae but that increasing the biomass of aquatic vegetation can suppress algal growth by reducing the availability of nutrients. Integrating probabilistic models with these qualitative relationships can reveal that there are two stable equilibria for the lake system: a clear state with dense aquatic vegetation and a turbid state with widespread algal blooms. In a managerial sense, it may be intuitive to simply reduce nutrient concentrations to suppress algal proliferation and to steer towards the clear state equilibrium. However, a hybrid model developed using cognitive computing reveals that it is necessary to reduce water depth as well so that turbidity decreases, and vegetation growth is adequately promoted. In this instance, cognitive mapping merges qualitative causal relationships with data science models to quantitatively guide water management, indicating the amount by which water depth and nutrient concentration should be decreased in order to achieve the favored equilibrium state.

Microsoft's Azure Cognitive Services allow users to infuse artificial intelligence, in the form of machine vision and decision making, into their applications without the need for machine learning expertise. In environmental modeling, which begins from sensor measurements and field observations, cognitive computing can be used to provide a machine with the sense perception needed to wrangle the input data so that the data analysis and interpretation steps become less intimidating.

9 Model Life Cycle Management

Successful models tend to be applied to multiple studies over an extended life (sometimes over decades), either as-is or with modifications and updates. Codes and formulations from a successful model are often re-purposed and used in a new generation of models. Consequently, it is important to think about modeling protocols over the life of the model or the long-term life cycle. This is shown schematically in **Figure 19**. As shown in the figure, the computer implementation of each model is initially based on a specified conceptual framework and model structure. This leads to a computer implementation which may be used for one or more model studies. As information from multiple model studies is accumulated, a more nuanced understanding of the strengths and weaknesses of the underlying model structure will develop. This may help to inform and improve the underlying conceptual framework (shown as a feedback loop below), and ideally, result in updates and revisions to the model for future applications. Over the long-term, individuals responsible for model development will likely change and, therefore, there is a need to adequately document the existing model and to develop an effective long-term plan for code maintenance and migration to newer software platforms. In the context of modeling, a long-term life cycle refers to model management and maintenance activities that enable its continued improvement and evolution over time.

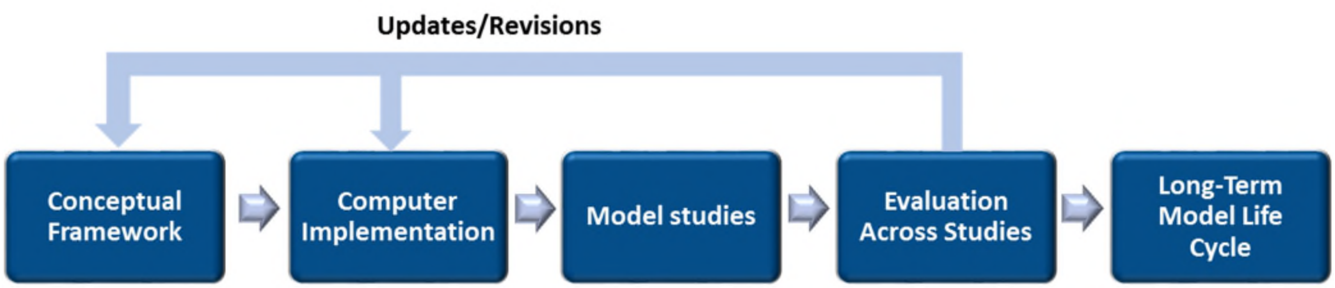

Figure 19. The life-cycle of a typical model.

Models represent large intellectual and financial investments, but in most instances, their long-term viability is unknown. Models are often developed to serve a specific need, and access to these models and related analyses rapidly diminishes over time. For key foundational models and related efforts, long term sustainability should be addressed early in its life cycle to make best use of the investments being made. This life cycle planning should identify the responsibilities, accountabilities, and resources needed to support a model over the long term, potentially over decades. This life cycle planning should also contemplate development of new versions and ongoing model evaluation. Because of the resources and long-term commitments required to sustain a model over time, this is an issue that should be brought to the attention of decision and policy makers early in the modeling process.

For any modeling effort to be successful, leadership is needed to provide motivation to participants and sustained funding support is needed to allow models to develop over its life cycle. Such efforts involve some risk in that the resulting tools may not work as intended, may take too long to develop, or may be too computationally complex to be of practical use. Sustained funding recognizes that most modeling efforts will take additional time and resources to be fully evaluated for real-world application. In most cases, leadership and commitment are likely to be present when the institutions' missions and the goal of the specific modeling exercise are well-aligned.

10 Next Steps in the Implementation of Modeling Protocols

Most modelers are aware of the concepts identified in the preceding chapters. However, many do not implement them for various reasons. This may be due to time and resource limitations associated with virtually all modeling studies; this may also be due to the lack of specific expectations in the broader community of modeling study participants. Thus, model sponsors may not know what specific and reasonable requests to make of modelers to guide a model study toward greater credibility and usefulness. This chapter outlines some activities to encourage broader adoption of these protocols. These protocols are not intended to be specific or prescriptive requirements, but to describe best practices that are expected to benefit the broader community of users

10.1 Using the Modeling Protocols

To encourage adoption of the best practices identified in this work, we provide three relatively compact summary sheets in the Executive Summary. The purpose of the first sheet (Checklist 1), designed as a checklist to be employed at inception of a modeling effort, is to enable various participants to agree on the basic features of the work to be done. The purpose of the second sheet (Checklist 2) is to evaluate and score a modeling exercise upon completion. The purpose of the final sheet (Checklist 3) is to evaluate the long-term sustainability of a modeling framework.

The first sheet is designed with *Yes/No* responses, although additional narrative information can be provided. While there are no correct answers associated with the model study pre-audit, the questions are designed to flag issues that may need to be resolved before significant modeling study resources have been expended.

The second sheet contains a list of questions that may be answered with narrative responses or with numerical scores. If the numerical scoring approach is used, a model study with a higher score is generally more desirable. A numerical scoring approach may be useful for comparing multiple model studies that employ the same type of domain modeling. However, this approach is of limited value when a unique or one-of-a-kind model study is to be evaluated. The questions provided in these sheets are offered as starting points to be modified as needed for specific agencies or applications. However, we expect many of the essential items will apply to most modeling studies.

The third sheet is focused not on modeling *per se*, but on questions that help evaluate the long-term sustainability of a model framework. It is not intended to evaluate a single study, but to assess whether the framework used in one or more studies is well supported into the future.

10.2 Targeted Outreach

Additional targeted outreach to different groups of users may be needed to enhance the utility of these protocols, as proposed below.

10.2.1 Modelers

By modelers we refer to model specialists who have the expertise to run, modify and maintain model frameworks and applications. It is important to get the buy-in of this group so that they can provide feedback on whether these protocols are practical, usable, and provide meaningful support toward the creation of high-quality model studies as outlined in CWEMF's goals for this work.

10.2.2 Model Sponsors

Model sponsors typically fund model studies, and have an important role in directing them, even though they will not have a hands-on role in implementing these protocols. Outreach to this group should highlight the benefits of these protocols in creating high-quality studies, and raise their expectations when studies are planned. The need for additional resources to adequately apply these protocols should be described.

10.2.3 Non-modelers

These protocols can be used to communicate the practice of modeling to non-modelers, and to help identify the ways in which they may be able to contribute to the development of better models, even without specialist knowledge. The executive summary can be used as a stand-alone document to highlight key concepts presented herein.

10.2.4 Students and New Modelers

Much of modeling is a craft, with key steps not adequately described in textbooks or other documents. This protocols document is an effort to fill this gap and to introduce new generations of modelers to the processes and procedures of modeling that have evolved over several decades of practice. Toward this end, focused outreach through CWEMF meetings and alternative presentation formats such as short videos should be considered.

10.3 Future Updates of the Modeling Protocols

Since this report is a "living document," it will need to be updated periodically, as the need arises. The authors recommend that CWEMF reconvene its Ad Hoc Modeling Protocols Committee at least once every three years to ascertain whether a partial or full update is needed.

11 References

Ajami, N.K., Duan, Q., Sorooshian, S., 2007. An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction. Water Resour. Res. 43 (1), W01403.

Baroni, G., Tarantola, S., 2014. A general probabilistic framework for uncertainty and global sensitivity analysis of deterministic models: a hydrological case study. Environmental Modelling and Software, 51, 26-34.

Bastin, L., Cornford, D., Jones, R., Heuvelink, G.B.M., Stasch, C., Nativi, S., Mazzetti, P., and Williams, M., 2013. Managing uncertainty in integrated environmental modeling: The UncetWeb framework. Environmental Modelling and Software, 39, 116-134.

Beven, K.J., Binley, A.M., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrol. Process. 6, 279–298.

Beven, K.J., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J. Hydrol., 249(1-4), 11-29.

Bishop, C., 2006. Pattern Recognition and Machine Learning.

Blair, G. S., Henrys, P., Leeson, A., Watkins, J., Eastoe, E., Jarvis, S., and Young, P. J., 2019. Data Science of the Natural Environment: A Research Roadmap. Front Environ Sci 7 (121).

Box, G. E. P., Draper, N. R., 2007. Response Surfaces, Mixtures, and Ridge Analyses.

California Water & Environmental Modeling Forum (formerly Bay-Delta Modeling Forum), 2000. Protocols for Water and Environmental Modeling.

California Department of Water Resources, 2016. Best Management Practices for the Sustainable Management of Groundwater: Modeling. Available online at: https://cawaterlibrary.net/document/sgma-best-management-practice-modeling/

Crout, N., Kokkonen, T., Jakeman, A.J., Norton, J.P., Newham, L.T.H., Anderson, R., Assaf, H., Croke, B.F.W., Gaber, N., Gibbons, J. and Holzworth, D., 2008. Chapter 2: Good modelling practice. Developments in Integrated Environmental Assessment, 3, pp.15-31.)

Daniel, E.B., Camp, J.V., LeBoeuf, E.J., Penrod, J.R., Dobbins, J.P. and Abkowitz, M.D., 2011. Watershed modeling and its applications: A state-of-the-art review. The Open Hydrology Journal, 5(1).

Dawson, C., Fringer, O., He, R., Ralston, D., Zhang, J., 2019. Final Report on NSF Workshop on the future of coastal and estuarine modeling.

Delta Stewardship Council, 2020. Integrated Modeling in the Delta: Status, Challenges and a View to the Future. January.

De Vos, M.G., Janssen, S.J.C., Van Bussel, L.G.J., Kromdijk, J., Van Vliet, J. and Top, J.L., 2011. Are environmental models transparent and reproducible enough? In Proceedings of the MODSIM 2011-19th International Congress on Modelling and Simulation on Sustaining our Future, Perth, Australia, 12-16 December 3011 (pp. 2954-2961).

Doherty, J. and Johnston, J.M., 2003. Methodologies for calibration and predictive analysis of a watershed model 1. JAWRA Journal of the American Water Resources Association, 39(2), pp.251-265.

Doherty, J., 2015. Calibration and Uncertainty Analysis for Complex Environmental Models. Watermark Numerical Computing, Brisbane, Australia. ISBN: 978-0-9943786-0-6.

Doherty, J.E. and Hunt, R.J., 2010. Approaches to highly parameterized inversion: a guide to using PEST for groundwater-model calibration (p. 2010). US Department of the Interior, US Geological Survey.

Duan, Q., Sorooshian, S., and Gupta, V.K.. 1992, Effective and efficient global optimisation for conceptual rainfall-runoff models. Water Resources Research, 28 (4) (1992) 1015-1031.

Ercan, M. B., Goodall, J. L., Castronova, A. M., Humphrey, M., and Beekwilder, N., 2014. Calibration of SWAT models using the cloud. Environ Modelling & Software. 62: 188-196.

Fustes, D., Cantorna, D., Dafonte, C., Arcay, B., Iglesias, A., and Manteiga, M., 2014. A cloud-integrated web platform for marine monitoring using GIS and remote sensing. Application to oil spill detection through SAR images. Future Generation Computer Systems. 34: 155-160.

Gaber, N., Foley, G., Pascual, P., Stiber, N., Sunderland, E., Cope, B. and Saleem, Z., 2009. Guidance on the development, evaluation, and application of environmental models. Report, Council for Regulatory Environmental Modeling.

Ganju, N.K., Brush, M.J., Rashleigh, B., Aretxabaleta, A.L., Del Barrio, P., Grear, J.S., Harris, L.A., Lake, S.J., McCardell, G., O'Donnell, J. and Ralston, D.K., 2016. Progress and challenges in coupled hydrodynamic-ecological estuarine modeling. Estuaries and coasts, 39(2), pp.311-332.

Granell, C., Havlik, D., Schade, S., Sabeur, Z., Delaney, C., Pielorz, J., Uslander, T., Mazzetti, P., Schleidt, K., Kobernus, M., Havlik, F., Bodsberg, N. R., Berre, A., and Mon, J. L., 2016. Future Internet technologies for environmental applications. Environ Modelling & Software. 78: 1-15.

Grayson, R. Blöschl, G., 2001. Spatial Patterns in Catchment Hydrology: Observations and Modelling, pp 404. Cambridge University Press, United Kingdom.

Gupta, H. V., Sorooshian, S. and Yapo, P.O., 1999. Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. J. Hydrologic Eng. 4(2): 135-143.

Harmel, R.D., Smith, P.K., 2007. Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. J. Hydrol. 337 (3–4), 326–336.

Hawking, S.W., 1988, A brief history of time: From the big bang to black holes. Bantam Books, New York, 1988.

Haynes, P., Hehl-Lange, S., and Lange, E., 2018. Mobile Augmented Reality for Flood Visualization. Environ Modelling & Software. 109: 380-389.

Hsieh, W. W., 2009. Machine Learning Methods in the Environmental Sciences.

Hu, Y., Cai, X., DuPont, B., 2015a. Design of a web-based application of the coupled multi-agent system model and environmental model for watershed management analysis using Hadoop. Environmental Modelling & Software 70, 149-162.

Hu, Y., Garcia-Cabrejo, O., Cai, X., Valocchi, A. J., DuPont, B., 2015b. Global sensitivity analysis for large-scale socio-hydrological models using Hadoop. Environmental Modelling & Software 73, 231-243.

Huberts, W., Heinen, S.G., Zonnebeld, N., van den Heuvel, D.A., de Vries, J.P.P., Tordoir, J.H., Hose, D.R., Delhaas, T. and van de Vosse, F.N., 2018. What is needed to make cardiovascular models suitable for clinical decision support? A viewpoint paper. Journal of computational science, 24, pp.68-84.

Hunt, R. J., Luchette, J., Schreuder, W. A., Rumbaugh, J. O., Doherty, J., Tonkin, M. J., and Rumbaugh, D. B., 2010. Using a Cloud to Replenish Parched Groundwater Modeling Efforts. Groundwater. 48(3): 360-365.

Hutton, P.H., Rath, J., Chen, L., Ungs, M.L., and Roy, S.B., 2015. Nine Decades of Salinity Observations in the San Francisco Bay and Delta: Modeling and Trend Evaluation, Journal of Water Resources Planning and Management, DOI: 10.1061 / (ASCE) WR.1943-5452.0000617.

Interagency Ecological Program (IEP), 2015. An updated conceptual model of Delta Smelt biology: Our evolving understanding of an estuarine fish. Available at: https://water.ca.gov/LegacyFiles/iep/docs/Delta_Smelt_MAST_Synthesis_Report_January%202015.pdf.

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

Jia, H., Xu, T., Liang, S., Zhao, P., Xu, C., 2018, Bayesian framework of parameter sensitivity, uncertainty, and identifiability analysis in complex water quality models. Environmental Modelling and Software, 104(2018) 13-26.

Kanevski, M., Timonin, V., Pozdnukhov, A., 2009. Machine Learning for Spatial Environmental Data: Theory, Applications, and Software.

Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., Kumar, V., 2017. Theory-guided data science: a new paradigm for science discovery from data. IEEE Transactions on Knowledge and Data Engineering 29 (10), 2381-2331.

Kavetski, D., Franks, S.W., Kuczera, G., 2003. Confronting input uncertainty in environmental modeling. In: Duan, Q., Gupta, H.V., Sorooshian, S., Rousseau, A.N., Turcotte, R. (Eds.), Calibration of Watershed Models. Water Sci. Appl. Ser. AGU, Washington, D.C., pp. 49–68.

Kim, G., J. Humble, P. Debois, J. Willis, 2016. The DevOps Handbook: How to Create World-Class Agility, Reliability, & Security in Technology Organizations. IT Revolution Press, LLC.

Kouwen, F., Schot, P. P., and Wassen, M. J., 2008. A framework for linking advanced simulation models with interactive cognitive maps. Environ Modelling & Software. 23(9): 1133-1144.

Lund, Jay R. and Richard N. Palmer, 1998. Water Resource System Modeling for Conflict Resolution, Water Resources Update, University Council on Water Resources, pp. 10-15.

Madsen, H., 2003. Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. Advances in water resources, 26(2), pp.205-216.

Maier, H. R., Kapelan, Z., Kasprzyk, J., Matott, L. S., Cunha, M. C., Dandy, G. C., Gibbs, M. S., Keedwell, E., Marchi, A., Ostfeld, A., Savic, D., Solomatine, D. P., Vrugt, J. A., Zecchin, A. C., Minsker, B. S., Barbour, E. J., Kuczera, G., Pasha, F., Castelletti, A., Giuliani, M., and Reed, P. M., 2014. Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions. Environ Modelling & Software. 62: 271-299.

McGovern, A., Lagerquist, R., Gagne, D. J., Jergensen, G. E., Elmore, K. L., Homeyer, C. R., and Smith, T., 2019. Making the Black Box More Transparent: Understanding the Physical Implications of Machine Learning. Bull Amer Meteor Soc. 100(11): 2175-2199.

McKinley, D. C., Miller-Rushing, A. J., Ballard, H. L., Bonney, R., Brown, H., Cook-Patton, S. C., Evans, D. M., French, R. A., Parrish, J. K., Phillips, T. B., Ryan, S. F., Shanley. L. A., Shirk, J. L., Stepenuck, K. F., Weltzin, J. F., Wiggins, A., Boyle, O. D., Briggs, R. D., Chapin, S. F., Hewitt, D. A., Preuss, P. W., and Soukup, M. A., 2017. Citizen science can improve conservation science, natural resource management, and environmental protection. Biological Conservation. 208: 15-28.

Moradkhani, H., Hsu, K.L., Gupta, H.V., Sorooshian, S., 2005. Uncertainty assessment of hydrologic model states and parameters: sequential data assimilation using the particle filter. Water Resour. Res. 41, W05012.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D. and Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE, 50(3), pp.885-900.

Nash, J. E., and Sutcliffe, J.V., 1970. River flow forecasting through conceptual models: Part 1. A discussion of principles. J. Hydrology 10(3): 282-290.

National Research Council (NRC), 2012. Assessing the reliability of complex models: mathematical and statistical foundations of verification, validation, and uncertainty quantification. National Academies Press.

National Research Council, 2007. Models in environmental regulatory decision making. National Academies Press.

Nevala, K., 2017. The Machine Learning Primer. SAS Institute, Inc.

Omrani, H., Parmentier, B., Helbich, M., Pijanowski, B., 2019. The land transformation model-cluster framework: applying k-means and the Spark computing environment for large scale land change analytics. Environmental Modelling & Software 111, 182-191.

Pianosi, F., Beven, K., Freer, J., Hall, J.W., Rougier, J., Stephenson, D.B., Wagener, T., 2016. Sensitivity analysis of environmental models: a systematic review with practical workflow. Environmental Modelling and Software, 79, 214-232.

Pipinellis, A., 2015. GitHub Essentials: Unleash the power of collaborative workflow development using GitHub, one step at a time. Packt Publishing.

Rath, J.S., Hutton, P.H., Chen, L. and Roy, S.B., 2017. A Hybrid Empirical-Bayesian Artificial Neural Network Model of Salinity in the San Francisco Bay-Delta Estuary, Environmental Modeling and Software, 93, 193-208, DOI: 10.1016 / j.envsoft.2017.03.022.

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., Timmer, J., 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics. 25 (15), 1923-1929.

Refsgard, J.C., van der Sluijs, J.P., Hojberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process – a framework and guidance. Environmental Modelling and Software, 22(2007) 1543-1556.

Richardson, J., Sallam, R., Schlegel, K., Kronz, A., and Sun, J., 2020. Magic Quadrant for Analytics and Business Intelligence Platforms. Gartner, Inc.

Romao, T., Correia, N., Dias, E., Danado, J., Trabuco, A., Santos, C., Santos, R., Camara, A., and Nobre, E., 2004. ANTS—Augmented Environments. Computers & Graphics. 28(5): 625-633.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global Sensitivity Analysis: The Primer. John Wiley & Sons Ltd, England.

Santhi, C, Arnold, J. G., Williams, J.R., Dugas, W.A., Srinivasan, R., and Hauck, L.M., 2001. Validation of the SWAT model on a large river basin with point and nonpoint sources. J. American Water Resources Assoc. 37(5): 1169-1188.

Shamoo, A., Resni, D., 2003. Responsible Conduct of Research. Oxford University Press.

Shoemaker, L.. Lahlou, M., Bryer, M., Kumar, D., and Kratt, K., 1997. Compendium of tools for watershed assessment and TMDL development. USEPA 841-B-97-006.

Solomatine, D.P., Dibike, Y.B. and Kukuric, N., 1999. Automatic calibration of groundwater models using global optimization techniques. Hydrological Sciences Journal, 44(6), pp.879-894.

Spear, R.C., Cheng, Q. and Wu, S.L., 2020. An example of augmenting regional sensitivity analysis using machine learning software. Water Resources Research, 56(4), p.e2019WR026379.

Tasdighi, A., Arabi, M., Harmel, D., 2018, A probabilistic appraisal of rainfall-runoff modeling approaches within SWAT in mixed land use watersheds, Journal of Hydrology. 564, 476-489.

Ter Braak, C. J. F., 2006, A Markov chain Monte Carlo version of the genetic algorithm differential evolution: Easy Bayesian computing for real parameter space, Statistical Computing, 16(3), 239–249.

Tetra Tech, 2006. Conceptual Model for Nutrients in the Central Valley and Sacramento-San Joaquin Delta, Technical Report prepared for Central Valley Regional Water Quality Control Board. Available at: https://www.waterboards.ca.gov/centralvalley/water_issues/drinking_water_policy/final_nutrient_report_lowres.pdf

Thiemann, M., Trosset, M., Gupta, H. and Sorooshian, S., 2001. Bayesian recursive parameter estimation for hydrologic models. Water Resources Research, 37(10), pp.2521-2535.

Trowbridge, P.R., Deas, M., Ateljevich, E., Danner, E., Domagalski, J., Enright, C., Fleenor, W., Foe, C., Guerin, M., Senn, D., and Thompson, L.., 2016 Recommendations for a Modeling Framework to Answer Nutrient Management Questions in the Sacramento-San Joaquin Delta. Report prepared for the Central Valley Regional Water Quality Control Board.

Tscheickner-Gratl, F., Bellos, V., Schellart, A., Moreno-Rodenas, A., Muthusamy, M., Langeveld, J., Clemens, F.H.L.R., Benedeti, L., Rico-Ramirez, M.A., Fernandes de Carvalho, R., Breuer, L., Shucksmith, J., Heuvelink, G.B.M., Tait, S., 2019. Recent insights on uncertainties present in integrated catchment water quality modelling, Water Res., 150, 368-379.

Tukey, J.W., 1962. The future of data analysis. The annals of mathematical statistics, 33(1), pp.1-67.

U.S. Environmental Protection Agency, 2002. Guidance for Quality Assurance Project Plans for Modeling, EPA QA/G-5M, on the internet at: https://www.epa.gov/quality/guidance-quality-assurance-project-plans-modeling-epa-qag-5m

U.S. Environmental Protection Agency, 2006. Guidance on Systematic Planning Using the Data Quality Objectives Process, EPA QA/G-4, on the internet at: https://www.epa.gov/sites/production/files/2015-06/documents/g4-final.pdf

U.S. Environmental Protection Agency, 2019. Basic Information about Citizen Science, on the internet at: https://www.epa.gov/citizen-science/basic-information-about-citizen-science-0

van Vliet, J., Bregt, A.K., Brown, D.G., van Delden, H., Heckbert, S. and Verburg, P.H., 2016. A review of current calibration and validation practices in land-change modeling. Environmental Modelling & Software, 82, pp.174-182.

Voinov, A., Bousquet, F., 2010. Modelling with stakeholders. Environmental Modelling & Software 25 (11), 1268-1281.

Voinov, A., Kolagani, N., McCall, M. K., Glynn, P. D., Kragt, M. E., Ostermann, F. O., Pierce, S. A., Ramu, P., 2016. Modelling with stakeholders -- Next generation. Environmental Modelling & Software 77, 196-220.

Vrugt, J. A., Gupta, H. V., Bouten, W., & Sorooshian, S., 2003. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resources Research, 39(8).

Vrugt, J., 2016. Markov chain Monte Carlo simulation using the DREAM software package: theory, concepts, and MATLAB implementation. Environmental Modelling and Software, 75 (2016) 273-316.

Walker, W.E., Harremoës, P., Rotmans, J., Van Der Sluijs, J.P., Van Asselt, M.B., Janssen, P. and Krayer von Krauss, M.P., 2003. Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. Integrated assessment, 4(1), pp.5-17.

Walters, C., 1997. Challenges in adaptive management of riparian and coastal ecosystems. Conservation Ecology [online]1(2):1.

Wan, Z., Hong, Y., Khan, S., Gourley, J., Flamig, Z., Kirschbaum, D., and Tang, G., 2014. A cloud-based global flood disaster community cyber-infrastructure: Development and demonstration. Environ Modelling & Software. 58: 86-94.

Wang, X., and A. M. Melesse. 2005. Evaluation of the SWAT model's snowmelt hydrology in a northwestern Minnesota watershed. Trans. ASAE 48(4): 1359-1376.

Willmott, C. J., 1981. On the validation of models. Physical Geography 2: 184-194.

Xu, J., Zhang, Q., Liu, S., Zhang, S., Jin, S., Li, D., Wu, X., Liu, X., Li, T., and Li, H., 2020. Ensemble learning of daily river discharge modeling for two watersheds with different climates. Atmos Scie Lett. 21: e1000.

Yang, J., 2011. Convergence and uncertainty analyses in Monte-Carlo based sensitivity analysis. Environmental Modelling & Software, 26(4), pp.444-457.

Zhang, X., Srinivasan, R., Arnold, J., Izaurralde, R.C. and Bosch, D., 2011. Simultaneous calibration of surface flow and baseflow simulations: a revisit of the SWAT model calibration framework. Hydrological Processes, 25(14), pp.2313-2320.

Appendix A: Attendees at Discipline-Specific Targeted Meetings Held in Davis, CA

Biological and Ecosystems Modeling
February 26, 2020

Name	Affiliation
Travis Hinkelman	Cramer Fish Sciences
Ed Gross	Resource Management Associates, Inc. (RMA) /UC Davis
Kenny Rose	University of Maryland
Lisa Lucas	USGS
Greg Reis	The Bay Institute
Dave Smith	U.S. Army Engineer Research and Development Center (ERDC)
Ben Geske	Delta Stewardship Council
John DeGeorge	Resource Management Associates, Inc. (RMA)
Mike Deas	Watercourse Engineering
Tariq Kadir	California Department of Water Resources

Hydraulics/Hydrodynamics/Water Quality Modeling
March 2, 2020

Name	Affiliation
Tara Smith	California Department of Water Resources
Deanna Sereno	Contra Costa Water District
Bill Fleenor	UC Davis
Chuching Wang	Metropolitan Water District
Ben Bray	East Bay Municipal Utility District
Willis Hon	East Bay Municipal Utility District
Todd Steissberg	The U.S. Army Engineer Research and Development Center (ERDC)
Steve Andrews	Resource Management Associates, Inc. (RMA)
Tony Donigian	Respec
Marisa Escobar	SEI
Nicky Sandhu	California Department of Water Resources
Joel Herr	Systech
John DeGeorge	Resource Management Associates, Inc. (RMA)
Josué Medellín-Azuara	UC Merced
Mike Deas	Watercourse Engineering

Groundwater and Integrated Surface Water-Groundwater Modeling
March 4, 2020

Name	Affiliation
Tariq Kadir	California Department of Water Resources
Can Dogrul	California Department of Water Resources
Tyler Hatch	California Department of Water Resources
Linda Bond	California Department of Water Resources
Jon Traum	U.S. Geological Survey
Charlie Brush	Hydrolytics
Ali Taghavi	Woodard and Curran
Matt Tonkin	SS Papadopulos Associates
Claudia Faunt	U.S. Geological Survey
Kathryn Koczot	U.S. Geological Survey
Randy Hanson	One-Water Hydrologic

Hydroeconomic Modeling and Economic Modeling
March 4, 2020

Name	Affiliation
Alvar Escriva-Bou	Public Policy Institute of California
Brad Franklin	The Nature Conservancy
Steve Hatchett	ERA Economics
Ray Hoagland	(Retired) California Department of Water Resources
Katrina Jessoe	UC Davis
Jonathan Kaplan	CSU Sacramento
Dan Liu	State Water Resources Control Board
Richard McCann	M-Cubed
Maura Allaire	UC Irvine
Kurt Schwabe	UC Riverside
Harrison B. Zeff	University North Carolina Chapel Hill
Spencer Cole	UC Merced
Alex Guzman	UC Merced
Josué Medellín-Azuara	UC Merced
Farhad Farnam	(Retired) California Department of Water Resources
Tariq Kadir	California Department of Water Resources

Surface Water Hydrology/Management and Reservoir Operations
March 10, 2020

Name	Affiliation
Rich Satkowski	CWEMF Monitoring Protocols Committee Lead
Tariq Kadir	California Department of Water Resources
Erik Reyes	California Department of Water Resources
Nancy Parker	U.S. Bureau of Reclamation
Ali Taghavi	Woodard Curran
Rob Leaf	Jacobs
Scott Ligare	State Water Resources Control Board
Helen Dahlke	UC Davis
Kathryn Koczot	U.S. Geological Survey
Jon Butcher	Tetra Tech

Appendix B: Inventory of Models

Model Category	Model Type	Model Name
Reservoir Operations Models	**Model Frameworks**	WEAP (Water Evaluation and Planning)
		WRIMS (Water Resource Integrated Modeling System)
		HEC-ResSim (Reservoir Simulation System)
	Delta Specific Models	CalSim II
		CalSim 3
		CALVIN
		SacWAM (Sacramento Water Allocation Model)
Hydrodynamics Models	**Model Frameworks**	Delft3D-FM (Finite Mesh)
		EFDC (Environmental Fluid Dynamics Code)
		RMA2
		SCHISM (Semi-implicit Cross-scale Hydroscience Integrated System Model)
		SUNTANS (Stanford Unstructured Nonhydrostatic Terrain-following Adaptive Navier-Stokes Simulator)
		TRIM/UnTRIM (Tidal, Residual, and Intertidal Mudflat/Unstructured)
	Delta Specific Models	DSM2 (Delta Simulation Model 2)
		ANN Model Emulators for DSM2
		FDM (Fischer Delta Model)
		RMA Bay-Delta 2D/1D Model (RMA2/RMA11)
		RMA3D San Francisco Estuary Model (UnTRIM)
Human Ecology and Economics Models	**Model Frameworks**	HAZUS-MH (HAZUS Multi-Hazard Model)
		IMPLAN (IMpact Analysis for PLANning)
		REMI (Regional Economic Models, Inc)
		SWAP (Statewide Agricultural Production Model)
	Delta Specific Models	DAP (Delta Agricultural Production Model)
		F-RAM (Flood Rapid Assessment Model)
Groundwater - Surface Water Models	**Model Frameworks**	IWFM (Integrated Water Flow Model) / IDC (IWFM Demand Calculator)
		MODFLOW (USGS Modular Groundwater Flow Model)
		MODPATH
		MT3D
		PHAST (PHREEQC And HST3D)
		STANMOD
		SUTRA (Saturated-Unsaturated TRAnsport)
		C2VSIM (California Central Valley Groundwater-Surface Water Simulation Model)

	Delta Specific Models	CVHM (Central Valley Hydrologic Model)
		CVHM-D (Central Valley Hydrologic Model - Delta)
Fisheries and Ecosystems Models	**Model Frameworks**	ELAM (Eulerian-Lagrangian-Agent Method)
		inSALMO (Improvement of Salmon Life-Cycle Framework Model)
	Delta Specific Models	Delta STARS (Survival, Travel Time, and Routing Simulation)
		DPM (Delta Passage Model)
		DSLCM (Delta Smelt Life Cycle Model)
		EFT (Ecological Flow Tools)
		ePTM (Enhanced Particle Tracking Model)
		IOS (Interactive Object-Oriented Simulation)
		SacPAS Fish Model
		SALSIM (Salmon Simulator)
		WRLCM (Winter Run Life Cycle Model)
Greenhouse Gas Emissions and Land Use Models	**Model Frameworks**	CANVEG
		DAYCENT (Daily CENTURY Model)
		DNDC (DeNitrification DeComposition)
	Delta Specific Models	PEPRMT-DAMM (Peatland Ecosystem Photosynthesis, Respiration, and Methane Transport – Dual Arrhenius Michaelis-Menten)
		SUBCALC
Water Quality Models	**Model Frameworks**	CE-QUAL-W2
		HEC-5 and 5Q (Hydrologic Engineering Center)
		HEC-RAS (Hydrologic Engineering Center's River Analysis System)
		HSPF (Hydrological Simulation Program FORTRAN)
		PHREEQC (pH-REdox-Equilibrium)
		SWAT (Soil & Water Assessment Tool)
		VIC (Variable Infiltration Capacity)
		WARMF (Watershed Analysis Risk Management Framework)
		RMA11
	Delta Specific Models	SBWQM (South Bay Water Quality Model)
		USRWQM (Upper Sacramento River Water Quality Model)
		RMA Bay-Delta 2D/1D Model (RMA2/RMA11)
Soil Chemistry and Salinity Models	**Model Frameworks**	Hydrus
		Watsuit
Consumptive Use Models	**Model Frameworks**	Cal-SIMETAW (California Simulation of Evapotranspiration of Applied Water)
		CIMIS (California Irrigation Management Information System) and AmeriFlux
		DisALEXI (Atmosphere-Land Exchange Inverse (ALEXI) flux disaggregation approach)
		ITRC-METRIC (Mapping of EvapoTranspiration with Internal Calibration)
		SIMS (TOPS Satellite Irrigation Management Support)
	Delta Specific Models	DETAW (Delta Evapotranspiration of Applied Water)
		DICU (Delta Island Consumptive Use)

Processing and Visualization Tools	Groundwater Vistas
	HEC-DSSVue (Hydrologic Engineering Center's Data Storage System Visual Utility Engine)
	ModelMuse
	PEST
	T-PROGS (Transition Probability Geostatistical Software)
	USGS Model Viewer
	Visual MODFLOW

The most up-to-date model inventory can be accessed online at: https://cwemfwiki.atlassian.net/wiki/spaces/MI/overview.

CWEMF
California Water & Environmental Modeling Forum