Integrated Modeling Support
Delta Stewardship Council Contract #17400

Memo 3. Challenges and Solutions for Model Integration and Related Data Needs

January 2020

Prepared for:

Delta Stewardship Council
Ben Geske, P.E., Project Manager

Project Team:

Tetra Tech, Inc.
Bachand & Associates
Cramer Fish Sciences
HydroFocus, Inc.
Pax Environmental
University of California at Davis
University of California at Merced

Acknowledgements

This work was performed under contract (#17400) between the Delta Stewardship Council (DSC) and Tetra Tech Inc. Mr. Ben Geske of the Delta Stewardship Council served as the project manager. The DSC's Integrated Modeling Steering Committee, consisting of modeling experts from state and federal agencies and private consulting firms, provided guidance on the study.

Project Team

Tetra Tech: Sujoy Roy, Paul Hutton, Katherine Heidel, Tom Grieb, and Ali Tasdighi

Bachand Associates: Philip Bachand and Yan Liang

Cramer Fish Sciences: Travis Hinkelman

Hydrofocus: Steve Deverel

Pax Environmental: Tom Lagerquist

University of California at Merced: Josué Medellín-Azuara and Anna Rallings

University of California at Davis: Anne Visser and Leslie Panyanouvong

Contents

Figures

Tables

Executive Summary

This memo evaluates the current state of challenges, potential solutions, and data needs within the context of Delta model integration. This information provides a foundation for specific recommendations for an integrated modeling strategy for the Delta.

Integrated modeling is defined as an approach where two or more models, typically with different areas of focus, are applied jointly in an analysis. This approach is often needed for system-level analysis of complex environmental problems that cross physical, chemical, biological, social and economic domains. 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 development of entirely new codes. Our review of major project initiatives in the Delta found that model integration was being used widely, notably in the physical, chemical, and biological domains, with growing opportunities in the economic domain, and emerging opportunities in the other social sciences. In general, integrated modeling in the Delta was found to provide utility when evaluating complex, high-stakes initiatives if supported by sufficient resources and if the missions and goals of the participating agencies or organizations are aligned to the modeling needs.

Our assessment of modeling needs suggests a wide variety of issues—pertinent today and in the foreseeable future—which could be addressed effectively through an integrated modeling framework. Identified modeling needs include continued support for regulatory actions under current laws; exploratory analyses and adaptation related to anticipated future conditions driven by climate change; developing better understanding of the interactions of different physical, chemical, and biological processes; and opportunities for more explicitly considering the dynamic role of humans in the landscape.

We identified challenges in model integration as part of this work. Although integrated modeling across different spatial and disciplinary domains can be beneficial in addressing complex environmental problems, the added complexity of getting two or more models to work together effectively raises some practical challenges. These challenges are grouped into two broad categories: institutional and technical. 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. 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. An assessment of these challenges shows that model integration is not driven by modelers alone. Even when the technical challenges of integration are solvable by modeling teams, successful development of integrated models will require other participants in the modeling process, such as model sponsors and other stakeholders, to address institutional challenges.

In our assessment of model integration challenges that arise around participating organizations and people, we identify different actions that can help stimulate the development of integrated models, including institutional commitment and leadership support, model community development, and education. Modeling communities can take the form of user groups (many of which are already in existence), a virtual community of practice, or a physical location for interested participants to work together (i.e. collaboratory). Community engagement across participating agencies is also fostered by various regional, state, and national forums that involve exchange among modelers and scientists. Institutional efforts for model integration also include education for current and future students, staff in participating organizations, as well as the broader stakeholder community. These institutional challenges, while distinct from technical challenges, are equally important to address for the long-term success of model integration in the Delta.

Several technological approaches to facilitate model integration were identified:

- Model documentation is an obvious and straightforward approach; this documentation should address model structure and processes and the data being exchanged between models. Documentation minimizes the opportunities for error in translation across models, a major concern in any model integration effort.
- User interfaces, while not essential for model integration *per se*, allows greater accessibility and understanding of data input and output needs, and is therefore beneficial for cross-disciplinary interaction.
- Data exchange standards are an essential element for creating frameworks that allow models to share information among one another in various dynamic formats. Several such data exchange frameworks are in active development in the environmental domain to promote efficient and transparent inter-model communication.
- Formal evaluation of uncertainty propagation in linked models is a technological approach that can promote more informed use of model results in decision making.

Such analysis can be highly computationally demanding and is currently the subject of research.

- Model emulation, an approach that replaces a complex model with a simpler approximation, reduces computational requirements. In many cases, emulators can be embedded within another model. Several emulation approaches are available, with many being used in the Delta.
- Adoption of big data approaches can facilitate integrated modeling. Related analysis tools are undergoing rapid development, especially in the commercial realm. Some environmental applications of these tools are beginning to appear, and given the potential utility of these tools for management and integrated data analysis, many future applications will likely develop. These likely developments include standalone models as well as hybrid models that combine data-based approaches with process-based models.

Overall, our review suggests that technological approaches to facilitate model integration are developing rapidly in the environmental domain and other related domains. These approaches offer many different avenues for linking models and creating new integrated modeling frameworks to support future decision-making needs.

Data needs across a range of Delta-relevant domains for model integration were discussed, with the goals of providing a general reference for modelers working across disciplines and identifying data gaps where appropriate. As part of this work, we evaluated several disciplines including: hydrodynamics, ecology, water quality, fish species, water budgets and consumptive use, agricultural economics, and socioeconomics. Large data collection efforts are ongoing in many of these areas; however, limited coordination is taking place across disciplines. As model integration becomes more commonplace across the Delta, coordinated sampling efforts in time and space will be needed to make the best use of these data for modeling.

Glossary

Term	Definition
Big data	Big data, although informal in origin, refers to data that are high in volume, velocity, and variety, requiring new technologies and techniques to capture, store, and analyze.
Black -box	In the context of modeling, an approach where the internal model structure is not necessarily visible to or interpretible by the user.
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.
Conceptual model	A high-level representation of inputs, interacting processes and drivers, and outputs for any kind of process (e.g., physical, biological, economic, etc.). Although a conceptual model may include quantitative information, it is often presented in non-quantitative form and serves to communicate the model structure in a transparent manner. A conceptual model may be developed as a communication tool following the completion of a modeling study, or, during the initiation of the project, the conceptual model serves as the basis for selection of or development of a quantitative model.
Downstream model	In the context of integration of models, this refers to a form of data exchange where outputs from one model (the upstream model) are fed to another (the downstream model), and where the outcomes of downstream model have no effect on upstream model.
Emulator	Computationally simplified model representations that use relationships between inputs and outputs. Emulators are typically developed to reduce the computational cost of model exploration.
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.

Term	Definition
Federated	A term in use in the data management literature, referring to datasets managed independently, within a common framework and with consistent standards. A similar approach may also apply to a set of models.
Feedback	In the context of model integration, this refers to the two-way exchange of data between models. Thus, where two models are integrated with feedback, the outputs from both models can serve as inputs to the other model.
Initial condition	The solution of a differential equation 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 need the definition of initial conditions.
Inline integration	In the context of model integration, this refers to a model structure where data exchange between two or more model components occurs within an integrated code, with minimal human processing of outputs from internal models. Similar to tight coupling.
Loose coupling	In the context of model integration, this refers to a model structure where data exchange between two or more model components occurs using output files at the completion of one model being fed to another model, often with human processing in between. Similar to offline integration.
Machine learning	In the context of environmental modeling, machine learning refers to a class of algorithms that are used to derive patterns or relationships between input and output data across different dimensions. The term training is often used for the process of calibrating a machine learning model. Most machine learning models are black-box representations of the data provided, and the underling relationships are generally not possible to infer directly.
Metadata	A set of data that describes and gives information about other data.
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.
Numerical model	Many quantitative models are represented by differential equations that cannot be solved exactly (i.e. analytically) because of domain or mathematical complexity. Numerical solutions (such as finite elements or finite differences) are commonly-used approaches to estimate the solutions of differential equations. Models that employ such numerical solutions are particularly common in the representation of physical and chemical processes, and are termed numerical models.
Offline integration	In the context of model integration, this refers to a model structure where data exchange between two or more model components occurs using output files at the completion of one model being fed to another model, often with human processing in between. Similar to loose coupling.

Term	Definition
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.
Tight coupling	In the context of model integration, this refers to a model structure where data exchange between two or more model components occurs within an integrated code, with minimal human processing of outputs from internal models. Similar to inline integration.
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 inputs and model parameters are known with different degrees of error, the goal of uncertainty analysis is to quantify the range of outputs in a modeling study.
Upstream model	In the context of integration of models, this refers to a form of data exchange where outputs from one model (the upstream model) are fed to another (the downstream model), and where the outcomes of downstream model have no effect on upstream model.
Validation	A term in common use in many modeling communities, validation refers to the process of applying a calibrated model to an independent set of observed data to assess whether the model fit is acceptable. A criticism of the term validation is that the process does not prove that a model is valid, but rather demonstrates performance over a limited range of conditions. The term evaluation is sometimes recommended as an alternative.

1 Introduction

There is greater recognition than ever before that human activities have a broad range of influences on natural systems, and conversely, human activities are affected by natural systems. However, interactions between environmental processes and human activities are complex, and the evaluation of these interactions transcends individual disciplines. Model representation of these interactions is often needed to support various decision-making processes (including facility planning, short-term forecasting, and regulation development and analysis) and science initiatives. Integrated modeling is conceived as a general approach to address these broad problems. We define integrated modeling as follows:

Integrated modeling is defined as an approach where two or more models, typically with different areas of focus, are applied jointly in an analysis. At its most general, the component models in an integrated modeling framework may focus on the same processes over different geographic areas or may originate from different disciplines.

In contrast, we define discipline-specific modeling as an approach that originates in a specific, mature field of study, with a focus on a limited set of processes, such as in hydrology, fluid mechanics and hydrodynamics, hydrogeology, biogeochemistry, economics, etc. It is recognized that, in many instances, sophisticated models already exist that integrate processes across disciplines, and thus the boundary between integrated and discipline-specific models is not a rigid one. Indeed, many good practices for developing integrated models may also apply to discipline-specific models, and vice versa.

Over the past two decades, integrated modeling has emerged as a sub-discipline within the larger field of environmental modeling. Rapid decreases in the cost of computer resources, including flexible resources such as cloud computing and storage, mean that integration of models to address interdisciplinary problems is now computationally feasible. There is a growing global literature on integrated modeling methodologies and applications. Key aspirations for the growing field of integrated modeling include common terminologies for variables across different disciplines, data management strategies to allow efficient integration, common standards for data exchange between models, institutional infrastructure to permit integration across agencies, and scientific understanding of new challenges that arise from integration, such as the propagation of uncertainty and the difficulties of calibration across multiple models.

This work is part of a larger study focused on the development of an integrated modeling strategy for the Sacramento-San Joaquin Delta (i.e. Delta), the region of interest for the Delta Stewardship Council. The Delta Science Program has prepared a Science Action Agenda, highlighting priorities over 2017-2021. A highly relevant priority action identified in this document (under "Modernize monitoring, data management, and modeling") is described as follows:

Advance integrated modeling through efforts such as an open Delta collaboratory (physical or virtual) that promotes the use of models in guiding policy.

The present work is a tangible step in the implementation of this priority action. It is expected that integrated modeling will provide decision makers with the best possible insight into multi-faceted environmental problems. The use of integrated modeling has been proposed for exploring adaptive management, a key component of long-term restoration planning in the Delta. Toward this end, we have summarized the current state of practice of Delta integrated modeling (Memo 2, *A Survey of Recent Integrated Modeling Applications in the Delta and Central Valley*). The goals of this memo include: providing a broad overview of integrated modeling and associated terminology, summarizing current and future integrated modeling needs, providing an overview of the challenges in model integration, and identifying potential solutions to address these challenges. This memo also discusses data needs for model development in different domains that are likely to be components of larger integrated modeling frameworks in the Delta. Best practices for model development, which apply equally to individual discipline-specific models and to integrated models, are discussed separately in Memo 4 (*Recommendations for Modeling Best Practices*).

1.1 Approaches for Model Integration and Associated Terminology

Three common approaches for model integration are shown in Figure 1. For the sake of clarity, this illustration assumes a simple case of two models. However, the concept can be generalized to a larger number of models.

- In Case 1, models may be linked to one another by means of external data transfer (i.e., through file exchange, typically performed with some manual intervention). This approach is also referred to as "off-line" coupling or "loose" coupling.
- In Case 2 (a more structured approach), models may be linked more directly through code. In this approach, data are internally exchanged at the computational time step level with minimal external intervention using a common data exchange format. This approach is referred to as "in-line" coupling or "tight" coupling. The benefit of this approach (relative to Case 3) is that required modifications to component model codes are limited to input-output routines.

- In Case 3, the models are re-written as a single code with internal data flows determined by the needs of the processes represented.

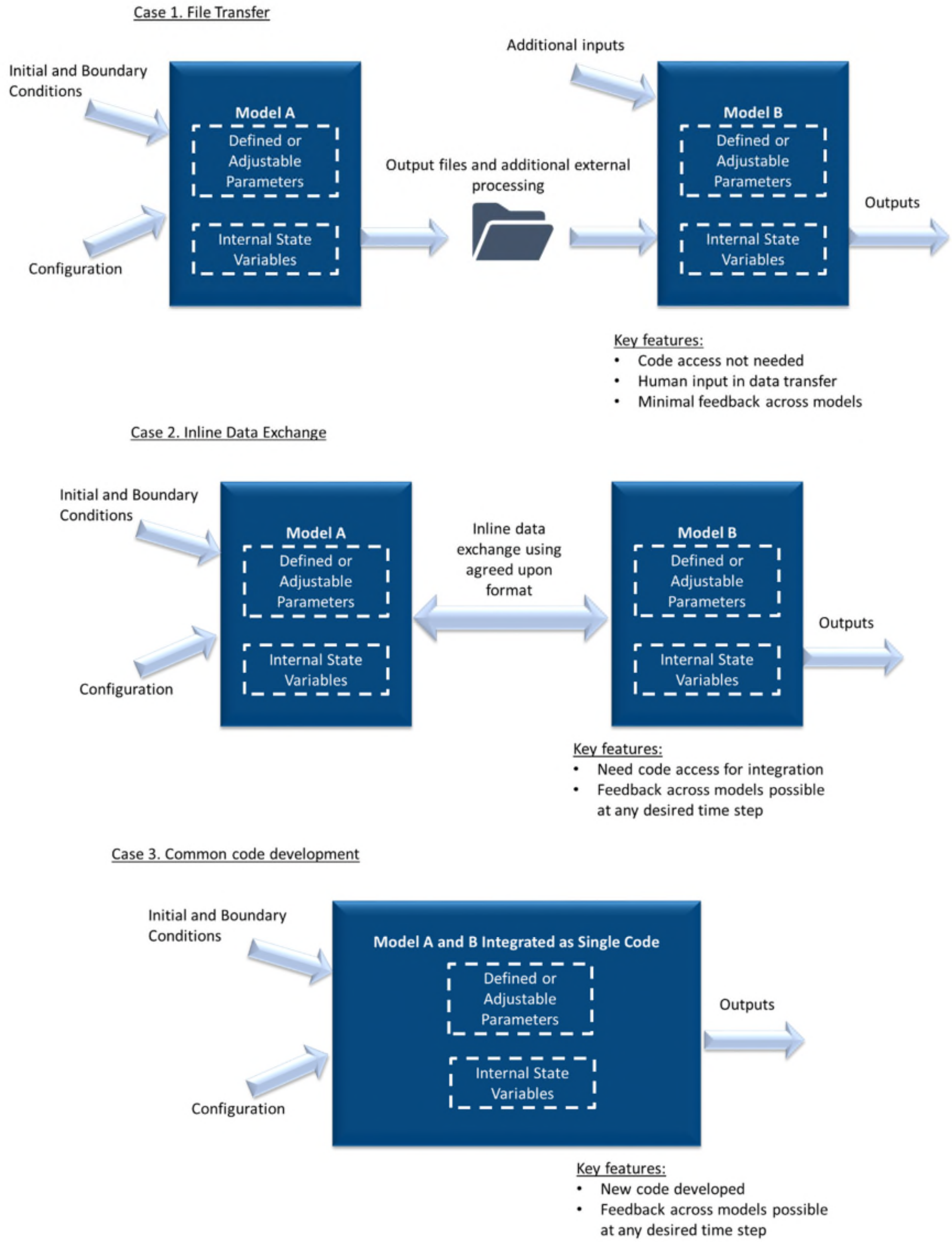

Figure 1. Three common approaches for model integration

In Cases 2 and 3, data are exchanged internally and not necessarily reported to the user and a single set of output data is produced. Cases 2 and 3 both require access to the original component model source codes. These types of integration are needed when there is feedback between two different models, i.e. where the results of each model influence the results of the other model. In our review of existing integrated modeling efforts, we found that most instances of integrated model applications in the Delta area fall under Case 1, with some notable exceptions that fall under Case 3.

We commonly refer to “upstream” and “downstream” models throughout this memo. In the context of model integration, an upstream model is one that provides input to a downstream model, and is typically used in the case where there is no feedback, i.e., the results of the downstream model do not influence the upstream model.

1.2 Types of Models to be Integrated

Table 1 summarizes a variety of modeling types or approaches used in the Delta. Selection of model type, as part of a model development process, is dictated by the potential availability of underlying theoretical frameworks and observations, the intended model use, and the technical discipline.

- Analytical or numerical models are often used where the underlying mechanisms can be explained through basic process representations. Numerical models, which solve differential equations over space and time, are in widespread use, especially in the flow and water quality domains. Over time, such models have tended to grow more complex, with greater spatial and temporal resolution, and associated computational demands. 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. Analytical solutions are also important for testing the computer implementation of numerical models, which are prone to solution errors.
- Statistical/empirical models are usually based on observed data and with limited underlying process representation. Larger datasets often improve performance of statistical models.
- Optimization-based models, notably water allocation models for the Central Valley (e.g. CalSim), have a relatively simple physical representation of processes (although the actual system may be very complex), and are focused around optimization of outcomes under specified constraints. In the case of CalSim, the outcomes are the water allocations to different users across the Central Valley and Delta, constrained by water availability, environmental flow requirements, and the hierarchy of water rights.
- Machine learning models, a class of statistical/empirical models, are identified here as a distinct model type because they offer a wide variety of emerging algorithms to find patterns or relationships in observed data. Unlike most statistical/empirical models, machine learning models may contain large numbers of fitting parameters that are not visible to a user.

- Agent-based models represent system with agents (e.g. organisms, individuals, or households) that have individual behavior and respond to external drivers or to each other.

The general task of model integration is to match component model input and output requirements within a single framework. Table 1 identifies the types of models that may be integrated and key considerations in getting these different types of models to work with one another.

1.3 Prior Synthesis Efforts

Integrated modeling strategies have been the focus of major cross disciplinary meetings in the U.S. and Europe over the past 20 years, including a large-scale workshop on Delta integrated modeling in 2015. A high-level summary of workshops that have influenced the current state of thinking on this subject is shown in Table 2. The results from these workshops are well documented, are represented in the literature, and are referred to throughout this memo. While there is a concern with creating overly complex models, a consensus that has emerged in these forums is that integrated modeling is beneficial to the study of complex environmental systems. However, this modeler-oriented perspective needs to be considered within a broader community of model sponsors, model users and other stakeholders. In this memo, informed by specific case studies summarized below (and described in greater detail in Memo 2), we take a neutral perspective on integrated modeling in the Delta and describe challenges and potential solutions for its wider implementation.

1.4 Ongoing Model Integration in the Delta

Our review of project initiatives in the Delta region (see Memo 2) revealed that model integration was being used widely, notably in the physical, chemical, and biological domains, with growing opportunities in the economic domain. We evaluated major project initiatives (see Table 3) by interviewing key participants and reviewing project information in the public domain. The most common applications evaluate a single environmental variable through space, such as when the flow of water is tracked for purposes of water supply and flood control from the upper watershed, through the man-made reservoir systems, into the Central Valley, and through the Delta and Bay. Models that consider water resources and economics (costs of flood protection or agricultural economics) are also in common use. Emerging applications consider the interaction of water flows with water quality and ecosystem processes. In general, integrated modeling approaches have been used to evaluate complex, high-stake Delta initiatives; these approaches are most effective when supported by sufficient resources and when the missions and goals of the participating agencies or organizations are aligned with the modeling needs. Insights gleaned from these previous efforts are incorporated throughout this memo.

Table 1. Types of models to be integrated

Model type	Feature	Key Considerations in Integration with Other Models
Analytical/Numerical	Solving a framework of process equations, either in closed analytical form or numerically; model parameters calibrated with observed data	Analytical models, because of the closed form nature of the solution of differential equations, often use limited spatial and temporal variation. Numerical models are often spatially and temporally detailed, with high-frequency outputs over fine grids (in 1-D, 2-D, or 3-D). Integration may need to match this scale of output by averaging over time or space.
Statistical/empirical	Limited process representation; model parameters calibrated with observed data	Less spatially detailed than analytical/numerical models; output often organized around scale of observations in the field. Other models may be constrained to work with this scale of output.
Optimization based	Focused on meeting key objectives under a range of input conditions	Constrained by optimization criteria, often defined at specific locations where compliance with specific targets is needed. Potentially less detailed representation than analytical/numerical models, and integration must align with the locations where optimization is focused on.
Machine-learning based	"Black-box" representation; trained on available data	Limited to specific locations or conditions, and other models needed to work with this constraint. Poor performance outside of training range must be considered in all phases of integration.
Agent-based	Represents behavior of organisms or populations (animal or human) in response to external factors.	Most experience is with exploring fish behavior with limited feedback with other models. Consideration of humans as an agent can create different feedbacks across models.

Table 2. Summary of integrated modeling workshops

Workshop Title	Date	Output
Environmental Software Systems Compatibility and Linkage Workshop	March 2000	Proceedings of the Environmental Software Systems Compatibility and Linkage Workshop (NRC 2002)
Integrated Modeling for Integrated Environmental Decision Making	January 2007	Integrated Modeling for Integrated Environmental Decision Making, EPA-100-R-08-010 (EPA 2008a)
Collaborative Approaches to Integrated Modeling: Better Integration for Better Decision making	December 2008	Workshop Report: Collaborative Approaches to Integrated Modeling: Better Integration for Better Decision-Making (EPA 2008b)
iEMSs 2010 Conference; Science session: Integrated Modeling Technologies; Workshop: The Future of Science and Technology of Integrated Modeling	July 2010	Integrated environmental modeling: a vision and roadmap for the future (Laniak et al 2013)
The International Summit on Integrated Environmental Modeling	December 2010	International Summit on Integrated Environmental Modeling Workshop Report (Moore et al 2012)

Workshop Title	Date	Output
Integrated Modeling for Adaptive Management of Estuarine Systems	May 2015	White paper on workshop (Medellín-Azuara et al 2016)

Table 3. Project initiatives evaluated as part of the integrated modeling survey (Memo 2).

Item	Project Initiative	Lead Agency	Project Description
1	California WaterFix	California Department of Water Resources (DWR)	Major proposed infrastructure project to construct tunnels under the Delta. Although this project is being re-configured, the modeling observations still pertain to the broader issues of integration.
2	Levee Assessment, Storage, Flood Management and New Infrastructure	DWR	Support for the following programs in the Delta and Central Valley: (1) Central Valley and statewide flood management planning, (2) Delta risk management planning and investment prioritization, (3) flood and ecosystem restoration feasibility investigations, (4) storage project economic justification and operation planning, and (5) Delta conveyance economic justification.
3	Socioeconomic Issues	Multiple	Challenges and opportunities surrounding integrated modeling as it relates to socioeconomic issues in the Delta.
4	Bay-Delta Water Quality Control Plan Updates	California State Water Resources Control Board (SWRCB)	New flow and salinity standards in the Delta being updated by the SWRCB.
5	Water Rights, Consumptive Use & Water Budgets	SWRCB	Consumptive use modeling and measurement for crops and other land use cover in the Delta.
6	Delta Smelt Biological Opinion	US Fish and Wildlife Service (USFWS)	Modeling of Delta Smelt behavior and population dynamics to support ongoing Biological Opinion re-consultation.
7	Central Delta Corridor/Future Carbon Markets	Central Delta Conservancy	Multi-agency effort to assess options for greater sustainability on publicly owned lands in the western and central Delta.
8	California EcoRestore	DWR	Multi-agency effort to restore 30,000 acres of habitat in a set of discrete projects across Delta islands.
9	Yolo Bypass Models	DWR	Water and environmental modeling by DWR and other agencies for the Yolo Bypass, a seasonally inundated floodplain used for flood protection, agriculture, fish populations, and migratory waterfowl.
10	Delta Methylmercury Total Maximum Daily Load Modeling	DWR	Evaluation of relationship between methylmercury loads and concentrations and water project operations.

Item	Project Initiative	Lead Agency	Project Description
11	CASCADEII Model Framework	U.S. Geological Survey (USGS)	Integrated model development and study of climate, hydrology, hydrodynamics, sediment, phytoplankton, bivalves, contaminants, marsh accretion, and fish populations.
12	AFRI Rice Agriculture Modeling	Multiple academic and consulting groups	Rice agriculture in the Delta to provide alternative income source with added benefits for subsidence mitigation, levee stability, and ecosystem services. Various aspects of land use were monitored and modeled.
13	Modeling for Climate Change Vulnerability Assessment and Adaptation Strategy for the Delta and Suisun Marsh	Delta Stewardship Council (DSC)	Project aims to 1) characterize climate change exposure, sensitivity, and adaptive capacity in the Delta to provides decision relevant information and 2) create adaptation strategies to support the achievement of the Delta Plan's coequal goals and to reduce impacts.
14	Managed Aquifer Recharge using Floodwater (FloodMAR)	DWR	Groundwater recharge using flood flows to increase water security and mitigate downstream flood risks. Modeling used to understand climate-driven surface water allocation and potential for groundwater recharge.
15	Franks Tract Restoration Feasibility	DWR	Hydrodynamic and water quality modeling to evaluate effects of different conceptual restoration designs.
16	Chinook Salmon Life Cycle Model	National Oceanic and Atmospheric Administration (NOAA) Fisheries	Mechanistic evaluation of juvenile Chinook salmon life cycle.

1.5 Summary

Integrated modeling is an approach where two or more models, typically with different areas of focus, are used together in an analysis. This approach can be applied to support analyses that cross physical, chemical, biological, social and economic domains. 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 codes. A review of major project initiatives in the Delta found that model integration was being used widely, notably in the physical, chemical, and biological domains, with growing opportunities in the economic domain, and emerging opportunities in the other social sciences. Our assessment of these projects supports the notion that integrated modeling in the Delta can provide utility when evaluating complex, high stake initiatives if supported by sufficient resources and if the missions and goals of the participating agencies or organizations are aligned to the modeling needs.

Chapter 2 presents a list of future integrated modeling needs. Chapter 3 describes commonly encountered challenges in developing integrated models; these challenges are classified as institutional and technical issues. Chapters 4 and 5 identify solutions to address institutional and technological challenges in model integration, respectively. Finally, Chapter 6 describes discipline-specific data needs within the context of integrated modeling.

2 Future Needs for Integrated Modeling

The 2017-2021 Science Action Agenda (DSC, 2017) identifies the advancement of integrated modeling as a goal to support policymaking in the Delta. Numerous organic efforts at model integration are already underway in California; nonetheless, identifying activities that may benefit from integrated modeling remains as important as ever.

A review of modeling needs suggests a wide variety of issues—pertinent today and in the foreseeable future—which could be addressed effectively through an integrated modeling framework. Given our experience working in different disciplines, we identify a list of future modeling needs in Table 4. The modeling needs include continued support for regulatory actions under current laws; exploratory analyses and adaptation related to anticipated future conditions driven by climate change; developing better understanding of the interactions of different physical, chemical, and biological processes; and opportunities for more explicitly considering the dynamic role of humans in the landscape. This list spans project areas that employ some form of integrated modeling for policy-oriented as well as research-oriented support and decision making. Many of these research-oriented project areas may evolve into policy-oriented project areas in the future. Based on our present understanding, these project areas will remain active in future decades, even as conditions in the Delta evolve and improvements in data gathering and computational capabilities are made.

Future modeling needs for project areas listed in Table 4 are expected to involve solutions drawn from a wide variety of disciplines which have their own, well-developed modeling frameworks. To address this range of topics, we recognize that it is not generally practical to develop single models encapsulating more and more relevant processes. These mega-models would be difficult to create and manage and would not make use of existing models and insights developed through them. Integration of available models is thus a reasonable alternative and can be advanced through one of the three approaches shown in Figure 1. While integration of available models is not without its challenges, the typical experience in the Delta and elsewhere is that the use of existing models as modules or building blocks (within a more complex model framework) is a practical solution for meeting future needs.

The information summarized in Table 4 provides the motivation for future investments to support integrated modeling in the Delta. While unanticipated needs may arise, the general methodologies for integration that are the focus of this work will continue to remain relevant.

Table 4. Example integrated modeling needs

Modeling Need	Description of Modeling Need
1. Development of Delta salinity standards.	Salinity is regulated at different compliance locations in the Delta by the State Water Resources Control Board to meet various ecological and human beneficial uses. These regulations are subject to regular updates; updates are currently in progress.
2. Development of biological opinions for key endangered aquatic species present in the San Francisco Estuary.	Biological opinions are developed by the U.S. Fish and Wildlife Service and National Marine Fishery Services to propose conditions for the sustainability of threatened and endangered species. Updates to these opinions are driven by conditions in the field.
3. Climate change impacts on water supplies, water demands, and flooding in the Central Valley and Delta.	Climate change impacts on sea level, precipitation volume and timing, and temperatures, are expected to have complex effects on agroecosystems, flooding potential, estuarine water quality, the water supply system, and municipal and agricultural demands. Climate change will also have a variety of impacts on communities dependent on the Delta. Modeling is needed to understand the range of inter-related impacts across these sectors.
4. Climate adaptation planning, costs and relationship to Delta communities.	Adaptation efforts include changes in water systems operations, regulatory actions, and engineering approaches. Changes are applied at different spatial scales. Modeling is needed to relate adaptation to future impacts (previous item above), explore changes in the Delta, and to develop cost estimates across strategies.
5. Implementation of Sustainable Groundwater Management Act (SGMA) impacts on groundwater-surface water systems, agricultural production and regional economics.	The management of groundwater across California is undergoing dramatic change as a result of the implementation of SGMA. The need for modeling is anticipated across groundwater and surface water basins, groundwater dependent ecosystems, as well as agricultural patterns and economic impacts.
6. Impacts of wetland restoration on Delta flows, water levels, water quality, ecosystems, and Delta communities.	Efforts to restore natural tidal wetlands on some Delta islands is envisioned as part of EcoRestore (see Memo 2). Potential impacts associated with these efforts are investigated through models, including direct impacts to Delta hydrodynamics, flooding, and water quality and indirect impacts on ecosystems and Delta communities.
7. Impact of Delta island subsidence on the future agricultural economy and water quality of the region.	Rapid subsidence on Delta islands, especially those with the highest organic content soils, poses levee failure risk. This risk affects Delta agricultural production, local communities, water quality, and the economy at large.
8. Impacts of new reservoir regulations and project operations on water quality and endangered species.	Changes in reservoir regulation for various reasons (e.g. FloodMAR and climate adaptation) have downstream effects on ecosystems and water quality that can be examined through models.
9. Effects of nutrients on Delta food webs and endangered species.	The effects of nutrients (specifically nitrogen and phosphorus species) on the aquatic ecosystem are of growing interest in the Delta and in downstream waters such as the Bay and coastal ecosystem. Effects include changes in algal communities and overall food web, harmful algal blooms, and low dissolved oxygen. These effects are evaluated through a mix of hydrodynamic, water quality, and food web models.

Modeling Need	Description of Modeling Need
10. Effect of emergency events such as levee failure or spills on water supply, water quality, project operations.	Emergency events in the Delta—driven by earthquakes, major storms, spills, and even sunny-day failures—may have broad ranging impacts across its different uses. Modeling across disciplines, ideally with structures set up in advance, are needed to evaluate the varied impacts.
11. Changes to water allocation given changes in hydrology (in the near term), regulations (in the medium term) and to climate change (in the long term).	Modeling project operations in conjunction with estuarine processes requires analysis across different time scales to evaluate responses due to hydrologic variability, regulatory changes for water quality and biological opinions, and changes in precipitation and mean sea level due to climate change.
12. Integrated management of flood peaks and groundwater recharge to improve groundwater sustainability.	Large flood flows that occur in some wet years are an opportunity to capture additional water supplies in California. Related planning activities require the integration of models for reservoirs, surface flows, and groundwater basins.
13. Effect of changing crop types on water use, water quality in Central Valley groundwater and impacts downstream.	Variations in cropping patterns across the Central Valley, driven by economics at the farm scale, as well as new regulations (such as SGMA noted above), have effects on water demands and water quality in surface and groundwaters. For long-term planning, integrated modeling is needed to evaluate these changes.
14. Impacts of innovation in monitoring, data collection, telemetry on water resources management in the Delta.	New data collection techniques driven in part by new sensor technologies on the ground, improved and more accessible remote sensing technologies, and new communication technologies potentially allow an entirely different perspective on monitoring. Data may be collected at much finer spatial and temporal scales and across a wider range of parameters, and with an associated need to assess with new integrated models.
15. Integrate biogeochemical processes across the Sierras, the Central Valley, Delta, Bay, and coastal regions.	There is growing scientific interest in biogeochemical processes which integrate flows, water quality, and ecosystem impacts over a large scale and requires understanding of human drivers at regional and global scales. These analyses require integration of larger earth system scale models with more localized models.
16. Sociohydrologic modeling to study the co-evolution of human and hydrologic systems.	Modeling typically assumes human behavior as fixed. However, there is growing research interest in incorporating human behavior as a variable in water resources modeling; this relatively new research field is termed sociohydrology. Integrated models of natural and human systems can aid exploration of the future evolution of communities across the Delta region.

3 Challenges for Integrated Modeling

Integrated modeling as a technique for addressing complex environmental problems has existed for at least two decades. As summarized in Chapter 1 and described in more detail in Memo 2, there are several notable examples of the use of integrated models for addressing a variety of Delta issues. The use of integrated models is expected to grow, in large part driven by stakeholder and decision-maker needs for developing a better understanding of the multi-faceted water resources problems they are required to address. Informed by the specific case studies in Memo 2 and the scientific literature on integrated modeling, we address here some of the practical challenges associated with developing true integration across model domains. While any modeling study may have challenges, the challenges described here are *specific to model integration* with two or more models. General modeling challenges and the practices that have evolved to address them are addressed in *Modeling Best Practices* (Memo 4).

Integration challenges are classified into two categories: institutional (“people”) issues and technical (“model”) issues. Possible solutions to these challenges are presented in Chapter 4 and form the basis for recommendations presented in the Synthesis Paper on Integrated Modeling (Memo 5).

3.1 Institutional (“People”) Issues in Model Integration

Institutional challenges involve people, from organizational management to technical staff performing the modeling studies to stakeholders engaged in working with the outcomes of a modeling study. We describe these issues under the following headings: setting, modeling expertise, stable funding sources, and stakeholder engagement and trust.

3.1.1 Setting

3.1.1.1 Model development within "siloed" institutions

Models are often developed in a "silo" environment. They are maintained by individual organizations (or groups within larger organizations) utilizing a focused expertise for a focused purpose with priorities set within the organization's institutional mission. As such, there is a potential barrier to the broader development and use of such model by specialists in different disciplines. An integrated model, especially where it crosses disciplinary boundaries, may not have an obvious institutional host or supporting funding stream for model maintenance and development. Similarly, there may be a lack of adequate data to support the modeling across a broader disciplinary or geographic domain. Managers responsible for developing and using such models must formally address this gap through proactive measures such as cross-training of staff in relevant disciplines, regular exchange of modelers from different areas, and finding computer resources (servers, databases) to allow joint development of the integrated model.

3.1.1.2 Delta as a setting with competing needs

The Delta is an epicenter of competing needs. The region's waterways serve as a major conduit for California's complex water system. The region's resources support competing agricultural, environmental and ecological beneficial uses and bolster local and regional economies. Debate among local communities, downstream users, environmental interests, and water managers continues in the search for common ground. Stakeholder agreement on many competing interests is difficult to find.

3.1.1.3 Project-driven modeling and model continuity

The majority of model integration efforts in the Delta and Central Valley—performed by agencies as well as academic institutions— are project driven, focused on specific outcomes with restricted timelines. Typically, modeling teams are focused on delivering specific results required for decision-making, and development and retention of the underlying capabilities related to model integration play a secondary role. Because time is an important driver, most integration efforts seek the path with the greatest probability of meeting deadlines. Therefore, existing models (both proprietary and public-domain) are used to the greatest extent possible. Specific tools and approaches developed for integration across models, including tools for pre- and post-processing, may not be fully documented or readily available for future applications. Institutional continuity of model integration is also a concern: when model integration is achieved for a specific study goal across different participating groups (such as agencies, universities, and private entities), there are no mechanisms to retain the integrated structure for the long term, and institutional knowledge is eventually lost.

3.1.2 Modeling Expertise

3.1.2.1 Discipline-specific training

Virtually all models in use today in the Delta emerged from specific disciplines such as hydraulics, hydrogeology, aquatic chemistry, and fisheries science. Complex models in these domains require a high level of discipline-specific training as well as model-specific knowledge. Availability of experts with the necessary background to perform these modeling analyses is often limited. Institutional expertise may also be a limiting factor for integrated modeling. Many specialized models are housed within specific organizations where there are appropriate mechanisms to hire, train, and maintain staff skills over time. Such a culture may work well for individual model development and use; however, it may pose a significant challenge for integration across models from different disciplines.

3.1.2.2 Rapidly changing science

The science underlying many models, especially those related to ecosystems and human-ecosystem interactions, is maturing and model representations continue to evolve. The potential benefits of integrating models with different levels of scientific maturity may be limited. An unpredictable amount of time may elapse between the science being peer-reviewed or published to putting that science into code.

3.1.2.3 Limited software engineering capacity

Today's models often have sophisticated graphical user interfaces (GUIs) at the front-end and at the back-end for processing output data or interacting with databases to store and manage data on remote servers. For computationally complex models, there is a need to use new hardware (e.g cloud servers or dedicated server clusters) to promote acceptable run times. These needs come with the need for increasing sophisticated software development. Software professionals with the skills to develop model interfaces are in demand from many fields and therefore are in short supply. Additionally, some large organizations may be required to hire staff from rigidly defined technical background or degrees, which can limit their agility in hiring computer scientists or software engineers.

3.1.2.4 Lack of team learning

Team learning, in the context of integrated modeling, is defined as a process where participants across different disciplines work together to implement an analysis and learn about each others' disciplines and mental models along the way. Systems thinking is another term that has been used in the literature to identify this process. Systems thinking is a conceptual framework – a body of knowledge and tools that can be developed to help make patterns clear (Senge, 1990). Where opportunities for team learning exist, there is a greater potential for generating acceptable solutions that address multi-faceted problems. Time and resource constraints, institutional structures, and even different terminologies can limit effective team learning in many problem-solving settings.

Pollino et al. (2017) described case-study examples that illustrate the benefit of engaging team learning. For example, in a case example of understanding groundwater-surface water interactions, an integrated model was developed collaboratively by a team that included expertise in hydrology, ecology, economics, governance, and social sciences. The inclusive and integrative approach to model development was time consuming; however, it provided significant value and greater insight into the problem and a platform for discussion among stakeholder groups.

3.1.2.5 Lack of a global expert or champion

When model integration occurs across different disciplines, there is rarely (if ever) a single expert to help interpret the global results and implications to a broader audience. Such expertise is needed in the communication of complex model outcomes in settings with stakeholder participation and is often accommodated through discipline-specific expert panels. This lack of global expertise is also encountered in interdisciplinary research work, but the work products are often reported in formal settings, such as research papers, and do not require the same level of stakeholder interaction.

3.1.3 Stable Funding Sources

Development of integrated models that serve a broad user community require resources beyond those needed for development of domain-specific models, as illustrated by the following example: Modelers in a particular domain, such as hydrology, may have the necessary background and training to develop and implement hydrologic models relatively efficiently. However, integration of a hydrologic model with a fish-behavior model needs more time and interaction with other models to develop a novel integrated framework. Furthermore, this integrated framework must mature over time to be credible and useful in a real-world setting. A stable and sustained funding effort can enable this process to occur and can provide the incentive for modeling teams to embark on such an exercise. In many situations, even when funding has been set aside for modeling studies, high-priority or short-term tasks can often exhaust modeling resources, and thus limit the development of longer-term and more novel approaches for integration.

3.1.4 Stakeholder Engagement and Trust

Stakeholders and other participants play a vital role in establishing the credibility of models and in supporting their use in important decisions. With the broader use of modeling in support of environmental decision-making, the key role of stakeholder engagement has been emphasized in several published modeling guidelines. Stakeholders should be involved in various phases of modeling, from inception to the final evaluation of results. Thus, constituents should be involved in the development of conceptual models, and ideally, should be able to investigate models independently. This participation in the modeling process may include, for example, the ability to develop independent model results with alternative scenarios that are a greater interest to the stakeholder constituency. This is especially true for decisions that involve the regulated and regulatory community where both parties should be able to independently run and evaluate models. An approach for stakeholder involvement is described in Memo 4 (Modeling Best Practices). However, in the specific context of integrated modeling, some challenges arise. For example, when multiple models are involved in analysis, stakeholders are expected to develop an understanding of different models and their interactions to appropriately interpret results. This is difficult for domain experts and may be particularly challenging for a broader group of stakeholders. An integrated model, especially a new application, may not have the same level of stakeholder acceptance that a single domain model may have developed after years of use. These aspects may adversely impact stakeholder engagement. Thus, the potential challenges of stakeholder engagement must be balanced with the benefits derived from developing integrated models.

3.2 Technical Issues in Model Integration

Integrating models that represent different technical disciplines and span different spatial and temporal domains presents a major and unique set of challenges. Here, we term them technical or “model” issues. Broadly, these issues relate to model compatibility, model and data accessibility, computational complexity and uncertainty.

3.2.1 Inter-Model Compatibility

Inter-model compatibility relates to spatial and temporal aspects of model runs. Relevant issues are discussed below.

3.2.1.1 Differences in model purpose

Temporal aspects of modeling in the Delta take two forms and are related to the underlying purpose of the model:

- **Time dependent models for understanding the evolution of systems.** This is a common form of modeling dynamic systems where a model is used to study the response of natural system variables that retain a memory over a long period of time (such as groundwater levels or fish populations). Thus, long term simulations are dependent on the initial conditions of the variables of interest, the sequence of natural drivers provided to the model, and the human drivers that may change over time. In such a simulation one might predict, for example, the groundwater levels in an aquifer given natural drivers (primarily climate) and human drivers (irrigated area and volume of pumping). The time variable associated with model output is often linked to a real time with observations, and historical observations—where the natural and human drivers are known—are the basis of calibration. A model of this type may be mechanistic or statistical.
- **Level-of-development models for planning.** These models are driven by hydrologic time series that represent the natural year-to-year variability of California climate but assume fixed conditions on the ground (e.g., land cover, regulations, water withdrawals, etc.) that represent specific time frames. In Delta modeling applications, system operations models such as CalSim take this form. The goal of such modeling is to understand the response of a defined system—including infrastructure, operations, and regulations—to a range of future hydrology. Relevant outputs from such models may be a range of values, such as water deliveries at a certain location, in response to a range of hydrologic inputs. This type of modeling is conceptually appropriate when the system being modeled effectively “re-sets” itself periodically and carries minimal memory of preceding years. For long-term simulations, outputs from such models are less sensitive to initial conditions and the sequence of inputs. Furthermore, time in such models represents a statistical sampling of the hydrology and the model response. Output from such models should not be compared directly to real-time observations.

The coupling of these two model forms (time dependent and level of development) is not straightforward, and the difference in the representation of time must be understood by model participants. Thus, when output from a level-of-development model is used to drive a time-dependent model, it is important to recognize that time in the level-of-development model may not reflect changing conditions over the long term. Another constraint for future projections is that time-dependent models may have a need for more varied data that is typically available from level-of-development models, which are primarily focused on hydrologic variables.

3.2.1.2 Differences in model time step

As models become more sophisticated and computational resources increase, finer time steps may be used. These finer time steps allow the representation of more temporally detailed processes and allow for greater spatial resolution. This is often needed for specific dynamic problems. However, upstream or downstream models may not have the same temporal discretization. Examples include: (i) the linking of the CalSim hydrologic model at a monthly time step with an estuary hydrodynamic model at a daily or sub-daily time step, (ii) the linking of a monthly groundwater model with an agroeconomic model at an annual time step, or (iii) a daily flow and water quality model with a fish-response model, requiring sub-daily inputs. The timestep for a typical model input or output, in many cases, may be tied to the observed data used for formulating and calibrating the model. Linking models with different time steps can necessitate artificial refinement in datasets that is not substantiated by the model output timestep or the observed data.

3.2.1.3 Differences in calibration and validation

Each model in an integrated modeling framework may have undergone an independent process of calibration and validation through which key parameters are adjusted to get the best fits to observations (see Memo 4 for a description of this process). The calibration and validation may be based on different time periods, different scenarios, different types of conditions (average versus extremes) that may limit the ranges and conditions over which models are credible. Coupling of models with independent calibration must consider whether there is compatibility in the conditions used. Furthermore, in some situations, there may be a need to re-calibrate an integrated model to improve the fits to observed data, which can be more challenging in an integrated framework. Special attention must also be paid to the values of shared model parameters, if such parameters are present in the component models.

3.2.1.4 Differences in time frames of published models

Some models have complex data requirements such that their set up and calibration is only performed for specific periods where all the relevant input data are available. In such instances, the extension of the model to different periods is time consuming and may be difficult to accomplish. This is a challenge when such models need to be integrated with other models that may have data availability over different periods. This consideration also applies when a proprietary model has been set up and run for a specific time period but is not readily available for use over other time frames. In other instances, some models contain embedded empirical data relationships that are tied to specific periods, and the generalization of such a model to other time periods is limited.

3.2.2 Data Availability and Exchange

Data commonly refers to alphanumeric values that are stored in some form, without regard to the origin of this information. In the environmental domain, the source of the data is also relevant, and observed data collected in the field are handled differently than data that are output from a model. Environmental model development, calibration, testing, and application is closely tied to the availability of relevant observed data. In the case of model integration, outputs from one model may serve as the input to another model. Where different types of data are needed to work with individual models in an integrated modeling framework, additional challenges arise as discussed below.

3.2.2.1 Geographic data limitations

Data collected for different domains are reported at different levels of spatial and temporal detail. Some types of physical data are collected at high frequency through automated sensors, on the order of minutes. With sensors becoming cheaper and more reliable, a more spatially intensive network can be envisioned. Other data, especially biological data, require considerable manual involvement, and may be reported at less frequent time scales, such as monthly and over a limited number of locations. Finally, economic and social data may be reported at annual time scales and specified over political boundaries. When models across these domains are to be integrated, a common basis for exchanging data must be found.

3.2.2.2 Management of model output

With the growing complexity of models, notably spatially resolved numerical models, large volumes of output data are generated that must be stored and processed into forms that are suitable for interpretation or as input to downstream models. Significant computer resources are needed to manage these data, especially if they are to be retained in a format that is accessible to a broader community. Furthermore, the manipulation of these data for input to downstream models can also be a time-consuming exercise.

3.2.3 Model Accessibility

When planning to conduct an integrated modeling study, reasonable accessibility of component models should be assured. While this need not mean that source code be available in all cases, other supporting documentation should be available to describe model conditions, scenarios, input-output file structures and units, etc. Model accessibility may be limited because of ongoing development, proprietary restrictions, and cost restrictions.

3.2.3.1 Searchability of Model-Related Information

To be able to use a model framework or a specific model application, the potential community of users should be able to access a certain amount of supporting information.

Relevant supporting information for a model includes: model source code or executable files, input data files and configuration files, sample output files, observed data files for calibration, and example scenario files. While some of this information can be re-created, it is most helpful if the relevant files from prior modeling efforts are easily available from public repositories and can be re-used or modified. A typical challenge is that such information from prior, dated projects are very difficult to find, and must be re-developed.

3.2.3.2 Documentation

Models are developed and applied over years with the participation of different individuals. The basis for specific model assumptions may be lost over time if not properly documented. In the long term, this lack of documentation may lead to an inappropriate use of the model by future users or may result in future users being unable to modify the model for a new application. Lack of model documentation may also be a challenge to a wider community of modelers who are trying to understand the model inputs, outputs, data needs, units, assumptions, etc. and integrate with other components. There are numerous examples in the Delta of older models being rendered of limited value because of insufficient or missing documentation.

3.2.3.3 Model versions in flux

Many models are in continual development, or the models’ development cycles may span several years. In such instances, even where the code is in the public domain, or is intended to be in the public domain, version control issues may limit accessibility for integration with other models.

3.2.3.4 Proprietary models

Non-proprietary software is usually preferred for stakeholder-based environment modeling because it reduces potential barriers to new users and encourages open dialog. However, nonproprietary models may not be available for a particular application or, if available, may not be adequately reliable. Furthermore, proprietary models, by having a mechanism to be sustained financially and independently of a government agency or academic institution, may provide new opportunities for enhancement and long-term viability. If proprietary models are used, they should be subject to the same rigorous quality control and peer review that might be expected of non-proprietary models. This appears to be true in many Delta applications, and the use of proprietary models is reasonably common. However, from the standpoint of model integration, the inability of proprietary models to be independently run is a challenge. In most instances, such integration is expected to be performed by utilizing results from pre-existing runs (which limits the types of scenarios that may be considered) or by the developers of the proprietary model. The latter situation may occur when the model developers are not just software vendors but an integral part of the modeling team.

3.2.4 Computational Complexity

Model complexity continues to grow, matching the trend in greater computer speeds. This phenomenon is particular true for numerical models that require gridded representations of systems in space and time. Several multi-dimensional hydrodynamic models of the estuary are computationally demanding, often placing an effective constraint on the length of a hydrologic sequence that can be evaluated. Models of such complexity also require specialized user expertise to operate. Both of these factors are a challenge to the potential integration of complex models within a larger modeling framework.

3.2.5 Propagation of Uncertainty Across Models

Models are simplifications of reality, thus making them subject to various forms of uncertainty. In environmental models specifically, these sources of uncertainty include: 1) parameters, 2) structure (model conceptualization), 3) initial state variables, 4) configuration and input variables, and 5) observation data used for training and testing the model. Further, the nature of uncertainty can be categorized into epistemic uncertainty and aleatory uncertainty or stochastic uncertainty (Walker et al., 2003). Epistemic uncertainties, which stem from a lack of knowledge, 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. Climate predictions over different time scales are perhaps the most common example of aleatory uncertainty in environmental models. Often any modeling application includes both epistemic and aleatoric uncertainties.

Although uncertainty is a well-recognized problem in modeling that must be addressed in any major modeling effort, it is of particular concern in integrated models. This is because the uncertainty in the outputs of one model result in uncertain inputs to a downstream model and this uncertainty continues to accumulate over a sequence of models. Lack of accounting for uncertainties will result in biased and unreliable model results which will directly affect the decisions made based on the modeling (Beven and Binley, 1992; Refsgaard et al., 2007; Bastin et al., 2013). Specific approaches to manage and evaluate uncertainty in modeling frameworks are described in the following chapter on solutions.

3.2.6 Limitations in Model Testing

Most models must undergo a variety of tests to assure that they are credible for the problem and the range of conditions that they are intended to evaluate. Memo 4 describes a series of steps that may be undertaken, depending on the model complexity and the importance of the decision being made using the model. Usually, rigorous testing of any model will require time and analyst resources. In the context of integrated modeling, it should be assumed that the component models have each been subject to testing and are individually considered credible. Even when component models are considered credible, integration may result in new issues that need to be considered. New issues may include propagation of uncertainty (discussed above) and occurrence of new conditions and new outcomes that have not been encountered in the component models. Best practices should include testing of integrated models in a manner consistent with component model. However, the added computational complexity of integrated models and the time and expertise taken to run them, especially when done through offline coupling (see Chapter 1 for description), severely limits the extent of testing that may be performed.

3.3 Summary

Although integrated modeling across different spatial and disciplinary domains can be beneficial in addressing complex environmental problems, the added complexity of getting two or more models to work together effectively raises some practical challenges. This chapter provides a summary of these challenges, informed by the general scientific literature and the specific case studies of integration in the Delta and Central Valley described in Memo 2. These challenges are grouped into two broad categories: institutional and technical. 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. 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. Based on these practical challenges in the implementation of integrated models, we provide specific solutions in the following two chapters, focusing on institutional issues and technical issues. This chapter illustrates that model integration is not driven by modelers alone. Even when the technical challenges of integration are solvable by modeling teams, successful development of integrated models will require other participants in the modeling process, such as model sponsors and other stakeholders, to address institutional challenges.

4 Institutional Approaches to Facilitate Integrated Modeling

Given the wide-ranging challenges associated with integrated modeling discussed in Chapter 3, we identify approaches to improve the institutional framework to facilitate integration of existing and future models. Institutional approaches involve people, including organizational managers, technical staff and educators. These approaches, which were compiled based on our experience, past work in the Delta, and documented successes in other systems, provide a foundation for the integrated modeling strategy presented in Memo 5.

4.1 Institutional Commitment and Funding

For any integrated modeling effort to be successful, leadership is needed to provide motivation to participants and sustained funding support is needed to allow novel integrated model frameworks to develop. Such efforts involve some risk in that the resulting tools may not work as intended, may take too much time to develop, or may be too computationally complex to be of practical use. Even when the integration effort is not a top-down driven exercise, leadership is needed to support modelers to go beyond existing modeling practices in creating new integrated applications. Sustained funding recognizes that most integration efforts will take additional time and resources to be fully evaluated for real-world application. In most cases, these factors (leadership and commitment) are likely to be present when the institutions’ missions and the goal of the specific integrated modeling exercise are well-aligned.

4.2 Community Development

Within the context of modeling and model integration, it is helpful to think of distinct 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. Indeed, the system in the Delta can be thought of as a "federated" system (borrowing a term from the data management literature) where modelers in different domains interact with one another, and are 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.

4.2.1 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.

Delta Modeling User Group – This user group was created by and receives ongoing support from 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 (DSM2UG).
http://baydeltaoffice.water.ca.gov/modeling/deltamodeling/dsm2usersgroup.cfm

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

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/

4.2.2 Virtual community of practice

The "virtual" or online community provides a vast network of development and support for modelers in the Delta. 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. The online code repository **GitHub** has allowed for the open storage and sharing of code, methods, tools and datasets. 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 these 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.

An important challenge that must 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.

4.2.3 Communication and physical collaboratory

The complexity and breadth of modeling in the Delta has brought to light many issues shared among modelers in the region. Although groups have formed organically to support the concurrent efforts of researchers within modeling domains, another identified approach is of a more centralized and coordinated social and physical infrastructure to support modeling in the Delta. This infrastructure, referred to as a collaboratory, provides a physical location for individuals to work together on a focused set of problems, along with the necessary virtual infrastructure to host related scientific information.

Two collaboratories have been established across California, including the Southern California Coastal Water Research Project (SCCWRP) and the San Francisco Estuary Institute (SFEI) (Medellin-Azuara et al, 2017). SCCWRP, founded in 1969, is a research institute that works to improve management of aquatic systems in Southern California and other regions. SCCWRP has been developing strategies, tools and technologies used broadly by the water management community. SFEI, founded in 1986 under a different name (Aquatic Habitat Institute), is similar to SCCWRP with a focus on San Francisco Bay. SFEI staff collect data and develop models and solutions for managing the Bay's aquatic resources. SFEI scientists are also involved in projects across the Bay watersheds, including the Delta. Both SCCWRP and SFEI are financially supported by the local wastewater discharger community as well as project-specific funds.

Integrated working groups have proven effective in other complex estuary management regions in the U.S. such as Chesapeake Bay, the Great Lakes, Long Island Sound, and the Gulf of Mexico. These groups, several of which have been active for decades, are generally funded through federal or state agencies. These working groups support monitoring, modeling, project implementation, and stakeholder engagement, i.e., modeling is not the sole focus of the collaboration – activities are summarized in Box 1 through 4. The Chesapeake Bay Program (CBP) has special relevance to the integrated modeling focus is this work. The CBP modeling workgroup has successfully united a diverse modeling community and developed a portfolio of models for the Bay, including a comprehensive watershed model and a physical space for collaboration and model development with US EPA (see Box 1).

Box 1. Chesapeake Bay Program

Chesapeake Bay was the first estuary in the nation targeted by Congress for restoration and protection. Following an initial research effort that identified excessive nutrients as the main driver of ecosystem impairment, the Chesapeake Bay Program was formed in the late 1970s with an agreement among the states bordering the Bay (represented by the governors of Maryland, Pennsylvania and Virginia, and the mayor of the District of Columbia), the U.S. EPA, and the Chesapeake Bay Commission. Delaware, New York and West Virginia joined the program in 2000. The program has an office in Annapolis, Maryland, and is staffed by employees from federal and state agencies, non-profit organizations and academic institutions (https://www.chesapeakebay.net/).

In 2010, the EPA established the Chesapeake Bay Total Maximum Daily Load (TMDL). The TMDL sets limits on the amount of nutrients and sediment that can enter the Bay and its tidal rivers to meet water quality goals from a multitude of point and non-point sources. Each of the member states has created a Watershed Implementation Plan to meet these pollution reduction goals by 2025.

Chesapeake Bay restoration is one of the largest such programs in the nation, and it is estimated that approximately $2 billion was spent on restoration in fiscal year 2017 from federal and state sources (https://chesapeakeprogress.com/funding).

Because of the variety and diffuse nature of nutrient sources to Chesapeake Bay, from land and from atmospheric deposition, modeling is an essential part of the process for quantifying loads. Modeling is also needed to evaluate biological impacts. The Chesapeake Bay Program Integrated Models workgroup consists of models for the airshed, watershed, estuary, key biota, and climate change. These integrated models assess effects of watershed management efforts on changes in nutrient and sediment loads delivered to the Bay, and the corresponding effects on Chesapeake Bay water quality and living resources.

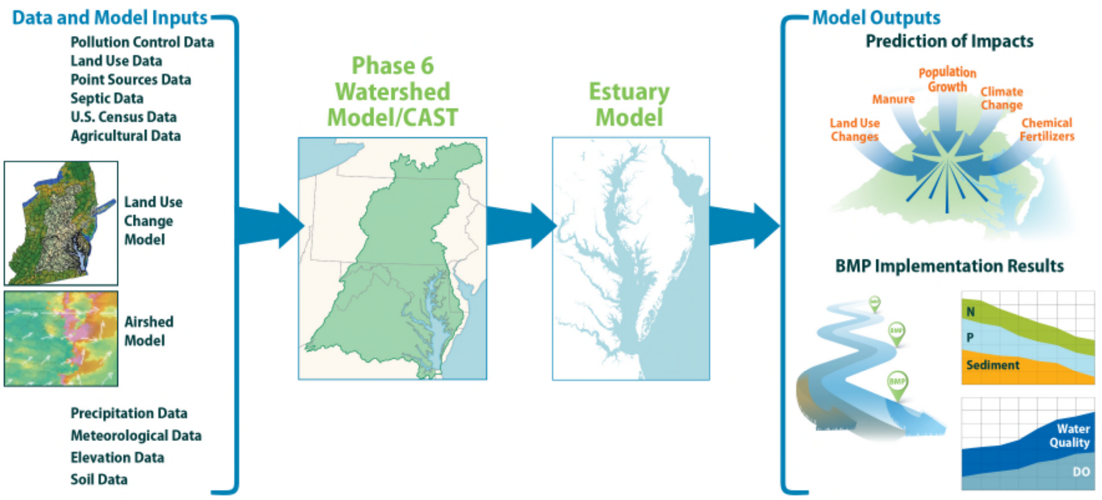

Figure Box 1. Models integrated for Chesapeake Bay nutrient loading (Source: Chesapeake Bay Program).

Box 2. Long Island Sound Study

Water quality in Long Island Sound has improved since the 1970s through a focus on point-source pollution. To continue the improvements, a cooperative effort – the Long Island Sound Study (http://longislandsoundstudy.net/) – was formed in 1985 by US EPA and the states of New York and Connecticut to focus on overall ecosystem health. The Study is a bi-state partnership consisting of federal and state agencies, user groups, concerned organizations, and individuals. In 1994, the Study developed a Comprehensive Conservation and Management Plan to protect and restore Long Island Sound (revised in 2015). The EPA Long Island Sound office is located in Stamford, Connecticut, with partners working from different locations in the Sound watershed. The Study supports and coordinates a variety of projects related to water quality and ecosystem monitoring, sea floor mapping, and modeling related to water quality, tidal marsh inundation, and climate change.

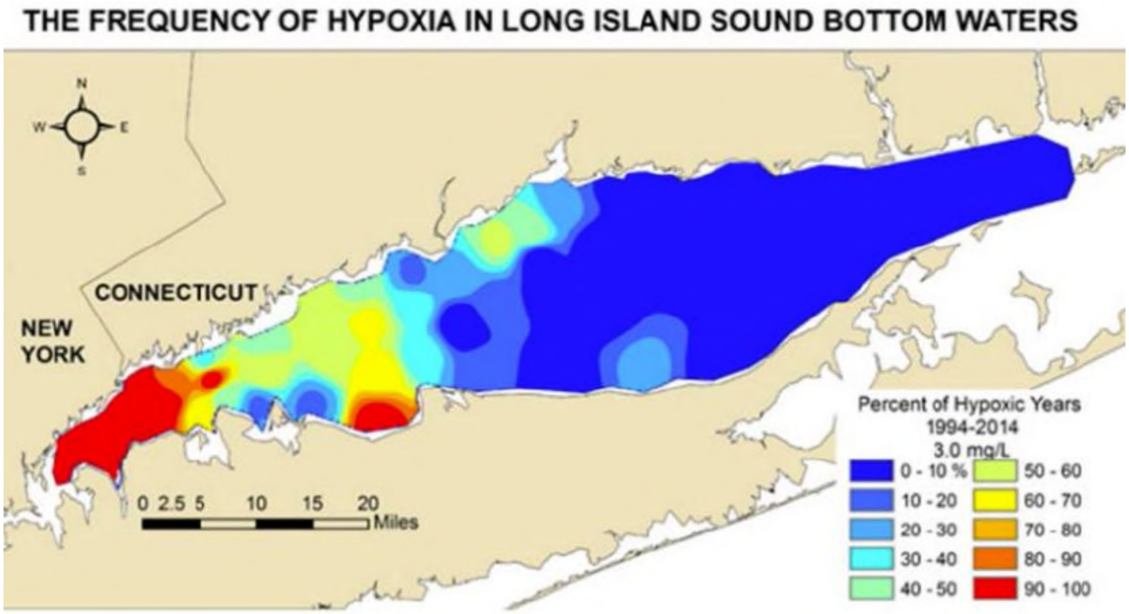

Figure Box 2. Hypoxia in Long Island Sound.

Box 3. The Center for the Integrated Modeling and Analysis of the Gulf Ecosystem (C-IMAGE)

The Center for Integrated Modeling and Analysis of Gulf Ecosystems (C-IMAGE) is a research consortium of 19 U.S. and international partners studying the effects of oil spills on marine environments (https://www.marine.usf.edu/c-image/). The C-IMAGE consortium received funding from the Gulf of Mexico Research Initiative (GoMRI) in response to the Deepwater Horizon blowout of 2010. Funding was initiated in 2011 and is now in its final phase. The center is housed in the University of South Florida.

The C-IMAGE research goal is to advance understanding of the processes and mechanisms involved in marine blowouts and their environmental consequences, ensuring that society is better-prepared to mitigate future events. Research has focused on the chemical and biological processes related to two major oil spills, the Deepwater Horizon event of 2010 and a spill of similar magnitude in the Bay of Campeche in 1979. The research includes chemical evolution and biological degradation of the petroleum/dispersant systems and subsequent interaction with coastal, open-ocean, and deep-water ecosystems, and the environmental effects of the petroleum/dispersant system on the sea floor, water column, coastal waters, beach sediments, wetlands, marshes, and organisms.

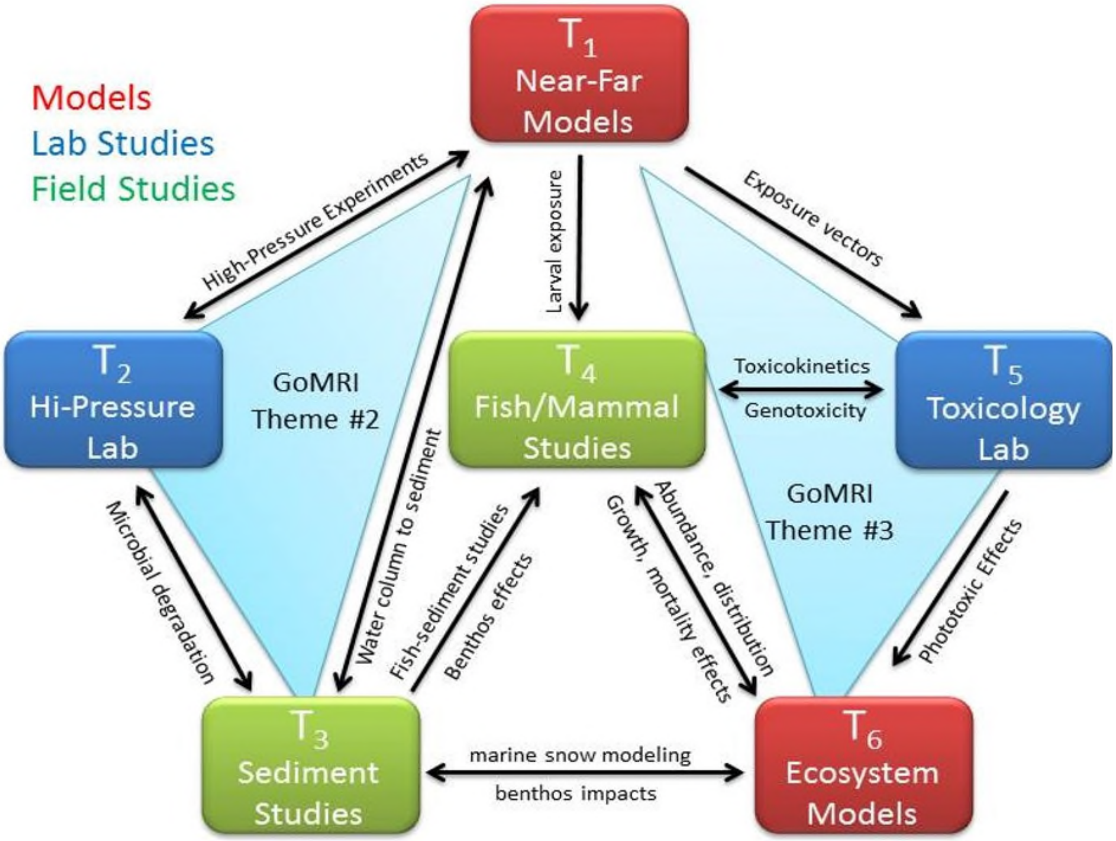

Figure Box 3. Modeling and data collection in lab and field studies as part of the C-IMAGE research program.

Box 4. Great Lakes Restoration Initiative

The Great Lakes Restoration Initiative (https://www.glri.us/) is focused on protection and restoration of what is the largest system of fresh surface water in the world. The Great Lakes contain 20 percent of the world's fresh surface water and span more than 750 miles west to east, with a 10,000-mile coastline.

The Great Lakes National Program Office (GLNPO), established in 1978, was the first EPA office with ecological rather than political or media boundaries. Its mission is to restore and maintain the chemical, physical, and biological integrity of the waters of the Great Lakes Basin ecosystem. Past and present problems in the ecosystem include excessive nutrient input (algal blooms, nuisance algae, hypoxia in Lake Erie and in large bays); bioaccumulative toxics, and invasive species. GLNPO programs include monitoring, Lakewide Management Plans (LaMPs), Areas of Concern, the Great Lakes Binational Toxics Strategy (GLBTS), the Great Lakes Legacy Act (to reduce contaminated sediments), the Cooperative Science and Monitoring Initiative (with Canada), and large-scale modeling programs. Work is coordinated through five-year action plans (the current plan is the third such plan) with a focus on the following areas: toxic substances and areas of concern; invasive species; nonpoint source pollution impacts on nearshore health; habitats and species and foundations for future restoration actions.

In addition to the EPA program office, monitoring and research functions related to water quality and ecosystem are independently performed by two other groups. The first is the National Oceanic and Atmospheric Administration (NOAA) Great Lakes Environmental Research Lab (https://noaaglerl.blog/) located in Ann Arbor, Michigan. The second is the U.S. Geological Survey's Great Lakes Science Center (https://www.usgs.gov/centers/glsc), also based in Ann Arbor, that has focused on biological research in the Great Lakes for more than 50 years.

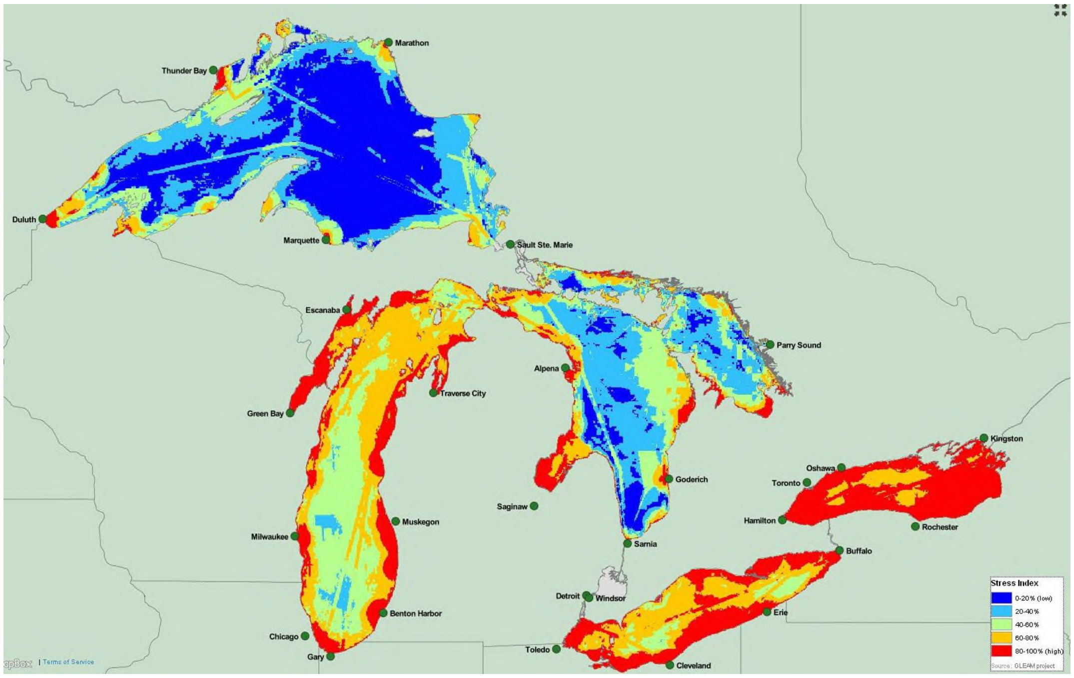

Figure Box 4. Stress in the Great Lakes mapped using 34 indicators (Source: University of Michigan).

In response to the growing interest for collaboration mechanisms in the Delta, UC Davis Center for Watershed Sciences hosted a National Science Foundation/Delta Science Council co-funded workshop intent on gaining consensus within the multi-disciplinary modeling community on how best to enhance the modeling efforts across the Delta (Goodwin et al, 2015). The primary recommendation resulting from this workshop was the formation of a Delta Modeling Collaboratory (DMC). Growing from the foundations of CWEMF, the proposed DMC would offer both an expansion to and an enhance of the existing virtual network. More importantly, the proposed DMC would provide the infrastructure required to advance the quality and role of Delta models, learning and collaboration.

The following vision of the DMC was laid out in Medellin-Azuara et al. (2017) although the future implementation may differ on specifics. The DMC's proposed primary role would be as a technical support center for modeling in the Delta (Medellin-Azuara et al, 2017). Through an association of university, agency, NGO, and private sector players, the DMC would provide support in several avenues:

- Centralized physical space intended for meetings, work, education, and in-house computational infrastructure.
- Technical staff and working teams focused on addressing existing and developing issues within and across model domains. As a virtual and physical forum on models, these teams would provide capacity for both the continual evaluation and updating of existing models, but also individualized technical support when possible.
- Educational resources for all levels of modeler as well as project managers and decision makers. Focusing on project-based learning (Thomas, 2000), the DMC would be providing a learning environment focused on applied problem-solving for current research project and management needs.

Although the concept had been proposed following the 2015 workshop, limitations in physical infrastructure and funding have not yet enabled the development of the DMC.

4.2.4 Regional forums

The Delta Science Program's **Integrated Modeling Steering Committee** (IMSC) was recently established to provide guidance and strategy for integrated modeling efforts in the Delta. The IMSC, along with the DSC, is the driver of the current study. As a consortium of agency representatives, researchers, and consultants, the IMSC builds on the increased interest in collaboration and Delta-specific solutions. As of 2018, the IMSC held monthly meetings to address key ecosystem modeling concerns and projects. As part of the updated Delta Science Plan (revised October 2018), the IMSC is intended as a stepping stone to a greater regional modeling community. The Delta Science Program has also provided an annual regional forum for Delta research through the Bay-Delta Science Conference, as has the San Francisco Estuary Partnership via the State of the Estuary Conference.. Both forums have created a venue for sharing and obtaining feedback on modeling for the Delta's agencies, researchers, and consultants.

The **Interagency Ecological Program** (IEP) has facilitated interagency research in the Bay and Delta since the 1970's. With a strong focus on fisheries, the program has been a critical collaborative organization for the development of monitoring and models. The organization has emphasized science, synthesis, and service in its operations and been an important component of stakeholder outreach in the region. IEP hosts an annual workshop, publishes quarterly newsletters, and provide strategic documents for ecological programming in the Bay/Delta area.

4.2.5 State/national forums

California Water and Environmental Modeling Forum (CWEMF) has been an integral component of California's modeling community since 1994. The mandate of the non-profit has been to increase usefulness of models, pool and distribute technical information, mediate technical disputes, and provide impartial peer reviews of models for the community. CWEMF hosts an annual meeting to allow for a physical forum on the state of modeling in California. These three-day events provide an opportunity for modelers to exchange ideas, highlight new approaches, and receive updates from model developers. In addition to the annual meeting, CWEMF hosts model-specific workshops and training opportunities.

Professional associations also provide forums on a state and national basis. The **Groundwater Resource Association of California** (GRA) is a network of managers and researchers focused on the topic of groundwater in the state. The GRA hosts numerous conferences and summits throughout the year, including technically-focused interdisciplinary forums. At a national level, **The American Society for Civil Engineers** (ASCE) includes several California chapters and a specialized **Environmental & Water Resource Institute** (EWRI), which hosts targeted forums, publications, and workshops. Similarly, the **American Water Resource Association** (AWRA) hosts several conferences and workshops each year throughout the U.S. and publishes the widely-read Journal of the American Water Resources Association. The National Science Foundation's **Consortium of Universities for the Advancement of Hydrologic Science** (CUASHI) serves as both an online and physical forum for national water resource scientists and holds a variety of annual events. The **International Association for Great Lakes Research** (IAGLR) hosts an annual conference on Great Lakes Research that provides an opportunity for workshops and sessions on cross-cutting modeling and analysis tools. The **American Geophysical Union** (AGU) hosts a major conference each year that allows for specialized groups from across the nation to gather in focused sessions, including significant representation from modelers. At a broad scale, these organizations offer forums for modelers to exchange knowledge and approaches to addressing regional issues. The creation of additional regional, state, and national forums which assemble users in a particular model domain would provide an ideal venue for providing domain-specific support, technical training, and workshops.

4.3 Education

Currently, the training of modelers is fragmented and dependent on existing expertise within an organization (Medellin-Azuara et al, 2017). Synergy between universities, public agencies, and private consulting is highly dependent on the experts or organizations involved and could be enhanced by increasing communication between domain experts (within public agencies and private consulting firms) and training institutions. Active feedback to university curriculums by domain experts, engagement of students through internships, and targeted workshops for novice modelers could improve training outcomes for the next cohort of modelers in the region.

4.3.1 Staff

Training is often provided within organizations that utilize a specific model or by modeling/professional associations. Many of the regional, state, and national organizations previously discussed offer model-specific workshops. These workshops provide time for direct interface with model experts, other domain modelers, and experts within the participant's own organization. With expert retirements affecting model use within organizations (Medellin-Azuara et al, 2017), these workshops fill an important training role. Organizations like CWEMF often offer modeling training workshops with regional and statewide applications often related to the Delta.

4.3.2 Students

The University of California (UC) and California State University (CSU) offer programs with foundations in modeling. Many programs offer opportunities for students to interact and participate in internships with regional, state, and federal agencies tasked with water resource management (e.g. USACE). Graduate programs often provide opportunities for mentorship with academic modelers and provide a training ground for many of the modelers in both public agencies and private consulting. However, with many models developed within agencies and consulting companies, graduates with strong modeling foundations often learn modeling within the workplace. Incorporating the following into course curriculum could enhance training in the university setting: cross disciplinary courses; introduction to the technical challenges of integrated modeling; and computer science training in integrated code development.

4.3.3 Stakeholders

Stakeholder education through training and workshops is a necessary part of the overall education framework. In the Delta, such education has been performed successfully and cost-effectively by CWEMF, the Water Education Foundation and by other state agencies.

4.4 Summary

This chapter focuses on model integration challenges that arise around participating organizations and people. We identify different actions that can help stimulate the development of integrated models, including institutional commitment and leadership support, model community development, and education. Modeling communities can take the form of user groups (many of which are already in existence), a virtual community of practice, or a physical location for interested participants to work together (i.e. collaboratory). Community engagement across participating agencies is also fostered by various regional, state, and national forums that involve exchange among modelers and scientists. Institutional efforts for model integration also include education for current and future students, staff in participating organizations, as well as the broader stakeholder community. These institutional challenges, while distinct from technical challenges, are equally important to address for the long term success of model integration in the Delta.

5 Technological Approaches to Facilitate Integrated Modeling

In this chapter we propose guidance and identify approaches that will enhance the ability of modelers to integrate existing and future models. Based on our review of present-day integrated modeling efforts, some of these proposals are already in use and are noted below. However, others are not, and this chapter is intended to serve as a point of reference for technological solutions to the integration of models. This chapter focuses on the challenges of model integration specifically; guidance to improve the robustness of modeling in general is presented in Memo 4, *Modeling Best Practices*. The approaches are discussed along the following themes: documentation and nomenclature standards, interfaces, uncertainty propagation, model data exchange standards, model emulation within integration frameworks, and big data analysis approaches.

5.1 Documentation and Nomenclature Standards

5.1.1 Documentation for model data exchange

Meta-data standards are needed for individual model inputs and outputs, similar to standards set for observed data. These metadata should include brief descriptions of the type, temporal and spatial scale, and units of the input/output data. Although such information is often available within model documentation, it may not be necessarily transparent to non-specialist users of the model. Specifically, implications of inconsistencies between the input/output data for different components of the integrated model and required data modifications should be explicitly communicated. Making this information clearly available will minimize the potential misuse of a model where it is being integrated to represent different processes.

5.1.2 Documentation for model processes

Documentation standards for model processes apply at two levels: first for general purpose model frameworks and second for specific applications to a geographic area. In some cases, the two are conflated, where a single model is developed for a specific geographic area.

Documentation for model frameworks needs to provide enough theoretical background on the processes being represented. Usually, presenting the general conceptual model is valuable for effectively communicating the processes modeled. Moreover, basic information on transformation of the conceptual model to a mathematical model, and then a potential computer model, should be presented clearly. For any model that is being used to support an environmental decision affecting the Delta ecosystem and communities, this level of information should be in the public domain, and there should be no mystery as to the processes being represented. In many cases, a typical user will not need to drill down to the implementation level of a model in the form of computer code, and this information need not be in the public domain. However, such documentation is needed to allow model maintenance, support, and improvements over an extended period of time where many different people may be contributing to the changes. Rigorous documentation is also important where a single individual or a small team is responsible for development and support of the model, when due to transitions and staff and retirements, there is an increased risk of loss of model background information.

For specific model applications (e.g. those applied to particular geographic areas or those applied for fundamental scientific understanding), there is an additional need to clearly document application-specific characteristics and methods such as: site conditions, model setup decisions, calibration and evaluation approach, and model uncertainty. These needs are outlined in more detail in Memo 4, *Best Modeling Practices*. It is important to point out that good practices in the development of individual models directly translate to greater ease of integration across models when the need arises.

5.1.3 Common nomenclature

Consistent terminology for similar model processes and data, including reporting units, is desirable to promote effective communication and minimize loss of information across models. It is unrealistic to expect that all models will use the same terminology, but it is reasonable to require that terminology be fully defined before it is used in each case, listing relevant information such as: linkage to specific theoretical or empirical framework, form of measurement, temporal and spatial frequency, etc. An effort to compile terminology for individual models will greatly facilitate communication among models and modelers. The wiki model inventory (see Memo 1 *Model Inventory*) is a good virtual location where key and commonly used terms can be defined with typical units and general context.

5.2 Interfaces

5.2.1 Front End Interfaces

While interfaces are not a central part of model integration, improved and accessible user interfaces almost always reduce barriers to entry and encourage broader adoption of models. Few public domain technical models in use in the Delta today have reasonably intuitive user interfaces, with the exception of tools being developed for support of the Sustainable Groundwater Management Act (SGMA) and selected general purpose tools developed by the U.S. Army Corps of Engineers for hydrology, hydraulics, and flood risk evaluation (see Memo 1 for overview). It is recognized that the development of user interfaces is a resource intensive task. However, there is a payback in wider use and greater stakeholder involvement over the long term.

5.2.2 Post-Processing Tools

Tools that efficiently and intuitively present model results greatly benefit the process of model development and testing and the utility to model users. As with front-end user interfaces, these back-end tools take resources to develop and are not central to model integration; however, they enable integration by encouraging adoption across a broader community.

5.3 Model Data Exchange Standards

Component models considered for integration, in most cases, will have independent histories, having been developed independently based on the subject area needs. In additional to the domain information embedded in each model, we expect the component models to differ in the following areas:

- spatial and temporal computational resolution,
- input data,
- programming language and development environment,
- units and assumptions,
- output results, and
- user interfaces for inputs and display or results.

The task of integrated modeling entails linking models with the above differences together into an operational model chain. In the simple case of manual exchange, the models are run separately in sequence and the outputs from one model is parsed to the next (see Case 1 in Figure 1). Upon completion of an upstream model run, outputs, following transformation if needed, serve as inputs to a downstream model. The component models are not modified and are run in a standalone manner. Typically, this sequence of runs is not iterative, i.e., one set of upstream model results are fed to downstream model (more than two models may be involved, but the same concept applies). As noted in Memo 2, *A Survey of Recent Integrated Modeling Applications in the Delta and Central Valley*, this is a common approach for many projects in the Delta today.

The more general requirement is of a fully integrated network with loops and feedbacks, where models or even modules within models pass data to each other dynamically. These integration approaches require interoperability to be addressed at technical, semantic, and dataset levels (Belete et al., 2017). Interoperable frameworks are a major focus of research in the field of integrated modeling. Key concepts in code development for data exchange in integrated environmental models and common data exchange platforms are discussed in further detail below.

5.3.1 Code Development for Data Exchange in Integrated Models

The approach of running models in sequence is adequate when there are no feedbacks between processes and where the decision-making process does not require an optimal analysis of multiple processes simultaneously. In some integrated modeling efforts, there may be a need for feedback among processes, such as between a reservoir-operations model and models of specific responses downstream of the reservoir, such that the operations can be modified to meet conditions downstream. Where dynamic feedback between components is an essential part of the conceptual model, some form of process integration at the code level is required.

A large portion of code development efforts for integrated models is concerned with data exchange and data manipulation which are fundamental to integrated modeling systems (Argent, 2004; Leimbach and Jaeger, 2005). These efforts are often constrained by technical and conceptual challenges. Technically, individual models are designed to serve as stand-alone components that serve unique purposes and goals. Conceptual challenges include resolving the different ways modelers and science domains represent data and knowledge. Inter-operability of models can also be provided at different levels. At the technical level, models should be able to ‘talk to each other’ which requires automating data exchange, making models jointly executable, and ensuring repeatability and reproducibility of model chain configuration and processing (Knapen et al., 2013). At the semantic level, models should 'understand each other' by identifying and, if possible, bridging semantic differences in an automated manner. After semantic reconciliation, the datasets should be compatible between the models. This often entails unit and format conversion, aggregation or disaggregation, interpolation, etc. to prepare the data for exchange between model.

Code development for integrated models often involves the following steps: (i) modification of individual model modules (subroutines) usually with the purpose of adding intermediary arrays to store variables that need to be passed between the models, (ii) creating new modules or subroutines that control the communication between the individual models at each integrated model iteration, and (iii) creating modules or subroutines for processing and writing the output data from the integrated model. This approach gets more complicated when the models to be integrated are developed in different programming languages. Under these circumstances, extensive programming skills are required to reconcile the inconsistencies that stem from different programming languages for the models. However, this level of programming skill may not be available in all cases. Another approach that can facilitate model integration is application of platforms that are developed for standardizing the data exchange between various models, as described below.

5.3.2 Platforms for Data Exchange in Integrated Environmental Models

Several model integration platforms (or frameworks) allow creation of dynamic feedbacks through a plug-and-play mechanism by connecting submodels and components of various models. The term plug-and-play in this context, does not necessarily mean a fully ready to use platform that various models can be plugged into and operate together. Rather, it implies that the level of effort to develop interactions and feedbacks between various components of the integrated model is reduced due to automation and standardization of data exchange procedures. There may still be a need for performing code revisions inside the individual models when these platforms are used, but the effort is considerably less than when models are integrated by only revising the model codes as discussed in Section 5.3.1. Intuitively, the ideal situation is to have a platform that is completely model-independent and can be applied to a wide range of models with different capabilities with minimal level of effort for modification.

Table 5 lists key features of commonly used integration platforms. This document provides an overview of these platforms, and the potential benefits of such an approach, although the specifics of implementation of each is beyond the scope of this work. Jagers (2010) summarizes the main differences between these platforms, and notes that there is wide variety of alternative solutions due to conflicting priorities (e.g., performance, ease of use and generality). It is important to recognize that when deciding on a platform for integrating existing models, there is often a tradeoff between convenience and reusability. For example, the effort required to standardize the interface of a legacy code for one of the platforms below can be substantial, but the resulting usability of the model can be greatly increased, since it may then be easily wrapped and combined with other models. A summary of the available platforms for model integration is provided below. The frameworks discussed range from model-independent frameworks such as the Open Model Initiative to discipline-specific frameworks such as Earth System Modelling Framework (ESMF).

- Initiatives such as Open Model Initiative (OpenMI) in hydrology (Blind et al., 2005) focus on standardizing the interface to give a clear vision of requirements and limitations. Recent adaptations to the OpenMI standard (Buahin and Horsburgh, 2018) pay particular attention to these common issues of interoperability: for example, allowing more abstract inputs and outputs, and permitting inputs which have no specific time frame, thus opening up the tools for use with non-time stepping models.
- The Community Surface Dynamics Modeling System (CSDMS), focused on earth system models, is a platform that employs state-of-the-art architectures, interface standards and frameworks that make it possible to convert stand-alone models into flexible "plug-and-play" components that can be assembled into larger applications (Peckham et al., 2013). The CSDMS model-coupling environment offers language interoperability, structured and unstructured grids, and serves as a migration pathway for surface dynamics modelers towards High Performance Computing.
- The Open Geospatial Consortium Web Processing Service (OGS WPS) Interface Standard provides rules for standardizing how inputs and outputs (requests and responses) for geospatial processing services, such as polygon overlay. The standard also defines how a client can request the execution of a process, and how the

output from the process is handled. It defines an interface that facilitates the publishing of geospatial processes and clients' discovery of and binding to those processes. The data required by the WPS can be delivered across a network or they can be available at the server.

- General purpose workflow tools such as Taverna, Kepler, Vis Trails, and Trident provide user-friendly GUIs within which modular processing or data entities can be arranged, inputs mapped to outputs and control/break conditions defined (Table 5). The resulting workflow chains can be stored, published, shared and exposed as encapsulated models, while the component models themselves must simply expose a document describing each process, and its inputs and outputs. Thus, these tools can be used as engines for interacting with other workflows, as well as compiled C code or R scripts.
- Finally, there are also a host of discipline-specific frameworks (Table 5) for combining models and controlling their execution, such as FRAMES (Framework for Risk Analysis of Multimedia Environmental Systems, Laniak et al., 2013), SME (Spatial Modelling Environment, Maxwell and Constanza, 1997), TIME (The Invisible Modelling Environment, Rahman et al., 2003), MCT (Model Coupling Toolkit, Larson et al., 2001), and ESMF (Earth System Modelling Framework, Collins et al., 2005). Many of these frameworks include standard modules for applications such as hydrological or climate modelling.

There is no existing one-size-fits-all platform for the wide range of integrated modeling potentially needed in the Delta, spanning natural and social science disciplines. Moving forward, however, a platform can be modified from the available options above for future use, and require participants to develop future models or modify existing models that meet a public specification. Borrowing from the experience of past efforts, a new system can be developed that is modular, extensible, and adaptable over the long term. For example, using web pages or web services, the system would provide an open, extensible dictionary of field types, a model registration service, unit conversion service, and data transfer capabilities—all pertaining to the Delta region. An extensible model file format would be defined to provide both general and model specific fields. When large amounts of data are produced, the model information would provide a database connect string and select statements. Models located locally or remotely could be chained together for coordinated execution – remote models launched using web services. These actions require specialized programming expertise to set up, and there are upfront costs for adoption. In the long run, however, they provide extensive opportunities for integrated modeling, and may be considered as the needs for such modeling mature in the Delta.

Table 5. A list of model integration technologies

Approach	Language(s) and service interfaces	Description	Key features	Reference
Open Model Initiative (Open MI)	C# or Java interfaces, wrapped C/Fortran	A collection of programming interfaces for components.	It consists of a list of function names or method signatures called initialize, update, input items, status, etc. that enable the model being wrapped to request data from other models and respond to requests for data.	Moore and Tindall (2005)
Community surface Dynamics Modeling System (CSDMS)	C, C++, Fortran, Java, or Python	Developed to simplify conversion of an existing model to a reusable, plug-and-play model component.	It has two levels of specification: (1) Basic Model Interface (BMI) developed to provide model metadata to the next level and (2) Component Model Interface (CMI) communicates with BMI functions as well as with Service Components and the CSDMS Framework	https://csdms.colorado.edu
Open Geospatial Consortium Web Processing Service (OGC WPS)	NA	Specifications on how inputs and outputs of geospatial services are handled.	It has three mandatory operations that should be implemented by all services: Get Capabilities, Describe Process, and Execute.	Schut and Whiteside (2007)
Kepler, Taverna, Vis Trails, and Trident	Java, PMML, WSDL, BPEL, wrapped C/Fortran, Python	Workflow tools providing user-friendly GUIs that can be used to process and arrange data entries, mapping inputs to outputs, and defining control/break conditions	The workflow can be stored, published, shared, and exposed as encapsulated models, while the component models themselves must simply expose a WSDL document describing each process, and its input and outputs.	More details in Bastin et al., 2013
FRAMES, TIME, SME, MCT, ESMF	Native C interface with bindings for Java, .NET, Fortran, VB6 and Python, Open MI, Wrapped C/Fortran, C++	Discipline-specific frameworks for combining models and controlling their execution	They include standard modules for hydrological or climate modelling. The recent versions generate wrappers and control code wrappers for model sequences based on standardized model metadata.	More details and comparisons in Bastin et al., 2013; Whelan et al., 2014

5.4 Uncertainty Propagation Across Integrated Models

All environmental models face challenges related to uncertainties. Approaches for analyzing uncertainty are described in Memo 4, *Modeling Best Practices*. These challenges are magnified in the case of complex integrated environmental models as described in Chapter 3. Uncertainty assessment in integrated modeling consists of two stages: 1) assessing uncertainties associated with individual models, and 2) assessing propagation of uncertainties from individual models through the integrated system. The decomposition of aggregated uncertainties is challenging due to multiple models being involved.

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 distribution of model parameters is 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, a 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 multi chain MCMC algorithm 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).

A significant challenge to applying sensitivity and uncertainty analysis techniques to integrated models is the high computational burden. Nearly all sensitivity and uncertainty analysis techniques require numerous model iterations. In the case of integrated models, this issue is exacerbated as several models are working jointly in each model iteration. Facing these high computational burdens, modelers must use manual techniques that use a very limited number of model runs or resort to more advanced computational techniques as described below.

5.5 Model Emulators to Represent Complex Models

When computationally intensive models are integrated, 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 6, 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 6. 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.

Algorithm	Description
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.
Artificial Neural Networks (ANNs)	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. ANNs 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.
Deep Learning	Software and hardware innovations since the early 2010s have greatly expanded the size and complexity of ANNs that are feasible to train in a reasonable amount of time. Specialized ANN architectures with many connected layers of nodes are referred to as "deep networks" and machine learning using these networks is deep learning. The distinction in terminology comes from the fact that the depth of these networks induces qualitatively different behavior compared to traditional ANNs. They can generalize beyond the training data much better and are able to extract relevant information from raw, unprocessed data much more successfully. In general, these networks must be trained on specialized hardware. It has become commonplace to rent cloud computing resources to train these models. Selection of an appropriate deep learning architecture should be guided by the particular applications. Thus, architecture selection requires some expert knowledge, even if the final model itself can handle raw, unprocessed data. For example, deep recurrent neural networks (RNNs) can be used to estimate relationships between and predict time series variables and convolutional neural networks (CNNs) can be used to process image-like data.

5.6 Data Analysis Frameworks in Support of Model Integration

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 R programming language is the most widely used non-commercial, or open source, programming environment 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 integrated environmental modeling realm, the big data concept primarily pertains to techniques to capture, process, analyze, and visualize large structured and unstructured 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 2). The inter/multi-disciplinary nature of the integration problem necessitates the merging of large, disparate datasets (model inputs/outputs) which eventually must be analyzed to make inferences about the system being modeled.

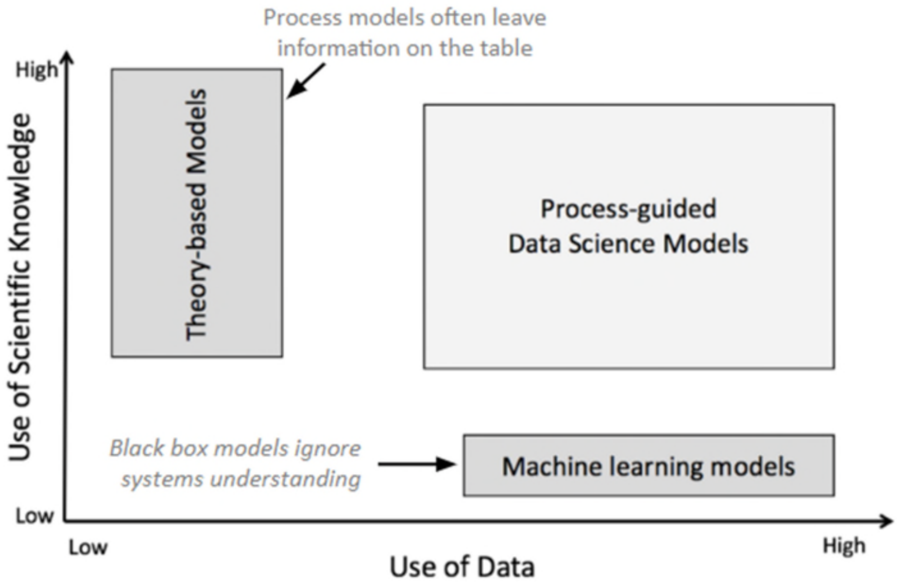

Figure 2. Big data analysis can help benefit from both black box and process-based modeling approaches. Modified from Karpatne et al. 2017.

5.7 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 section 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. Its graphical wizard generates native code. 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.
- **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.
- **R-Programming:** 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 superb visualization capabilities which is 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.
- **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.

- **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 of 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.

5.8 Summary

This chapter discusses several technological approaches to facilitate model integration:

- Model documentation is an obvious and straightforward approach; this documentation should address model structure and processes and the data being exchanged between models. Documentation minimizes the opportunities for error in translation across models, a major concern in any model integration effort.
- User interfaces, while not essential for model integration *per se*, allows greater accessibility and understanding of data input and output needs, and is therefore beneficial for cross-disciplinary interaction.
- Data exchange standards are an essential element for creating frameworks that allow models to share information among one another in various dynamic formats. Several such data exchange frameworks are in active development in the environmental domain to promote efficient and transparent inter-model communication.
- Formal evaluation of uncertainty propagation in linked models is a technological approach that can promote more informed use of model results in decision making. Such analysis can be highly computationally demanding and is currently the subject of research.
- Model emulation, an approach that replaces a complex model with a simpler approximation, reduces computational requirements. In many cases, emulators can be embedded within another model. Several emulation approaches are available, with many being used in the Delta.
- Adoption of big data approaches can facilitate integrated modeling. Related analysis tools are undergoing rapid development, especially in the commercial realm. Some environmental applications of these tools are beginning to appear, and given the potential utility of these tools for management and integrated data analysis, many future applications will likely develop. These likely developments include standalone models as well as hybrid models that combine data-based approaches with process-based models.

Overall, our review suggests that technological approaches to facilitate model integration are developing rapidly in the environmental domain and other related domains. These approaches offer many different avenues for linking models and creating new integrated modeling frameworks to support future decision-making needs.

6 Data Needs for Integrated Modeling

This chapter outlines data needs for the following model domains: hydrodynamics, ecology, water quality, fish species, water budgets and consumptive use, agricultural economics, and socio-economics. These data needs, based upon evaluation of individual models that may constitute components in an integrated modeling framework, support the construction and testing of model studies and are intended to serve as a reference for modelers from different domains. In many cases, these data are being collected across the Delta, although not always in a coordinated manner and not covering the same spatial and temporal extents. Where appropriate, we identify key gaps in the currently collected data that affects our understanding of the Delta system. In addition to the domain-specific data described below, for long term modeling over decadal time scales, there is also a need for consideration of climate data. Such data are being developed available statewide (e.g., through the Cal-Adapt website, https://cal-adapt.org/) and can provide a set of consistent scenarios for impact modeling in different domains. Climate scenario data have been described extensively elsewhere and the reader is referred to the Cal-Adapt website for more specific information.

This chapter is largely focused on the types of data currently used for developing and testing models. However, it is important to recognize that technology is driving data collection as well as accessibility over time. Thus, new sensors and communication technologies for spatially and temporally resolved data collection, LiDAR[1] and remotely sensed geospatial data, geophysical data, and citizen-sourced data are adding to our capabilities. Similarly, there are a variety of new data management tools (described in the previous chapter) that enhance data access and support the tasks of analysis and

---

[1] For Light Detection and Ranging, used for high resolution elevation mapping using puled laser beams.

interpretation. Furthermore, Delta management is being broadly influenced by population trends, climate change pressures, and societal expectations are broadly affecting Delta management, translating into new management priorities such as water security and greenhouse gas emissions. Over each of the domains identified below, there is a need to continually assess new opportunities and prioritize data requirements.

6.1 Hydrodynamics

Delta hydrodynamic and water quality transport models require a wide variety of data. Hydrodynamics models generally compute water flows, stages, and velocities in one- to three-dimensional framework. Water quality variables such as temperature and salinity in estuarine settings are also typically calculated using hydrodynamic models, whereas other water quality parameters, such as nutrients and contaminants, many of which are reactive in the environment, require additional specialized models. Data needs vary to some degree based on the dimensionality and formulation of the individual model. For purposes of this discussion, data needs are defined in terms of simulating a historically-observed condition. However, data needs are similar when simulating synthetic planning-level scenarios. Data needs are classified here into two broad groups based upon our findings in this study: time dependent data and time independent data.

6.1.1 Time Dependent Data

Time dependent data are provided to the models as time series and include boundary conditions, operation schedules, and validation data. The necessary time step for these data will vary for different hydrodynamic models. Each data type is discussed below.

6.1.1.1 Boundary conditions

Delta boundary conditions may be specified for flow, water quality (including temperature and salinity), water levels at the downstream tidal boundary, wind speed and direction, and air temperature. Flow boundary conditions are specified at riverine inflow locations, at internal diversion locations (agricultural and urban), and at internal return flow locations. Common practice employs separate models to generate agricultural diversion and return flow data, as these are not generally measured. Water quality boundary conditions are specified at riverine inflow locations, at internal return flow locations, and at the downstream tidal boundary. Water quality from return flows are not generally measured; thus, these data are typically specified by synthetic time series. Water quality at the tidal boundary may be specified as a constant, time independent value, or a time varying values for models with spatial domains that extend to the Pacific Ocean. Finally, wind and air temperature boundary conditions are specified at a regional scale.

6.1.1.2 Initial conditions

Delta initial conditions are typically generated by a “cold start”, i.e. nominally initializing the model and then “spinning up” the model until a representative set of conditions are achieved in the system. A spin-up period of six to 12 months is common. An accurate specification of initial conditions, i.e. a “warm start”, is recommended when the simulation period is shorter than the time required for these initial conditions to be “flushed out”. A short-term forecast is an example of an application when accurate initial conditions are necessary and a warm start should be used.

6.1.1.3 Operation schedules

Operating schedules or rules are required for Delta flow management facilities, including the Delta Cross Channel, the Suisun Marsh salinity control structure, and temporary barriers that are seasonally installed in south Delta channels.

6.1.1.4 Validation data

Flow, water quality and water level data are needed at internal Delta channel locations to validate historical model simulations. While a rigid data requirement does not exist, a model validation is most convincing when demonstrated over multiple years and over a wide range of hydrologic conditions.

6.1.2 Time Independent Data

Time independent data needed for Delta hydrodynamic models include geometry data and model parameters. Each is discussed below.

6.1.2.1 Geometry data

Delta hydrodynamic models require data to characterize a variety of physical features that influence flow patterns. First, the model needs data to define the broad spatial domain. This is done through a network for one-dimensional models or through a model grid for multi-dimensional models. Ideally, the model network or grid is geo-referenced. Next, hydrodynamic models need data to define physical features of the Delta channels; these data are referred to as bathymetry data. Bathymetry data are typically processed to define channel characteristics at a number of cross sections. Hydrodynamic models also need data to define land elevations, including levees. Elevation data are typically processed through digital elevation models to provide adequate spatial resolution. Finally, Delta hydrodynamic models require data to characterize locations and dimensions of channel barriers and gates.

6.1.2.2 Model parameters

Delta hydrodynamic and water quality models require a variety of constants to characterize physical processes; specific needs are unique to each model. Common model parameters include hydrodynamic and transport constants (e.g. roughness and dispersion coefficients), climatic constants (e.g. evaporation rates), and water quality rate constants (e.g. degradation rates). Model parameters are not typically determined by direct measurement. Rather, they are typically estimated through literature review or through a model calibration process.

6.2 Ecology: Food-webs and Fisheries

The Delta is a highly dynamic system with a diverse and complex food web that changes among locations and time of year. Feeding behavior varies as prey availability shifts and can depend on the size of the predator. We are not aware of a food web model that encompasses the entire Delta food. Thus, physical and biological descriptions below are based on conceptual models for the Delta and food web models from other systems.

6.2.1 Physical Data

The key physical drivers of food web structure and dynamics include residence time, turbidity, depth, stratification, and salinity. The amount of time spent in a physiographic region (i.e., residence time) is important to both phytoplankton and zooplankton. Increased residence time can be related to increases in water temperatures, biomass accumulation, and nutrient retention. Decreased residence time moves non-motile organisms and particles through the system more quickly, spreading them to other parts of the Delta. Food-web production at different trophic levels is thought to depend on nutrient recharge, biomass accumulation and advection, water temperatures, and trophic response time; these conditions are all directly related to residence time. Turbidity in the Delta reduces the photic zone, having an inhibitive effect on phytoplankton which forms the base of the food web. A reduction in the amount of primary production limits the amount of food available for higher trophic levels. Water depth similarly affects primary production because photosynthesis declines with depth. Stratification affects primary productivity by keeping phytoplankton in the photic zone. Salinity can affect the food web through its influence on location of stationary benthic organisms. For example, the invasive bivalves, C. fluminea and C. amurensis, respond along low salinity and freshwater boundaries to changes in flow regimes in complex ways that depend on life stage and other factors; salinities that might eliminate one of these species could provide hospitable habitat for the other.

6.2.2 Biological Data

Food web models link organisms by their feeding relationships. For simplicity, species are often placed in functional groups. Links among functional groups are based on diet information that are traditionally collected through scat or stomach content analysis. 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 (e.g., at the scale of the Delta) 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.

6.3 Water Quality

The direct measurement of water properties is often considered to be indicative of water quality. However, strictly speaking, water quality is defined by a beneficial use or an interacting entity or process (i.e. fishery or ecological processes) in conjunction with measured intrinsic water properties. For example, if water quality is defined by a harmful algal bloom (HAB) prevention goal, then nutrient levels would be a relevant water quality property. Below we discuss key water properties as defined by important processes and beneficial uses for the Delta, the root causes for gaps within water quality data, current and future efforts to address these gaps, data needs for water quality modeling and some examples of Delta specific implications.

6.3.1 Important Water Properties by Beneficial Use and Ecological Process in the Delta

Some important Delta ecological processes and beneficial uses, along with associated water properties, are presented in Table 7. Nutrients, salinity and pesticides are the most common Delta water quality constituents affecting three to five of the ecological processes and human beneficial uses listed in Table 7). Nutrients are a nuisance to water delivery and recreation in the Delta, as they often lead to the formation of algal blooms (Mioni, 2012). HABs can contribute cyanotoxins that negatively impact drinking water supply, aquatic species and water-based recreation. Some nutrients, such as ammonia, are toxic to fish at sufficiently high concentrations. High salinity concentrations negatively affect urban and agricultural water supplies (Medellín-Azuara et al., 2014), two very important beneficial uses for Delta waters. Anthropogenic influences on salinity levels (e.g. agricultural practices and upstream reservoir operation) promote reversed spatial and dampened seasonal salinity gradients in the Delta (Hutton et al., 2015); these salinity changes have been shown to change the dynamics of aquatic food-webs, fish spawning and populations (Mount et al., 2012). Organophosphorus, organochlorine and pyrethroid-based pesticides influence aquatic species toxicity and bioaccumulation and impact drinking water quality. Other water properties, such as mercury (California Environmental Protection Agency, 2019), selenium, organic carbon and disinfection byproducts are also important considerations for toxicity, bioaccumulation and drinking water supply. Among the processes and uses identified in Table 7, drinking water is the most heavily affected beneficial use, with nearly every listed water property having some sort of regulatory standard (U.S. EPA, 2018).

Table 7 is certainly not a complete or static list. Many water properties can potentially hinder water use and impact ecological processes, and we are continually discovering new chemicals that have significant impacts. Chemicals of emerging concern (CECs) is a transitory category; these substances are often unregulated or have only recently been detected due to advances in technology. This category currently includes pharmaceuticals, personal care products, flame retardants, emerging pesticides, and commercial and industrial chemicals (California Water Boards, 2017; Weston and Lydy, 2010). Endocrine disruptors, another group of CECs, cause unnatural shifts in gender distributions among aquatic species. New beneficial uses will continue to be defined and important ecological processes will continue to be uncovered as scientific discovery and societal needs evolve over time (California Regional Water Quality Control Board, 2018).

Table 7. Water quality parameters: Important water properties by process and beneficial use in the Sacramento-San Joaquin Delta Example integrated modeling needs, now and in the future

Important Water Property Affecting Process or Beneficial Use	**Affected Process or Beneficial Use**					
	Harmful Algal Blooms (HABs)	Impacts aquatic species[α]	Bioaccumulators in aquatic species[β]	Municipal and Drinking Water Supply	Agricultural and Industrial Water Supply	Recreation & Sport
Algal Blooms				x[8]	x[8]	x[8]
Bromide				x[9]		
Color				x		
Algal toxins[1]		x		x		
Dioxins			x	x		
Dissolved Oxygen (low levels)		x				
Electrical Conductivity (EC)					x	
Mercury & Methylmercury			x	x		
Microorganisms[2]				x[10]		x[13]
Nutrients[3]	x	x[5]		x[11]	x	x[14]
Odor				x		
Organic Carbon				x[9]		
Pesticides[4]		x[6]	x[7]	x		
pH		x		x		
Polychlorinated biphenyls (PCBs)			x	x		
Radionuclides		x		x		
Salinity		x		x	x	
Selenium		x	x	x		
Suspended Solids	x	x		x		
Temperature	x	x				
Total Dissolved Solids (TDS)					x	
Turbidity	x	x		x		
Trace Substances & Heavy Metals				x[12]		
Chemicals of Emerging Concern (CECs)		Edocrine distruptors, Pesticides, Pharmaceuticals, Personal care products		Dioxane, Flame retardants, Hexavalent chromium, Nitrosamines, Perchlorate, Pharmaceuticals, Personal care products		

[α] Negative impacts include hindered survival, reproduction, development, loss of habitat and migration corridors.

[β] Concern for food consumption.

[1] Produced by HABs.

[2] Viruses, bacteria, protozoa, parasites.

[3] Nitrogen (ammonia, nitrate), phosphorous (total P, phosphate).

[4] Herbicides, insecticides, fungicides.

[5] Ammonia.

[6] Organophosphorus pesticides (diazinon, chlorpyrifos), pyrethroid-based pesticides (bifenthrin, cyfluthrin, cypermethrin, lambda-cyhalothrin).

[7] Organochlorine pesticides (dichlorodiphenyltrichloroethane (DDT), chlordane, dieldrin).

[8] Hinders transport of water supplies through water delivery systems or water sport recreation.

[9] Disinfection byproduct precursors.

[10] Cryptosporidium and giardia.

[11] Nitrates in drinking water, nutrients cause odor.

[12] Arsenic, boron, cadmium, copper, cyanide, iron, lead, selenium, zinc.

[13] Fecal coliform and giardia.

[14] Causes algal blooms, invasive species and waterweed growth.

6.3.2 What Limits Water Quality Data?

Delta water quality data are plentiful; however, they exist in various formats, originate from scattered sources, and their connectivity to beneficial uses or ecological processes is not always clear. When there is an inability to connect water properties to beneficial uses or processes, information gaps result. Regardless of the type of water property or causal relationship being considered, there are three main root causes for gaps: monitoring, understanding and connectedness, and communication and coordination.

Gaps in monitoring can result from an incomplete or poorly targeted set of water properties. For example, important parameters such as water flow are often not collected concurrently with water property measurements, resulting in an incomplete dataset that can limit the usefulness of water quality data. A Delta source assessment of total and methylmercury found a balanced total mercury budget and an unbalanced methylmercury budget; 50 percent of the incoming methylmercury was not accounted for in the known and measured outgoing processes (California Regional Water Quality Control Board Central Valley Region, 2008). This discrepancy was likely due to incomplete monitoring of processes and events, such as missing data during events of higher methylmercury loss from the Delta. However, it is possible that methylmercury losses resulted from unknown process that were not monitored.

Monitoring efforts need to be of appropriate spatial and temporal coverage to be informative; it is unlikely a few or narrowly scattered datapoints can represent complex ecosystem processes in the Delta. Tidal effects, seasonal precipitation patterns, reservoir operations, irrigation schedules and withdrawal schedules are important considerations for water quality monitoring and can greatly affect spatial and temporal variability. The spatial and temporal requirements for data collection will depend on the specific goals of an investigation or the questions put forth. In the case of HABs, their occurrences are still poorly predicted despite current monitoring of temperature and nutrients in the Delta. Researchers believe spatial and temporal considerations contribute to formation of temporary seeding grounds where HABs seed before dispersal. More spatially and temporally comprehensive toxin monitors were put into place to identify and characterize these seeding grounds (Mioni, 2012).

6.3.3 Current Status and Future Efforts in the Delta

The current state of water quality data is inadequate for aiding adaptive management decision-making in the Delta (Delta Independent Science Board, 2018; Johnson et al., 2010). Future efforts to narrow data gaps will guide us towards a more comprehensive understanding of the Delta and development of effective management actions.

Dispersed data sources and incomplete documentation often make the use of historically collected water data a time-intensive endeavor. There is an increasing acknowledgement across academic and state entities of the need for more comprehensive monitoring programs and an organized, integrated and transparent data management system (Cantor et al., 2018; Delta Independent Science Board, 2018). Actions resulting from this acknowledgement have included the development of new or improved environmental monitoring (Delta Stewardship Council, 2018; Jabusch et al., 2018; California Water Boards, 2018; Aquatic Science Center, 2011). Research communities are also moving towards making previously and currently collected data more useable and accessible through data repositories (NERC, EDI data portal) and data-focused journals (e.g. Data in Brief).

Water properties and quality are transient parameters that are influenced by hydrologic, watershed, coastal, and atmospheric inputs as well as physical, biological and chemical processes occurring within the aqueous medium. This inherent transient nature makes data synthesis and interpretation difficult. To make water quality data more manageable, clear and concise goals must be defined to target specific parameters, and the spatial density and temporal frequency they are to be measured. An example of this type of strategy is the "use case" definition (Cantor et al., 2018), where multiple use cases are framed around the questions of "who needs what data in what form to make what decisions."

6.3.4 Modeling Data Needs and Implications for Modeling

Water quality models used in the Delta can be divided into two broad categories:

- Speciation/distribution-type models (Geochemical models, chemical models, transport models) that address the fundamentals of water properties, but not necessarily water quality. Examples of these include HEC-RAS, PHREEQC, USRWQM and WARMF.
- Discipline- or purpose-specific model. This category of models is more aligned with the definition of water quality provided here. Examples of these include CE-QUAL-W2, SBWQM and SWAT. These models have built-in functions that directly apply the simulated water properties to a beneficial use or ecological process; thus, water quality is defined within the model.

Depending on the model category and the specific model function, different types of data are needed to properly develop, calibrate, validate and run the model. The data needs for these two model categories are quite different, aside from concentration data that is generally needed for both types of models. Speciation-type models needs center on microscale data (e.g. kinetic, thermodynamic, chemical, elemental), while discipline-specific models needs center on macroscale data (e.g. hydrologic, meteorological, geometric, economic, agronomic, etc.). The availability of these two types of data source vary; microscale databases are often easily obtainable while macroscale data requires some synthesis by users.

Model development, calibration and validation are limited by available datasets. For example, a modeling effort undertaken to study the effects of air and water temperature on Delta aquatic species found that – even for commonly measured and easily obtainable data such as air temperature – temporal gaps limited the robustness of the analysis (Wagner et al., 2011). The author also noted that because of spatial limitations associated with the available datasets (Sacramento and San Joaquin River data only), the model would likely have reduced applicability to other locations within the Delta. Data gaps can also impact the accuracy of model boundary specification; Wright (2018) found this to be the case for sediment transport models, limiting physical processes for which these models could be constrained.

6.4 Fish Species

6.4.1 Water Operations Data

Delta fish models often incorporate statistical relationships between operations and fish parameters rather than differentiating the different mechanistic pathways that water operations may have on fish. The primary operations included in fish models are DCC and HORB operations and CVP and SWP exports.

6.4.2 Physical Data

Management operations in the Delta primarily affect the physical properties of the system (e.g., flow, exports, temperature, etc.) and many of the dynamics in Delta fish models are driven by the effect of changes to the physical system on biological parameters. 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; not daily or monthly changes in flow, exports, or X2. 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. Thus, daily or monthly flow at various points throughout the Delta is the most common data need for fish models.

Other physical data included in Delta fish models include temperature, salinity, turbidity, depth, and habitat features. Temperature is commonly included in life-cycle models because of the importance of temperature in egg survival and incubation duration. Depth and habitat features are primarily used for assessing habitat suitability and capacity.

6.4.3 Biological Data

Numerous biological data sources are used to inform parameters and relationships in Delta fish models. Life-cycle models typically require data on habitat capacity, survival, fecundity, spawn timing, sex ratios, and maturation rates. If data are limited, a single parameter value may be applied to all life stages and geographic regions of the model. If data are more extensive, parameters may be specified for different life stages and regions and, ideally, many of those parameters are functions of physical data (e.g., flow, temperature). Detailed movement models might require data on vertical and horizontal distributions of fish, proportion of fish entering different routes or structures, swimming behavior, and predator encounters. In these models, the data often are too limited to specify model parameters that depend on individual fish attributes (e.g., size, age, sex, hatchery origin). Moreover, data may not be available to inform each parameter and, instead, the parameter space is searched for combinations that match patterns in observed movement data. In individual-based models, many physiological, behavioral, and habitat parameters are required for parameterization. Such data-intensive models are rarely able to be fully parameterized to a specific system but instead rely on data from laboratory studies, different systems, and related species.

Evaluation of life-cycle models typically involves hind-casting and comparing to time series of observed abundances, possibly separated by life stage and region. Model evaluation requires effective techniques to convert sampled abundance to population estimates. Data limitations may limit evaluation of many fish models; under these circumstances models are evaluated based on expert review of model parameters, model relationships, and patterns observed in the model output.

6.4.4 Data Gaps

Previous Delta fish modeling efforts have illuminated significant knowledge gaps. Below, we briefly highlight some of the identified data gaps for Chinook Salmon and Delta Smelt models. For both species, a broad modeling need has been identified to couple individual-based modeling with 3-D hydrodynamic modeling. Modeling efforts of this magnitude will result in large data demands, thereby magnifying currently-identified data gaps.

6.4.4.1 Chinook Salmon

Advances in acoustic tag technology have allowed for detailed tracking of smolts and have greatly improved our understanding of route-specific survival and movement behavior of emigrating juvenile salmonids through the Delta. However, the minimum size requirements for tag implantation creates a data gap on fry and parr movement and survival. In particular, we need a better understanding of factors broadly affecting fry rearing and migration, and, more specifically, factors affecting habitat suitability, growth, survival, and residence time in potential rearing habitats such as off-channel habitat, floodplains, and estuaries. While many acoustic telemetry studies are conducted with hatchery fish, we need a better understanding of whether data from hatchery fish adequate reflect behavior and survival of wild-origin fish. Additional knowledge gaps include the effects of turbidity, salinity, contaminants, and nutrients on salmon behavior and survival.

6.4.4.2 Delta Smelt

The inability to effectively sample Delta Smelt in multiple habitat types and the inefficiency of the sampling gear has limited our ability to adequately assess the abundance and distribution of Delta Smelt. Key data gaps include effective population size, density-dependence, carrying capacity, microhabitat use, spawning areas, size-specific fecundity, growth, and mortality. Key knowledge gaps include effects of flow regimes on Delta Smelt vital rates (via effects on habitat), effects of non-native predators, and effects of Delta exports on population dynamics.

6.5 Consumptive Use/Evapotranspiration

Consumptive use is water that is lost from a watershed via evapotranspiration (ET). Evapotranspiration is the sum of evaporation from soil and transpiration from plants, measured in linear units over time, which can be multiplied by land area and time period to calculate a volume. Consumptive use and evapotranspiration are often used interchangeably and are typically the largest component of an agricultural region’s water balance.

Estimation of Delta-wide consumptive use has historically been hampered by scarce land-use and meteorological data. By overcoming these data obstacles, Medellín-Azuara et al.

(2018) accomplished a substantive leap in quantification of Delta consumptive use. These authors used prominent methods for estimating ET which included models, field data collection efforts, current land use data and remote sensing to develop multiple Delta-wide ET estimates.

The study estimated total annual evapotranspiration from crops in the Delta Service Area at 1,445 thousand acre-feet (TAF) in 2015 and 1,379 TAF in 2016. The mean departures of individual estimates from the ensemble mean were about 91 TAF for the Delta Service Area in both years, representing roughly 6.3% and 6.6% of the estimated ensemble means for 2015 and 2016.

6.5.1 Land Use

Land use can be derived from either a non-spatial (county census) or spatial (map) methods. Commodity production can be assessed using **County Agriculture Commissioner's Reports** (CCAC). These annual census data can provide a non-spatial snapshot of crop areas, gross revenues, and yields per county. However, they may be too coarse for spatially-explicit models (e.g. consumptive use, Medellin-Azuara et al, 2018).

Land use/land cover (LULC) maps are a snapshot of land use across a study area and provide critical information for agricultural modeling. LULC maps are often derived from satellite imagery through classification processes. The accuracy of maps is dependent on both the spatial resolution of the source imagery as well as the classification and validation methods. Incompatibilities between map/model crop classes, inaccuracies in classification, and coarse spatial resolution can all contribute to error propagation in spatially explicit models.

In the Delta region, at least two products have been used in agricultural modeling to provide spatially explicit land cover data – USDA NASS CropScape and Land IQ Land Use. Formerly, DWR maintained county-level land use surveys, with various iterations for the Delta in 1976, 1991, 1993, and select counties therein between 1976-2015 (https://water.ca.gov/Programs/Water-Use-And-Efficiency/Land-And-Water-Use/Land-Use-Surveys). Nevertheless, budget constraints and repurposing of funding to related programs has discontinued this effort.

The **USDA NASS CropScape** (https://nassgeodata.gmu.edu/CropScape) is a commonly used annual land cover datasets developed by the USDA. The greatest benefit of this product is their frequency, consistency, and availability. CropScape is updated annually and freely available through the USDA. However, this product has a coarse resolution and variable accuracy depending on crop type.

Land IQ (https://www.landiq.com/remote-sensing) is a high spatial resolution map produced by LandIQ, LLC and provided to the public by DWR. Although a high quality product, the commercial product is only produced on demand. As such, it provides a narrow snapshot of land cover in the Delta, the most recent products being the DWR 2014-2015 land use datasets. Data is available for download through the DWR Land Viewer application.

6.5.2 Data Gaps

Medellin-Azuara et al (2018) identified several areas where additional data and analysis are needed:

- ET estimates for idle agricultural land are highly uncertain and further investigation is needed.
- ET estimates for non-agricultural land use classes (including floating, riparian, and native classes) indicate their consumptive use rates may potentially be higher than those of irrigated crops in the Delta. Because most methods are primarily developed for agricultural land uses, these estimates require more examination. Refinement of land use information for non-agricultural land use classes could also be done through collaboration with other state agencies and research groups to better differentiate between specific natural vegetative types and potentially invasive species. This information is important for including restored landscapes and programs in regional water balances. Moreover, research at UC Berkeley has demonstrated that large acreages of wetlands in the Delta create a cooling effect over the long term which can lead to a reduction in ET rates relative to newly created wetlands and isolated small wetlands.
- The report presents some clear discrepancies between the field campaign ET and the modeled ET estimates. The long-term value and credibility of ET estimation will eventually require a better understanding of this difference between field and model results. Some strategies for resolving these differences include:
    - Conduct a field campaign focusing on detailed paired comparisons with a few modeled estimates, with uncertainty analyses of measurements and modeled ET estimates;
    - Involve multiple water experts in the field campaign, including independent networks such as FLUXNET-AmeriFlux, DWR, and other organizations and expert groups;
    - Explore the use of additional field-obtained data in modeling ET estimates, and compare the outcomes of additional field calibration and validation efforts;
    - Establish an ET program with some minimal base funding to maintain collaboration and advancement of ET quantification in the Delta.
    - Continue to improve quantification of non-agricultural consumptive use using field methods and updated models.
- Medellin-Azuara et al (2018) identified additional Delta-island water-balance and quantification issues related to evapotranspiration and evaporation. Specifically, isotopic data from throughout the Delta demonstrate that Delta island groundwater is partially evaporated channel water (HydroFocus, 2015). These isotope data demonstrate that seepage from surface-water channels onto Delta islands evaporates and thus contributes to consumptive use. This evaporation process, in conjunction with flow to drainage ditches, drives seepage. Modeling of the water balance on Delta islands (e.g., Deverel et al. 2017; Siegfried et al. 2014) and effective estimates of ET of applied water will be improved by better quantification of the shallow groundwater evaporation term.

6.6 Agricultural Economics

Models focused on the economics of agriculture use a variety of datasets in order 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.

6.6.1 Cost and Return Studies

These studies present the practices typical for a specific commodity in a specific region of California for a specific year. 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. Cost and return studies provide a breakdown of costs associated with labor, materials, equipment, and contract services to a high level of detail. 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. Although agricultural operations may vary greatly in practices and sales, the studies provide a streamlined method for understanding the economics of a commodity at the scale typically required by regional economic models.

Currently, cost and return studies are developed by the University of California Agricultural Issues Center (https://coststudies.ucdavis.edu). California is divided into 9 regions and 50 commodities including fruit, vegetable, field, tree, and vine crops, as well as beef cattle, sheep and other animals. Each dataset is updated periodically as needed.

6.6.2 Water Use

Water use is an important component of agricultural costs in California. To that end, DWR has collected data on water use and application methods for the state. The land and water use dataset (https://water.ca.gov/Programs/Water-Use-And-Efficiency/Land-And-Water-Use) produces annual, county-specific estimates irrigated crop acreages, crop evapotranspiration (ETc), effective precipitation (Ep), evapotranspiration of applied water (ETaw), and applied water (AW) for 20 crop categories. Every 10 years, DWR also collects data on crop irrigation methods and water use. The dataset is used in conjunction with agricultural models, such as SWAP (e.g. State Water Control Board, 2012).

6.7 Socioeconomics

Models focused on socioeconomic and development issues need a variety of datasets to link social and economic outcomes to land use and policy discussions and the associated changes and practices therein. Modeling challenges include limited access to proprietary data sources and geographic aggregation issues. Socioeconomic data are discussed below in the context of specific models in which they are embedded.

6.7.1 Regional Economy & Economic Change

Understanding regional economic changes or the impact of potential changes are often derived from forecasting models; these models are based on data from multiple data sources that are both publicly available and proprietary in nature. Forecasting models provide a snapshot of a local economy, offer a prediction of the future of a regional economy, and can foretell the effects on that same economy when a change is implemented. Such models, variously referred to as econometric models, input-output models, or computable general equilibrium models, can represent long term general equilibrium between supply and demand in a regional economy as prices, production, consumption, imports, exports, and other changes occur to stabilize the economic system.

In the Delta region two primary models have been used to model economic conditions and monitoring changes – REMI and IMPLAN. Each of these products combine data from multiple data sources to provide the user with expected forecasts.

6.7.1.1 REMI

REMI is a proprietary model that represents inter-industry relationships. The industry structure of a particular regional economy is captured within the model, as well as the transactions that occur between industries. The REMI model can represent long term general equilibrium between supply and demand as prices, production, consumption, imports, exports, and other changes occur to stabilize the economic system. The model incorporates data on energy, the environment, public policy, taxation and economic data gathered from the Bureau of Economic Analysis, The Bureau of Labor Statistics, the Department of Energy, the Census Bureau and other public sources. The greatest benefit of this product is that it is a mature model and data is updated annually. However, this product is proprietary and not publicly available and assembling necessary data to carry out similar input-output analyses is difficult.

6.7.1.2 IMPLAN

IMPLAN is a proprietary forecasting model that provides a snapshot of a local economy and facilitates the assessment of the economic impact of projects/actions in that economy. In doing so, the IMPLAN model estimates the “direct” and “multiplier” effects of economic changes on yearly revenues, employment, and “value added”. Here, value-added can be understood as the difference between revenues and the cost of non-labor business expense. Value added is the primary measure of the value of economic activity in a region and includes compensation for employees as well as income to businesses and landowners.

IMPLAN utilizes data from the system of national accounts for the U.S. based on data collected by the U.S. Department of Commerce, the U.S. Bureau of Labor Statistics and other federal and state government agencies. Data are collected for 528 distinct industry sectors which are classified on the basis of the primary commodity or service produced. Data sets are produced for each county in the U.S., allowing analyses at the county level and for geographic aggregation.

Data provided by the IMPLAN model includes outputs and inputs from other sectors, value-added, employment, wages and business taxes paid, imports and exports, final demand by households and government, capital investment, business inventories, marketing margins, and inflation factors (deflators). Data on the technological mix of inputs and levels of transactions between producing sectors are taken from detailed input-output tables of the national economy.

Although a high quality product, IMPLAN tends to provide upper bound estimates in relation to the annual economic loss from reducing a particular economy activity. Because the Delta region does not correspond to standard geographic areas available for most of the datasets used in the product, users are often tasked with constructing an area that roughly corresponds to the Delta from zip codes. This may lead to the model experiencing larger multiplier effects and reduce the certainty of model results.

6.7.2 Cost Benefit Analyses

Cost benefit Analysis (CBA) is an approach to estimate the strengths and weaknesses of alternatives and is often used to determine options which provide the best policy choice or option to achieve the desired outcomes while preserving savings. These models utilize a mixture of data sources to compare potential (or completed) courses of actions and/or to estimate the value against the cost of a decision, project, or public policy. CBA studies have two main applications: 1) to determine whether or not a decision is sound by understanding the extent to which its benefits outweigh its costs, and 2) to provide a basis for comparing investments (or decisions) by comparing the total expected cost of each option with its total expected benefits.

In the Delta Region, two CBA models used include HAZUS and F-RAM. Each of these models combine data from multiple data sources to provide the user with an understanding of the risks and benefits associated with public policy decisions.

- HAZUS, a risk assessment software package built on GIS technology, estimates multiple types of risks including: flooding, hurricanes, coastal surges, tsunamis, and earthquakes. The model estimates risk in three steps: 1) by calculating the exposure for a selected area; 2) characterizing the level or intensity of the hazard affecting the exposed area; 3) using the exposed area and the hazard to calculate the potential losses in terms of economic losses, structural damage, etc. HAZUS utilizes data from multiple sources including: national land cover data, HURDAT, USGS National Elevation Data set, U.S. Census Data, GIS Data Layers for Flood Forecasting, National Operational Hydrologic Remote Sensing Center data, and data from the U.S. Geological Survey. The model estimates multiple outputs including: physical damage to residential and commercial buildings, schools, critical facilities, and infrastructure; economic loss, including lost jobs, business interruptions, repair, and reconstruction costs; social impacts, including estimates of shelter requirements, displaced households, and population exposed to scenarios. Although the model is updated annually and consistently, it remains a work in progress and absolute value estimates should be interpreted with care.
- F-RAM is a model that provides a method for the rapid and consistent evaluation of floodplain management measures in a benefit-cost analysis framework. Two key concepts of the F-RAM are optimal knowledge and appropriate precision. F-RAM was developed to determine levee rehabilitation priorities within the San Joaquin River Basin during a task order by URS for DWR. F-RAM is an economic model that is used to measure one component driving investment decisions (floodplain

management measures) and cannot capture aspects of public safety, equity, or political factors that must be integrated into any analysis that seeks to comprehensively understanding a measure's costs and benefits. Moreover, while F-RAM is a mature and frequently updated tool, the tool is designed for relative (rather than absolute) comparison. As such, absolute estimates of flood damages that are derived from these models should be treated with caution when negotiating investment cost-sharing or considering new public policy issues.

6.8 Coordination in Data Collection

Given our focus on integrated modeling, data collection in a single discipline should not be considered in isolation. Understanding the interconnectedness between various natural drivers, stressors and endpoints of interest will improve data collection and advance our understanding of interacting processes. The performance of integrated models and insights developed through them will be enhanced by the availability of coordinated data collection across domains. Inter-disciplinary communication and collaboration is crucial for the success of such efforts. More collaborative data collection efforts and synergistic data management system (with consideration to quality assurance and quality control) are needed to improve the usefulness of data. Creating co-beneficial relationships, where all participants contribute a level of effort equivalent to their benefits gained, is important to facilitate communication and coordination. A critical gap in communication exists between researchers and policy makers; it has been noted that the frequency of interactions in the form of conferences and workgroups is not meeting science-based policy making needs, and a more regular system of communication would improve efforts for increasing water supply reliability and ecosystem health in the Delta (Tennefoss, 2018).

6.9 Summary

This chapter presents data needs across a range of Delta-relevant domains for model integration, with the goals of providing a general reference for modelers working across disciplines and identifying data gaps where appropriate. The specific disciplines covered in this chapter include: hydrodynamics, ecology, water quality, fish species, water budgets and consumptive use, agricultural economics, and socioeconomics. Large data collection efforts are ongoing in many of these areas; however, limited coordination is taking place across disciplines. As model integration becomes more commonplace across the Delta, coordinated sampling efforts in time and space will be needed to make the best use of these data for modeling.

7 References

Aquatic Science Center. 2011. The Pulse of the Delta: Monitoring and Managing Water Quality in the Sacramento – San Joaquin Delta. Aquatic Science Center: Oakland, CA. Available at: https://www.sfei.org/sites/default/files/biblio_files/2011_ASC_PulseOfTheDelta_final_0.pdf (accessed Feb. 15, 2019).

Argent, R.M., Perraud, J.-M., Rahman, J.M., Grayson, R.B., Podger, G.M., 2009. A new approach to water quality modelling and environmental decision support systems. Environmental Modelling & Software 24, 809-818.

Armstrong, C.W., Ford, R.W., Riley, G.D., 2009. Coupling integrated Earth system model components with BFG2. Concurrency and Computation: Practice and Experience. 21, 767-791.

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 (2013) 116-134.

Becker, B.P., Schuttrumpf, H., 2011. An OpenMI module for the groundwater flow simulation programme Feflow. J. Hydroinformat. 13 (1), 1-12.

Belete, G., Vionov, A., Morales, J., 2017. Designing the distributed model integration framework – DMIF, Environmental Modelling and Software, 94 (2017) 112-126.

Beven, K., Buytaert,W., Smith, L.A., 2012. On virtual observatories and modelled realities (or why discharge must be treated as a virtual variable). Hydrol. Process. 26 (12), 1905-1908.

Blind, M. and Gregersen, J.B., 2005. Towards an open modelling interface (OpenMI) the HarmonIT project. Advances in Geosciences, 4, pp.69-74.

Blythe, J.N., Dadi, U., 2012. Knowledge integration as a method to develop capacity for evaluating technical information on biodiversity and ocean currents for integrated coastal management. Environ. Sci. Policy 19, 49-58.

Box, G.E.P., and Draper, N.R., 1986. Empirical model-building and response surface. John Wiley & Sons, Inc. New York, NY, USA.

Buahin, C., and Horsburgh, J.S., 2018. Advancing the open modeling interface (OpenMI) for integrated water resources modeling, Environmental Modelling and Software, 108 133-153.

California Environmental Protection Agency. Sacramento – San Joaquin Delta Estuary TMDL for Methylmercury. Available at: https://www.waterboards.ca.gov/rwqcb5/water_issues/tmdl/central_valley_projects/delta_hg/archived_delta_hg_info/staff_report_jun06/delta_title_exec_sum_toc.pdf (accessed Feb. 19, 2019).

California Regional Water Quality Control Board Central Valley Region. 2018. The water quality control (Basin plan): The Sacramento River Basin and the San Joaquin River Basin, fifth ed. Available at: https://www.waterboards.ca.gov/centralvalley/water_issues/basin_plans/sacsjr_201805.pdf (accessed Feb. 15, 2019).

California Regional Water Quality Control Board Central Valley Region, 2008. Delta Methylmercury TMDL, Draft Report for Public Review. Available at: https://www.waterboards.ca.gov/centralvalley/water_issues/tmdl/central_valley_projects/delta_hg/archived_delta_hg_info/staff_report_feb08/tmdl_ch6.pdf (accessed Feb. 16, 2019).

California Water Boards, 2017. Constituents of Emerging Concern (CECs) and the Water Boards. Available at: https://www.waterboards.ca.gov/water_issues/programs/swamp/cec_aquatic/docs/cecinit_info_item_gtg_final.pdf (accessed Feb. 15, 2019).

California Water Boards. 2018. Surface Water Ambient Monitoring Program. Available at: https://www.waterboards.ca.gov/water_issues/programs/swamp/mission.html (accessed Feb. 15, 2019).

Cantor, A., Kiparsky, M., Kennedy, R., Hubbard, S., Bales, R., Cano,L., Guivetchi, K., McCready, C., and Darling, C., 2018. Data for Water Decision Making: Informing the Implementation of California's Open and Transparent Water Data Act through Research and Engagement. Center for Law, Energy & the Environment, UC Berkeley School of Law, Berkeley, CA. 56 pp. Available at: https://doi.org/10.15779/J28H01 or law.berkeley.edu/datafordecisions (accessed Feb. 15, 2019).

Collins, N., Theurich, G., Deluca, C., Suarez, M., Trayanov, A., Balaji, V., Li, P., Yang, W., Hill, C. and Da Silva, A., 2005. Design and implementation of components in the Earth System Modeling Framework. The International Journal of High Performance Computing Applications, 19(3), pp.341-350.

Delta Independent Science Board. 2018. Water quality science in the Sacramento-San Joaquin Delta, chemical contaminants and nutrients. Available at: http://deltacouncil.ca.gov/sites/default/files/2018/07/2018-7-26-FINAL-Delta-ISB-2018-Water-Quality-Review.pdf (accessed Feb. 15, 2019).

Delta Science Council, 2018. Delta Science Plan (draft update, October 12, 2018). http://www.deltacouncil.ca.gov/sites/default/files/2018/10/2018-10-22-20181012-DeltaSciPln-Draft-ISB-REVIEW-101218.pdf

Delta Stewardship Council. 2017. Science Action Agenda 2017-2021: A Collaborative Road Map for Delta Science. Available at: https://scienceactionagenda.deltacouncil.ca.gov/sites/default/files/2017-2021-SAA-final-Sept2017.pdf (accessed May 1, 2019).

Delta Stewardship Council. 2018. Delta Monitoring Enterprise Review: Request for proposals. Available at: http://www.deltacouncil.ca.gov/request-proposals-rfp-delta-monitoring-enterprise-review (accessed Feb. 15, 2019).

Deverel, S. J, Leighton, D. A, Lucero, C., and Ingrum, T., 2017. Simulation of Subsidence Mitigation Effects on Island Drain Flow, Seepage, and Organic Carbon Loads on Subsided Islands Sacramento–San Joaquin Delta. San Francisco Estuary and Watershed Science, 15(4). Retrieved from https://escholarship.org/uc/item/4q340190

Donatelli, M., Rizzoli, A.E., 2008. A design for framework-independent model components of biophysical systems. In: iEMSs 2008 Int. Congr. Environ. Model. Softw, vol. 2. iEMSs, Barcelona, Catalonia, pp. 727-734.

Dozier, A.Q., David, O., Arabi, M., Lloyd, W., and Zhang, Y., 2016. A minimally invasive model data passing interface for integrating legacy environmental system models. Environmental Modelling and Software, 80 265-280.

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.

Eichelmann, E., Hemes, K. S., Knox, S., Oikawa, P., Chamberlain, S., Sturtevant, C., Verfaillie, S., and Baldocchi, D., 2018. The effect of land cover type and structure on evapotranspiration from agricultural and wetland sites in the Sacramento–San Joaquin River Delta, California, Agricultural and Forest Meteorology, 256: 179 – 195.

EPA (US Environmental Protection Agency), 2008a. Integrated Modeling for Integrated Environmental Decision Making. EPA-100-R-08-010. Washington, DC. Office of the Science Advisor. http://www.epa.gov/CREM/library/IM4IEDM_White_Paper_Final_(EPA100R08010).pdf

EPA (US Environmental Protection Agency), 2008b. Workshop Report: Collaborative Approaches to Integrated Modeling: Better Integration for Better Decision-Making. December 10–12, 2008. Phoenix, AZ http://www.epa.gov/crem/integrated-modeling-workshop2008.html

EPA (US Environmental Protection Agency). 2018. National Primary Drinking Water Regulations: Ground Water and Drinking Water. Available at: https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations (accessed Feb. 15, 2019).

Goodwin, P., Enright, C., Medellin-Azuara, J., Lund, J.R., and Bray, B., 2016. A Workshop on Community Integrated Environmental Models May 21-23, 2015 in Davis, California: Project Outcomes Report. https://watershed.ucdavis.edu/files/biblio/Outcomes_Report_20160405.pdf

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. Environ. Modell. Software, 70 (2015), 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. Environ. Modell. Software, 73 (2015), 231-243.

Hutton, P.H., Rath, J.S., Chen, L., Ungs, M.J. and Roy, S.B., 2015. Nine decades of salinity observations in the San Francisco Bay and Delta: Modeling and trend evaluations. Journal of Water Resources Planning and Management, 142(3), p.04015069. Available at: https://ascelibrary.org/doi/pdf/10.1061/%28ASCE%29WR.1943-5452.0000617 (accessed Feb. 26, 2019).

HydroFocus, Inc., 2015. Groundwater Quality Assessment Report, San Joaquin County and Delta Water Quality Coalition Service Area. Available at https://www.waterboards.ca.gov/centralvalley/water_issues/irrigated_lands/water_quality/coalitions_submittals/sanjoaquin_delta/ground_water/2015_0427_sjdwqc_gar.pdf

Jabusch, T., Yee, D., Ross,J., Franz, A., and Heberger, M., 2018. Delta Regional Monitoring Program. San Francisco Estuary Institute-Aquatic Science Center. Available at: https://www.waterboards.ca.gov/centralvalley/water_issues/delta_water_quality/delta_regional_monitoring/wq_monitoring/2018_1119_drmp_qapp.pdf (accessed Feb. 15, 2019).

Jagers, H.R.A., 2010. Linking data, models and tools: an overview. In: Proceedings of iEMSs (International Environmental Modelling and Software Society).

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.

Johnson, M.L., Werner, I.N.G.E., Teh, S.W.E.E. and Loge, F., 2010. Evaluation of chemical, toxicological, and histopathologic data to determine their role in the pelagic organism decline. Final report to the California State Water Resources Control Board and Central Valley Regional Water Quality Control Board. University of California, Davis. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.364.7339&rep=rep1&type=pdf (accessed Feb. 15, 2019).

Karpatne, A., Atluri, G. Faghmous, J., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, S., Kumar, V., 2017. Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data, in IEEE Transactions on Knowledge and Data Engineering, vol. 29, no. 10, pp. 2318-2331, 1 Oct. 2017.

Knapen, R., Janssen, S., Roosenschoon, O., Verweij, P., De Winter, W., Uiterwijk, M., Wien, J.E., 2013. Evaluating OpenMI as a model integration platform across disciplines. Environmental Modelling and Software, 39 (2013) 274-282.

Laniak, G.F., Olchin, G., Goodall, J., Voinov, A., Hill, M., Glynn, P., Whelan, G., Geller, G., Quinn, N., Blind, M. and Peckham, S., 2013. Integrated environmental modeling: a vision and roadmap for the future. Environmental Modelling & Software, 39, pp.3-23.

Larson, J.W., Jacob, R.L., Foster, I. and Guo, J., 2001. The model coupling toolkit. In International Conference on Computational Science (pp. 185-194). Springer, Berlin, Heidelberg.

Lee, G.F., Jones-Lee, A. 2004. Overview of Sacramento-San Joaquin River Delta water quality issues. Available at: http://www.gfredlee.com/SJR-Delta/Delta-WQ-IssuesRpt.pdf (accessed Feb. 15, 2019).

Leimbach, M., Jaeger, C., 2005, A modular approach to integrated assessment modeling. Environ. Model. Assess. 9, 207-220.

Lloyd, W.J., David, O., Ascough, J., Rojas, K., Carlson, J., Leavesley, G., Krause, P., Green, T., Ahuja, L., 2011, Environmental modeling framework invasiveness: analysis and implications. Environmental Modelling and Software, 26 (10), 1240-1250.

Maxwell, T. and Costanza, R., 1997. An open geographic modeling environment. Simulation, 68(3), pp.175-185.

Medellin-Azuara, J. Lund, J.R., Goodwin, P., Enright, C., Bray, B., Argent R., Ariyama, J., Bratton, J.F., Burau, J., Chotkowski, M., Escriva-Bou, A., Lee, J., Lindley, S., McWilliams, M., Peckman, S., Quinn, N., Senn, D., Siegel, S., Wolfe, J., 2017. Integrated Modeling of Estuarine Systems: Lessons for the Sacramento-San Joaquin Delta. https://watershed.ucdavis.edu/files/content/files/Integrated_Environmental_Modeling_Policy_Brief_20170222r1.pdf

7. References

Medellín-Azuara, J., Howitt, R.E., Hanak, E., Lund, J.R. and Fleenor, W.E., 2014. Agricultural Losses from Salinity in California's Sacramento-San Joaquin Delta. San Francisco Estuary and Watershed Science, 12(1).

Medellín-Azuara, J., Lund, J.R., and Goodwin, P., 2016. Integrated Modeling for Adaptive Management of Estuarine Systems. Final Report of a Workshop at UC Davis. National Science Foundation Award Number: 1464440. 44p. Available at: https://watershed.ucdavis.edu/doc/integrated-environmental-modeling/white-paper-integrated-modeling-estuarine-systems. Last visit April 12, 2019.

Medellín-Azuara, J., U, P., Jin, K.T., Jankowski, Y. et al., 2018. A Comparative Study for Estimating Crop Evapotranspiration in the Sacramento-San Joaquin Delta. Center for Watershed Sciences, University of California Davis. https://watershed.ucdavis.edu/project/delta-et

Mioni, C., 2012. What Controls Harmful Algal Blooms and Toxicity in the Sacramento-San Joaquin Delta?. Available at: https://escholarship.org/uc/item/3qf633v9 (accessed Feb. 18, 2019).

Moore, R., A. Hughes, N. Gaber, G. Geller, P. Glynn, G. Laniak, A. Voinov, AND G. Whelan. International Summit on Integrated Environmental Modeling. U.S. Environmental Protection Agency, Washington, DC, EPA/600/R/12/728, 2012. Available at: https://cfpub.epa.gov/si/si_public_file_download.cfm?p_download_id=509013&Lab=NERL

Moore, R.V., Tindall, C.I., 2005. An overview of the open modelling interface and environment (the OpenMI). Environ. Sci. Policy 8 (3), 279-286.

Mount, J., Bennett, W., Durand, J., Fleenor, W., Hanak, E., Lund, J. and Moyle, P., 2012. Aquatic Ecosystem Stressors in the Sacramento-San Joaquin Delta. San Francisco: Public Policy Institute of California. Available at: https://www.ppic.org/content/pubs/report/R_612JMR.pdf (accessed Feb. 16, 2019).

NRC (US Nuclear Regulatory Commission), 2002. Proceedings of the Environmental Software Systems Compatibility and Linkage Workshop. NUREG/CP-0177, PNNL-13654. Hosted by The U.S Nuclear Regulatory Commission, Professional Development Center. Rockville, MD March 7–9, 2000. http://www.nrc.gov/reading-rm/doccollections/nuregs/conference/cp0177/

Omrani, H., Parmentier, B., Helbich, M. and 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, pp.182-191.

Peckham, Scott D.; Hutton, E.W.H.; Norris, B., 2013. A component-based approach to integrated modeling in the geosciences: The Design of CSDMS, Computers & Geosciences, Special Issue: Modeling for Environmental Change. 53: 3–12. doi:10.1016/j.cageo.2012.04.002.

Pollino, C.A., Hamilton, S.H., Fu, B., Jakeman, A.J., 2017. Integrated approaches within water resource planning and management in Australia—theory and application in (Hart and Doolan, ed) Decision Making in Water Resources Policy and Management, Elsevier

Rahman, J.M., Seaton, S.P., Perraud, J.M., Hotham, H., Verrelli, D.I. and Coleman, J.R., 2003. It's TIME for a new environmental modelling framework. In MODSIM 2003 International Congress on Modelling and Simulation (Vol. 4, pp. 1727-1732). Modelling and Simulation Society of Australia and New Zealand Inc. Townsville.

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.

Rizzoli, A.E., Davis, J.R., Abel, D.J., 1998. Model and data integration and re-use in environmental decision support systems. Decision Support Systems 24, 127-144.

Schut, P., Whiteside, A., 2007. OpenGIS Web Processing Service. OGC Project Document.

Senge, Peter M., 1990, The Fifth Discipline, The Art and Practice of the Learning Organization, Doubleday

Shamoo, A.E., Resnik, B.R., 2003. Responsible Conduct of Research. Oxford University Press.

Siegfried, L. J, Fleenor, W. E, & Lund, J. R., 2014. Physically Based Modeling of Delta Island Consumptive Use: Fabian Tract and Staten Island, California. San Francisco Estuary and Watershed Science, 12(4). Retrieved from https://escholarship.org/uc/item/3t82s21b

State Water Resources Control Board, 2012. Substitute Environmental Document in Support of Potential Changes to the Water Quality Control Plan for the San Francisco Bay-Sacramento/San Joaquin Delta Estuary: San Joaquin River Flows and Southern Delta Water Quality. Public Draft. December. (ICF 00427.11.) Sacramento, CA. https://www.waterboards.ca.gov/waterrights/water_issues/programs/bay_delta/bay_delta_plan/water_quality_control_planning/2012_sed/

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.

Tennefoss, A. 2018. Shared Science for the Sacramento-San Joaquin Delta. Available at: https://epm.ucdavis.edu/sites/g/files/dgvnsk296/files/inline-files/Aston%20Tennefoss%20Practicum_FINAL_1.pdf (accessed Feb. 16, 2019).

Thomas, J. W., 2000. A review of research on project-based learning. San Rafael, CA: Autodesk Foundation. http://www.bobpearlman.org/BestPractices/PBL_Research.pdf

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 Research, 150, 368-379.

Valcke, S., Balaji, V., Craig, A., DeLuca, C., Dunlap, R., Ford, R.W., Jacob, R., Larson, J., O'Kuinghttons, R., Riley, G.D., Vertenstein, M., 2012. Coupling technologies for earth system modelling. Geosci. Model Dev. 5, 1589-1596.

Vitolo, C., Scutari, M., Ghalaieny, M., Tucker, A., Russell, A., 2018. Modeling Air Pollution, Climate, and Health Data Using Bayesian Networks: A Case Study of the English Regions. Earth and Space Science, 5, 76-88.

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, 273-316.

Wagner, R.W., Stacey, M., Brown, L.R. and Dettinger, M., 2011. Statistical models of temperature in the Sacramento–San Joaquin Delta under climate-change scenarios and ecological implications. Estuaries and Coasts, 34(3), pp.544-556.

Weston, D. P., and Lydy, M.J., 2010. Urban and agricultural sources of pyrethroid insecticides to the Sacramento-San Joaquin Delta of California. Environmental science & technology 44, no. 5: 1833-1840.

Whelan, G., Kim, K., Pelton, M.A., Castleton, K.J., Laniak, G.F., Wolfe, K., Parmar, R., Babendreier, J., and Galvin, M., 2014. Design of a component-based integrated environmental modeling framework. Environmental modelling and software, 55(2014) 1-24.

Wright, S. 2018. Delta sediment measurements to support numerical modeling of turbidity. Available at: https://ca.water.usgs.gov/projects/2011-05.html (accessed Feb. 16, 2019).