United States Environmental Protection Agency

EPA/600/R-12/030 | March 2012 | www.epa.gov/gateway/science

Assessing the Challenges Associated with Developing an Integrated Modeling Approach for Predicting and Managing Water Quality and Quantity from the Watershed through the Drinking Water Treatment System

SCIENCE

Office of Research and Development

Assessing the Challenges Associated with Developing an Integrated Modeling Approach for Predicting and Managing Water Quality and Quantity from the Watershed through the Drinking Water Treatment System

by

Sandra C. Karcher
Department of Civil and Environmental Engineering
Carnegie Mellon University
Pittsburgh, PA 15213-3890
sck@andrew.cmu.edu

Jeanne M. VanBriesen
Department of Civil and Environmental Engineering
Carnegie Mellon University
Pittsburgh, PA 15213-3890
jeanne@andrew.cmu.edu

Christopher T. Nietch
U.S. EPA, Office of Research and Development
National Risk Management Research Laboratory
Water Supply Water Resources Division
Water Quality Management Branch, 26W MLK
Cincinnati, Ohio 45268
Nietch.Christopher@epamail.epa.gov

Contract No. EP-C-09-041
Work Assignment No. 1-17

National Risk Management Research Laboratory
Office of Research and Development
U.S. Environmental Protection Agency
Cincinnati, OH 45268

March 2012

Disclaimer

The U.S. Environmental Protection Agency, through its Office of Research and Development, funded and managed, or partially funded and collaborated in, the research described herein. It has been subjected to the Agency's peer and administrative review and has been approved for publication. Any opinions expressed in this report are those of the author(s) and do not necessarily reflect the views of the Agency, therefore, no official endorsement should be inferred. Any mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Abstract

Natural and engineered water systems interact throughout watersheds (e.g., at water intakes, wastewater outfalls and water pipe breaks of all kinds), and while there is clearly a link between watershed activities and the quality of water entering the engineered environment, surface water and drinking water are considered distinct operational systems. As a result, the strategic approach to data management and modeling within the two systems is very different, leading to significant difficulties in integrating the two systems in order to make comprehensive watershed decisions. In this paper, we describe a highly-structured data storage and exchange system that integrates multiple tools and models, describing both natural and engineered environments to provide a scientifically based, economic tool for assessing the impact of land use policy decisions on ecosystems and on the treatability of the water for human use. Our underlying objective in presenting our conceptual design for this water information system is to challenge the current paradigm for modeling water systems, and to advocate for moving towards the standardization of data storage and transfer protocols within the water science community.

Executive Summary

Engineered water systems (e.g., drinking water and wastewater treatment plants) and natural water systems (e.g., streams, rivers) interact throughout watersheds in a variety of ways. For example, natural waters enter engineered systems at drinking water plant intakes, treated water from wastewater plants is reintroduced into natural systems at outfalls, and water is exchanged between the two systems through leaking and/or broken pipes. Decisions regarding the management of natural systems upstream of drinking water treatment plant intakes (e.g., non-point source runoff from agriculture and livestock, and discharges from mining operations) affect water movement and biogeochemical processes, altering water conditions that then affect ecosystems and the treatability of water for human use. Clearly, there is a link between watershed activity and source water impairment; however, despite this connection, the typical response of watershed managers (e.g., reduce inputs from distributed multiple users) and drinking water plant operators (e.g., add additional treatment or alter processes) are made independently. Either as a consequence of the historical conceptual isolation of natural water systems from engineered/built systems, or as a result of managing the two separately, surface water and drinking water are considered distinct operational systems, and there is no standardized way of storing and sharing water data, and no single tool that can be used to model the quantity and quality of water as it moves through the natural and engineered water environments.

This report describes the state-of-the-practice in water information processing, and sketches the framework for the type of water information system (WatIS) that will be needed to manage water resources holistically. The proposed WatIS is fundamentally a system of models, communicating with a master data repository and integrated together with a robust user interface. Unlike the traditional method of linking models in series and cascading data from one model to the next, the proposed WatIS will allow data to flow in multiple directions. A well-structured database and a comprehensive data management strategy will be essential for the long-term success of the WatIS.

Data needed in modeling are discussed in detail in Section 2 of this report. The prominent models used for simulating water movement and biogeochemical processes in water systems are discussed in Section 3**,** along with model integration methods. The number and type of models needed in the WatIS will depend on the capabilities of the models and the specific problem to be addressed. This report focuses on modeling surface water, for which four types of processes will likely need to be simulated. A schematic of these processes is shown in **Figure ES-1**.

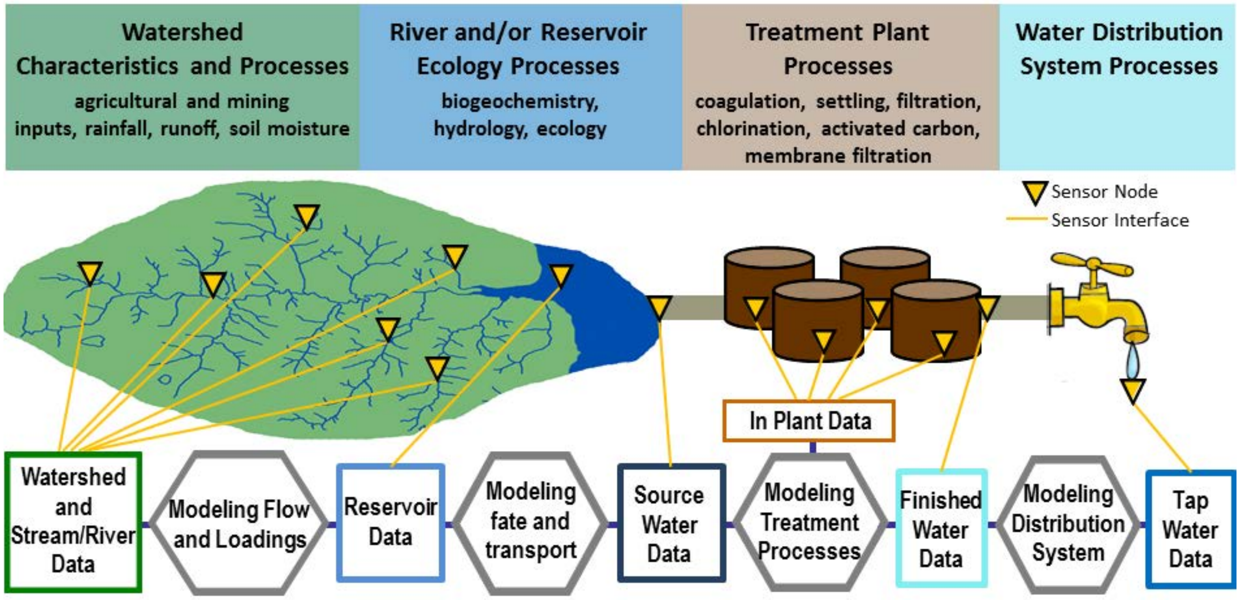

Figure ES-1. Schematic Relating Physical Water Systems to a Modeling Network.
The schematic represents the way water moves through natural and engineered systems, and shows the locations where data are collected and models are needed in the water information system.

As shown in **Figure ES-1**, water moves from left to right, starting in the watershed and working its way through the system until it exits as tap water. Along the way, information on the quantity and quality of the water is collected via sensors and/or grab-samples. These data are assimilated by the models and used to make predictions regarding the quality of the water as it travels from the watershed through the distribution system.

As part of this project, popular modeling tools were explored. As a result of the conceptual isolation of the natural system models and the built system models, the key parameters of the two types of models are often significantly different. For example, a watershed model may predict temperature, nutrient levels, and the algal biomass concentration in the source water, while the drinking water treatment plant requires the concentrations of taste and odor (T&O) precursors like geosmin and 2-methylisoborneol (2-MIB). There is no direct, easily incorporated translation from watershed parameters to drinking water intake parameters. Thus, to integrate these models, *treatability translation* tools will need to be included in the WatIS. These tools will require expert knowledge regarding the controlling physical, chemical, and biological processes and the accuracy of the computational algorithms will hinge on access to water quality monitoring in the watershed and on water quality monitoring at the drinking water treatment plant intake in order to inform the relationship among the different water quality terms. In addition, the operation and control of engineered water systems are typically managed using a supervisory control and data acquisition (SCADA) system. Thus, to alter treatment processes in real-time to maximize the quality of finished drinking water while minimizing treatment costs, the SCADA system must be fully integrated into the WatIS.

One of the applications of the WatIS is the development of a scientific framework for evaluating the relative costs and benefits of implementing changes in the watershed (i.e., the implementation of best management practices (BMPs)) with changes in the drinking water plant (i.e., using activated carbon). By incorporating these cost/benefit features, the decision space for maximizing the production of high quality drinking water at the least cost can be expanded to include decisions made in the watershed. To accomplish this goal, cost/benefit information must be aggregated and incorporated into the system, and cost/benefit computational tools must also be developed and added.

In the model section of the report, Section 3**,** a direct comparison is made between the proposed design of the WatIS that allows for multidirectional data flow via a master data repository and the traditional method of cascading data from model to model. The differences in these two approaches are summarized in **Figure ES-2**.

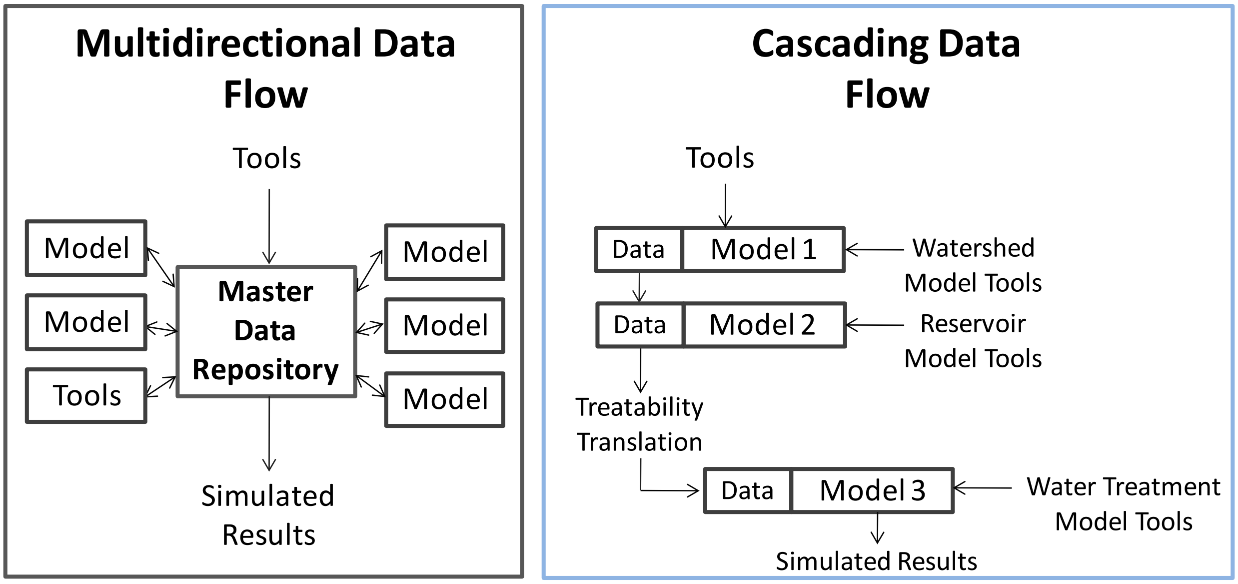

Figure ES-2. A Side-by-Side Comparison of Data Flow and Data Transfer Methods. The left image shows multidirectional data flow via a master data repository. The right image shows the more traditional method of cascading data downward through the models.

The benefits to the WatIS shown on the left are: (1) data can be accessed from all models in the system and data flow is multidirectional, (2) the data structure is common to all the models, encouraging data standardization among researchers, (3) the results of the model simulation and all the associated metadata can be stored in the data repository, and (4) the structure allows for a 'plug and play' model development. This approach will eventually reduce work for those using the models, but will require an extensive effort on the part of the model developers to transition to standardized data structures, and will require a long-term commitment to maintaining the master data repository and the user interface.

The cascading data approach offers the advantage of being able to be developed in pieces, one model at a time, but cannot overcome the limitation of unidirectional data flow. Multidirectional data flows are necessary to integrate real-time applications and facilitate adaptive management. Furthermore, without a master data repository, modelers will be required to learn data storage schemas for all the models they want to use, and will have no way of tracking or documenting the inputs and outputs of an integrated model simulation.

Adopting multidirectional flow and the unifying structure of the WatIS master data repository will expedite a ‘plug and play’ nature for model inclusion, facilitating the development of data analysis tools (and/or links to commonly used tools), the inclusion of treatability translation tools, and the incorporation of models for comparing costs/benefits, leading to a robust, integrated water information system for managing the quantity and quality of water across the natural and built environments.

Acknowledgements

The authors wish to thank the U.S. Environmental Protection Agency and the U.S. National Science Foundation for supporting this research. In addition, the authors would like to acknowledge the contributions of the members of the East Fork Watershed Cooperative, including: Elly Best, Michael Elovitz, Matt Heberling, and Hale Thurston from the USEPA-ORD-NRML; Steve Anderson and Lori Hillman from the USDA; Paul Braasch, Eric Heiser, Hannah Lubbers, John McManus, and Kevin Saunders from Clermont County; Michael Preston from UC Clermont College; and Lisa Underwood from the USACE. Additional contributors include Joel Allen and Jerry Waterman from the USEPA, and Sri Panguluri and Balaji Ramakrishnan from Shaw Environmental, Inc. The authors also wish to acknowledge Shaw Environmental, Inc. for their partnership in the project.

List of Figures

List of Tables

List of Acronyms

Acronyms	*Description*
AFW-ERPIMS	Air Force Wide Environmental Resources Program Information Management System
BASINS	Better Assessment Science Integrating Point and Nonpoint Sources
BMPs	Best Management Practices
DBP	Disinfection By-Products
DEM	Digital Elevation Model
DOC	Dissolved Organic Carbon
DWRS	Pennsylvania's Drinking Water Reporting System
DWT	Drinking Water Treatment Plant
EFW	East Fork Watershed
EN	National Environmental Information Exchange Network
ESRI	Environmental Systems Research Institute
GIRAS	Geographic Information Retrieval and Analysis System
GIS	Geographic Information System
HIS	CUAHSI has developed the Hydrologic Information System
HUC	Hydrologic Unit Codes
NED	National Elevation Database
NCDC	National Climatic Data Center
NLCD	National Land Cover Database
NOAA	National Oceanic and Atmospheric Administration
NOM	Natural Organic Matter
NSF	National Science Foundation
NWIS	National Water Information System
ODM	Observations Data Model
PASDA	Pennsylvania Spatial Data Access
SCADA	Supervisory Control and Data Acquisition
SSURGO	Soil Survey Geographic
STATSGO2	State Soil Geographic
STORET	EPA STOrage and RETrieval Data Warehouse
T&O	Taste and Odor
TDS	Total Dissolved Solids
TOC	Total Organic Carbon
US	United States
USACE	United States Army Corps of Engineers
USEPA	United States Environmental Protection Agency
USGS	United States Geological Survey
WaterML	Water Markup Language
WatIS	Water Information System
WQX	Water Quality Exchange
XML	Extensible Markup Language

Table of Contents

1 Introduction and Background

This report presents an assessment of the state-of-the-practice in water information management and modeling, and proposes a framework for the type of water information system (WatIS) that would be needed to: (1) understand the impacts of changing land use on the quality of receiving water as it pertains to the treatability of water for human use, (2) evaluate the socioeconomic implications of policy choices regarding the management of natural and engineered systems in an integrated way, and (3) enable the integration of data from multiple water systems to facilitate changes to engineered systems in real-time, and to identify long-term changes needed to achieve water quality and quantity goals.

1.1 Background of the Challenges of Implementing a Water Information System

The quality of the water that flows into a drinking water treatment plant is affected by the policies governing the water and land use upstream of the plant and by the way these policy choices influence the biogeochemical and ecological processes in the water systems. These dynamic processes often have indirect effects. For example, the application of fertilizer to farmland can lead to nutrient runoff (nitrogen and phosphorus) that enters the water system and leads to increases in primary productivity of algal species. The resulting changes in the ecology may increase concentrations of certain chemicals in the water, and these can lead to taste and odor (T&O) problems in finished drinking water. Similarly, discharges from mining operations can produce water that is high in dissolved chemicals (e.g., sulfate and chloride) that affects ecosystems and changes the quality of finished drinking water. Clearly, there is a link between watershed activities and source water impairment; however, despite this connection, the typical response of watershed managers (e.g., reduce inputs from distributed multiple users) and drinking water plant operators (e.g., add additional treatment or alter processes) are made independently. Either as a consequence of the historical conceptual isolation of natural water systems (watersheds, streams, rivers, lakes) from engineered/built water systems (drinking water and wastewater treatment plants), or as a result of managing the two separately, surface water and drinking water are considered distinct operational systems.

The divide that exists between those working in natural systems and those working in engineered systems affects the way that systems' information is managed. Professionals typically work in one system or the other, and become familiar with the unique features and complexities of their system. This makes the exchange of information between the two groups complicated, and affects the way tools are developed to model water systems. Historically, when a tool was needed to model system processes, it was developed to answer a specific question in one system, with little thought given to integrating and sharing models among multiple water systems. As a legacy of the way water systems have evolved, there is no standardized way of storing and sharing water data, and no single tool that can be used to model the quantity and quality of water as it moves through the natural and engineered water environments (Horshburgh et al. 2009).

1.2 Overview of Considerations in Water Information Management

In the domain of environmental science and engineering, models provide an organizing and integrating framework for fundamental knowledge on environmental processes and interactions.

In this sense, they serve as a repository for advances in scientific understanding of the complex processes that occur in natural and engineered systems. Environmental models also provide a basis for predicting changes to the environment in response to human activities. As such, when properly formulated, tested, and corroborated with observed data, they can provide a foundation and focus for decision support in the development of environmental policy (USEPA 1989a; Small 1997; Jakeman et al. 2006; Liu et al. 2008).

1.2.1 Modeling the Physical World

To provide a scientific basis for environmental policy decisions, a water information system must incorporate *data*, such as: (1) data uploaded from sensors deployed throughout natural and built environments, (2) data gathered through parameter analysis of individual grab samples collected from locations throughout natural and built environments, (3) data that are estimated by performing model simulations, and (4) data gathered from experiments designed to improve the parameterization and biogeochemical details of water system models, and must also include *models* that can: (1) be seamlessly integrated and used to simulate the physical and biogeochemical processes occurring in natural and engineered systems, and (2) predict the quality of finished drinking water at some future point in time when the parcel of water that was sampled in the watershed has moved through the treatment process. **Figure 1-1** shows a simplified schematic of the relationships among the physical water systems and the models.

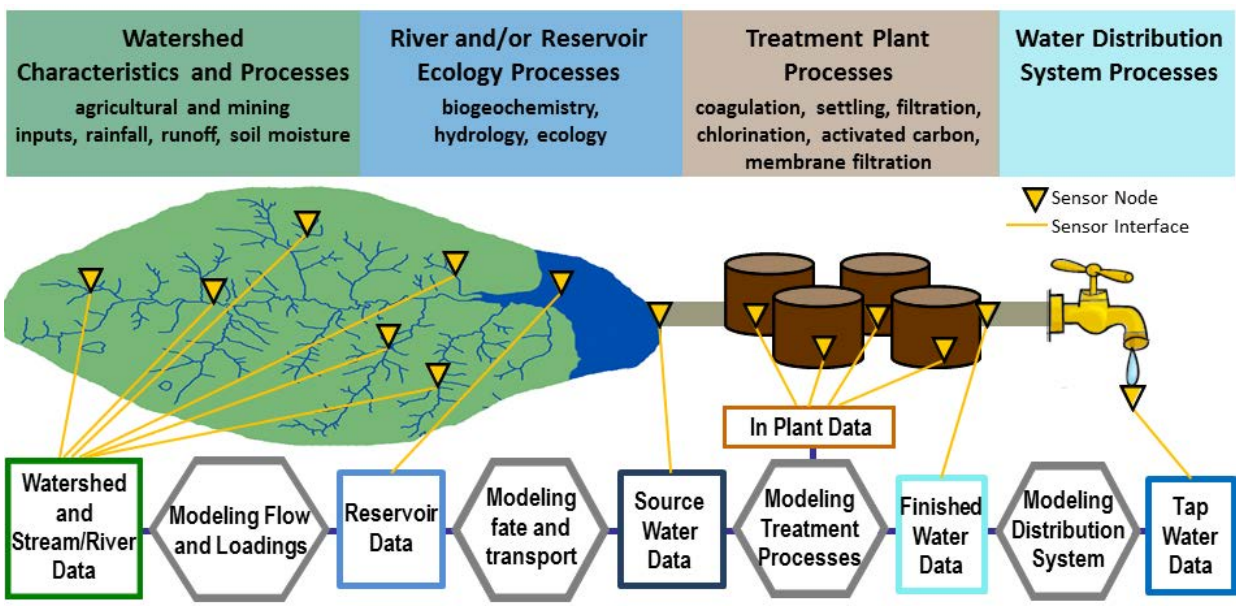

Figure 1-1. Relationship among the Physical Water Systems and Model Processes. Physical systems include a watershed, a receiving waterbody (e.g., river or reservoir), a drinking water treatment plant, and a finished water distribution system. Data are collected at various places in the systems, are stored in the water information system, and are used to update model parameters and confirm the predictive capability of the models.

In this conceptualization, water begins in the streams and tributaries of the watershed, where sensors capture data on water quantity and quality. A watershed-based model with land use features is used to simulate the runoff that leads to water quantity and quality changes in this part of the system. This water then enters a reservoir, where sensors and regular grab-sample monitoring programs provide additional data. These data inform a surface water model that incorporates extensive biogeochemical processes that are likely to occur in the water. These models generate profiles of water flow and quality, predicting the condition of the water at the drinking water treatment plant (DWTP) intake. The water continues to flow through the plant and the distribution system, where it is monitored using sensors and grab-samples. A series of unit operational models predict the final quality of the water using statistical relationships between source water surrogates and finished water parameters of interest. These finished water characteristics are also measured and sensed before entering the distribution system. In a fully functional WatIS, a multi-model simulation could be performed using sensor data gathered from locations some distance upstream of the DWTP intake, and the results of the simulation could be used to suggest and/or implement changes to the drinking water treatment process in real-time.

The schematic in **Figure 1-1** shows water flowing from the watershed through to the drinking water treatment plant and does not show the return flow of water through sewers to wastewater treatment plants and back into natural systems. Since the focus of this research is on laying a framework for the WatIS rather than on predictive modeling, some parts of the water cycle, even those that directly affect surface water quality, are not explicitly discussed. In the full deployment of the WatIS, all the pieces of the water cycle that impact the quality of the simulation will need to be incorporated and handled appropriately.

1.2.2 Components of a Water Information System

The traditional approach to integrating modeling is to transfer, or cascade the output from one model into input for the next model in series. This is the method used in the Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) tool (USEPA 2001; Kittle et al. 2006), as well as the approach discussed by other teams working on water models and their associated cyberinfrastructure (e.g., Finholt and VanBriesen 2007; Cuddy and Fitch 2010). While the cascading data method could be adapted for use in a robust water information system, it is limited by the restriction that data only flow in one direction. To allow for multidirectional data flow, an alternative method for integrating models in the WatIS is shown in **Figure 1-2**.

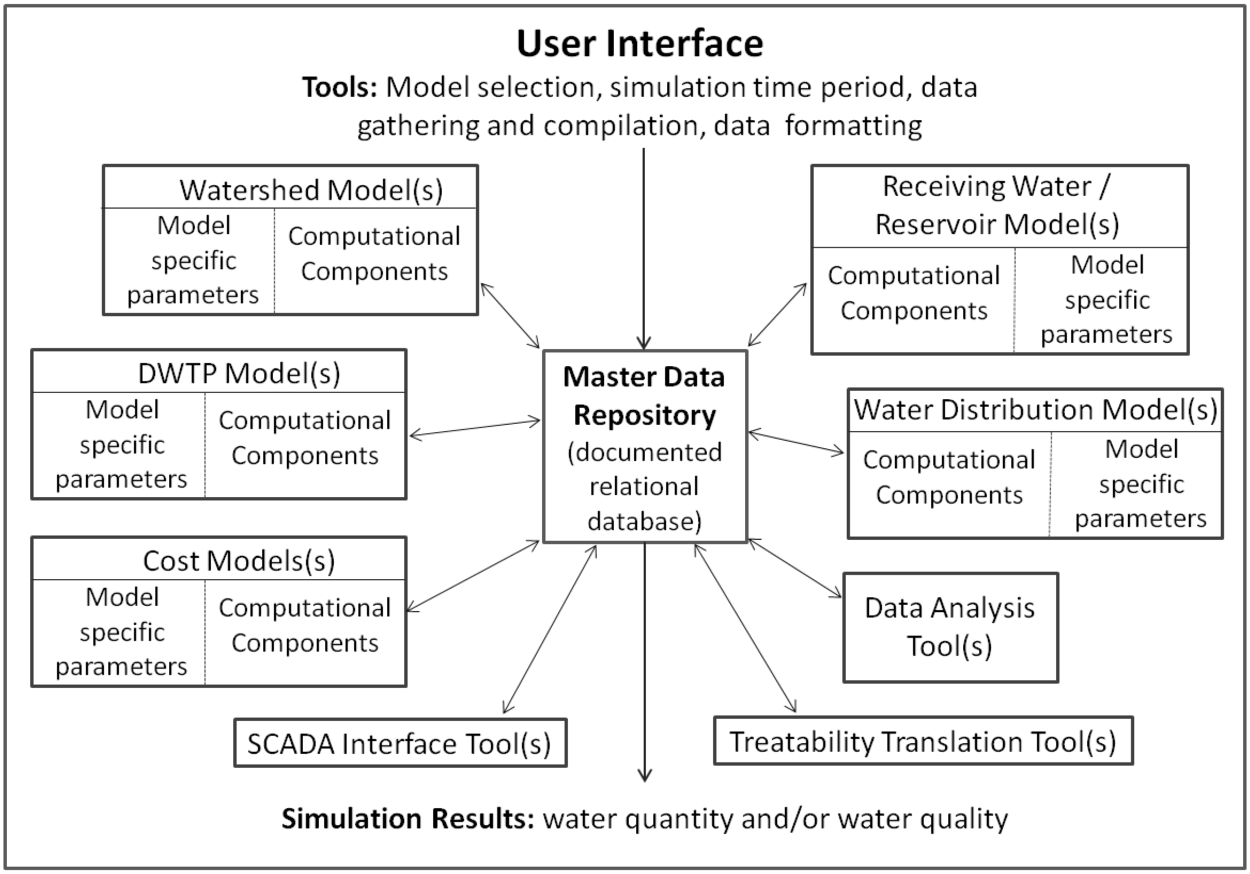

Figure 1-2. The Essential Components of a WatIS Needed for Multidirectional Data Flow. The WatIS consists of a master data repository and a group of models and tools. All models read from and write to a shared master data repository. Computational parameters and algorithms unique to a specific model may be stored separately; however, an indication of the parameter values and algorithms used in an integrated model simulation, enough to rerun the simulation at a later time must be stored in the data repository for use in documentation and optimization. The number of models that could be included in the WatIS is unlimited. The user interface keeps track of information common to all models, and, along with the results of the simulations, writes all metadata to the master repository for archiving, reviewing, reporting, and for comparing the results of multiple simulations.

There are three essential components of the WatIS: (1) the user interface, (2) the master data repository, and (3) the models and tools. The user interface simplifies the process of running a simulation of multiple models by providing a common look and feel as a frontend for all the models, and by holding information that is common to all the models, such as the dates of the simulation period. It also writes all the information to the master repository after the simulation is completed, giving a single input/output interface.

Prior to the 1980's, it would not have been possible to implement a master data repository approach to managing water system data due to limitations of computer hardware and software, but developments over the past 30 years have paved the way for a new paradigm in managing water resources data. The development of data management system software has made it possible for tools to be built on top of a data structure. One such tool, ARCINFO, introduced in the 80's by the Environmental Systems Research Institute (ESRI) (ESRI 2011), and subsequent versions of GIS tools have transformed the way researchers conceptualize and work with watersheds. With ESRI driving efforts to standardize geospatial data, it is now possible to take the next step in standardizing data needed to model water systems holistically. With large, multiagency collaboration projects such as the National Ecological Observatory Network (2011) and Ocean Observatories Initiative (2011) already underway, development of an archival system is essential if these data are to be available to current and future modelers.

1.3 Study Area for Problem Assessment

Although the selection of a specific location is not necessary for assessing the level of effort required to design and develop the WatIS, the East Fork Watershed (EFW) in Ohio is used here as a basis for framing the discussion of the challenges. The EFW is an ideal area for studying the database management issues associated with environmental modeling, and also for studying the interconnectivity between activities in the watershed and the quality of drinking water treatment plant (DWTP) source water due to (1) the amount of data collected as part of a watershed monitoring program, (2) an extensive monitoring effort conducted by the DWTP, and (3) the close collaborative relationship between professionals working in the natural and engineered systems within the watershed.

The East Fork Watershed makes up the lower 30 percent of the Little Miami Watershed. The Little Miami Watershed is a 1,710 square mile (4,429 square kilometer), fourth-level hydrologic unit (code 05090202) watershed in Southwestern Ohio (USGS 2010). A National Scenic River, the Little Miami flows almost 107 miles (172 km) from the Dayton-Springfield area to Cincinnati, where it drains into the Ohio River (Hedeen 2010).

There is substantial interest in source water protection in the EFW, with a program focusing on the water quality of Harsha Lake (aka East Fork Lake), a flood control run-of-the-river reservoir that is used for recreation as well as the source water for Clermont County's Bob McEwen Water Treatment Plant. Excess algal growth has been reported to occur in Harsha Lake in response to agricultural fertilizer use in the basin, and at certain times of the year, the herbicide atrazine can be detected above drinking water standards in the water intake, requiring removal by activated carbon (Hedeen 2010). Changes in drinking water treatment (e.g., addition of activated carbon or changes in coagulation and settling parameters) that are designed to control algal-derived taste and odor and/or pesticide problems can lead to a cascade of additional changes within the drinking water process dynamics.

To address in-stream water quality, stakeholders in the EFW are actively engaged in evaluating the feasibility of implementing a water quality trading program in the watershed. Water quality trading allows facilities with higher pollution control costs to purchase environmentally equivalent credits from other sources at a lower overall cost. In the EFW, there is an effort underway to determine if water quality objectives could be achieved by allowing wastewater treatment plants to purchase credits from farmers located upstream of the plant who implement pollution control techniques on their land.

The complexity of the anthropogenic impacts on the water quality in the EFW highlight the need for an integrated understanding of the water processes in the watershed, in streams and tributaries, in the reservoir, and in water treatment plant unit operations.

2 Database Management in the Water Information System

Modeling is a data intensive process, thus data management is an essential part of a water information system (WatIS). Ideally, a master data repository would exist where researchers and consultants collecting data within a watershed could (and would) report their results. The repository would be standardized, well documented, and maintained on an ongoing basis. Presently, water professionals collect data and store it in whatever format best suits their application. This can range from a pile of printouts in a filing cabinet to an electronic database with associated metadata for broad use by additional researchers. Many data originators lack the skills needed to reformat their data for broader community use, or do not have time and/or support to dedicate to data management. The management of environmental data must be a multidisciplinary effort, and must include environmental scientists, water professionals, and information technology specialists (WERF 2001; NSF 2007; Horshburgh et al. 2008; Horshburgh et al. 2009; Dozier et al. 2009).

2.1 Moving Data to a Shared Database

The amount of work and resources required to gather, compile, organize, and store data in a way that is meaningful is often significantly under estimated. This level of effort, shown schematically in **Figure 2-1**, grows as increased standardization and documentation are included.

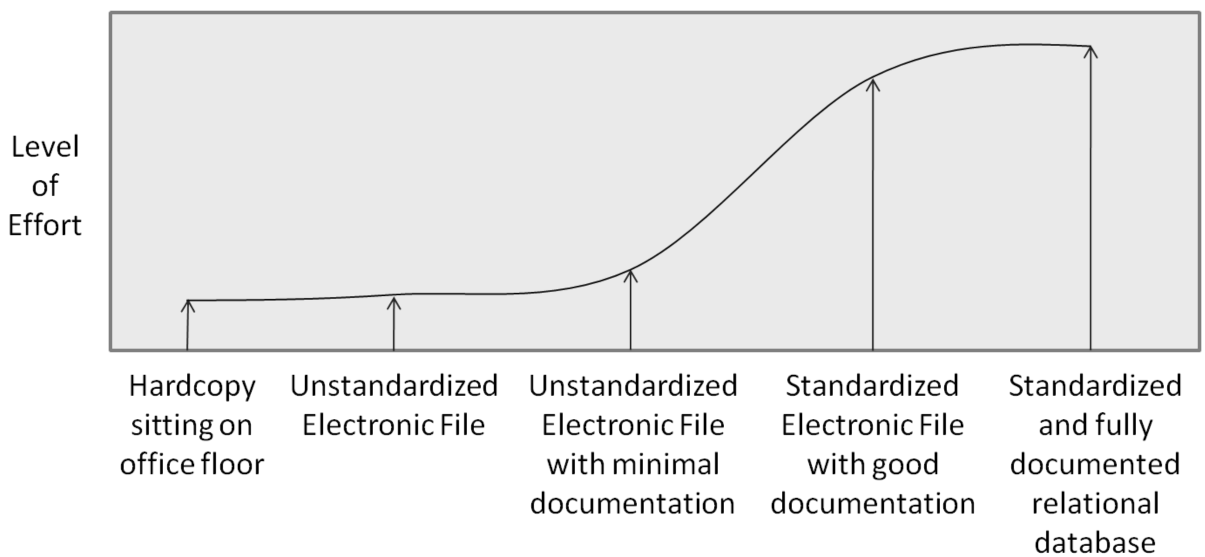

Figure 2-1. Level of Effort Required to Transition to a Multiuser Data Environment. It takes a significant investment of effort to move data from an unstandardized electronic format into a standardized, fully documented database. Without external motivation and support, it is typical for data collected as part of research studies and field sampling programs to reside in undocumented and unstandardized formats.

As shown in **Figure 2-1**, the least effort is required to store data in some kind of hardcopy format, such as a logbook, or in a non-standardized electronic file, such as an undocumented Excel worksheet. Data stored in this manner are difficult to share without extensive communication between the original researcher and the next data user. The most difficult part of organizing data so that they can be shared with others is standardization and documentation. One of the reasons why standardizing and documenting is so labor intensive is that it forces the data originator into the role of an information technology specialist. Transitioning to a standardized, documented database requires several critical steps: (1) selecting a framework for storing data, (2) determining all the important information about the data to be stored, and (3) preparing extensive documentation. These can be onerous tasks, and their completion is not typically of high importance to the original researcher. Consequently, the results of many field sampling programs sit in hard copy reports in file cabinets, or in Excel files that can only be understood by the original researcher.

2.1.1 A Move Towards Standardization

The understanding of the fundamental importance of managing environmental data is not new, and several databases have been developed to enable data management for water and/or environmental systems. Some have been developed with a specific purpose other than ongoing data sharing. For example, the Air Force Wide Environmental Resources Program Information Management System (AFW-ERPIMS), a database maintained by the Air Force Center for Engineering and the Environment, was designed to hold data collected as part of Air Force environmental projects (AFCEE 2009) and reduce the amount of sampling that needed to be done when different contractors were working at the same site. Other data systems have been developed specifically to keep the public informed. For example, Pennsylvania's Drinking Water Reporting System (DWRS) allows users access to water quality sampling data and the violation history of public drinking water facilities (PADEP 2011).

In an attempt to begin the process of standardizing data in the research community, the National Science Foundation (NSF) implemented a policy requiring that all proposals due after January 18, 2011 contain a data management plan (NSF 2011). The data management plan is to explain how the handling of data will conform to the NSF's *dissemination and sharing of research results* policy. While this is a good start, the NSF only requires that the researcher be able to *explain* their results (NSF 2011). Thus, the project data will still likely fall somewhere on the left side of **Figure 2-1**; meaning that future data users will need to communicate with the original researcher to know how to appropriately use the data. The best way to avoid this situation is to provide the original researcher with a well-structured and well-documented database for their data and with resources to assist them with data formatting and data loading.

2.1.2 Importance of Knowing the Target Format

When formatting data for community use, there are two extremely important pieces of a well documented database: (1) a *data definition dictionary*, and (2) a set of *valid values* (sometimes called a *controlled vocabulary* (Horshburgh et al. 2008, Gaber et al. 2008)). A data definition dictionary details the way data are to be organized and stored. In the simplest sense, it can be thought of as similar to defining the format of a table. A table includes columns and rows; with the column headings indicating what information is to be stored in that column. Consider a table that lists the location of major universities in the United States. The *data definition dictionary* specifies, by column, what information will be included in the table (for example, column three is to contain the abbreviation for the state in which the university is located). The *controlled* *vocabulary* details the acceptable abbreviations for each state. Using recognized abbreviations for each state reduces confusion when transferring data from user to user, reduces the possibility of misspelling the name of the state when entering it into the database, and reduces the amount of electronic storage space needed. A table of valid values, along with their corresponding expanded descriptions can be stored in the database and can be associated to the abbreviated values for display and/or reporting.

Providing data collectors with a data management structure (complete with valid values) *a priori* can significantly reduce the effort required for them to load their data into a documented database. When a target format for data storage is available, individuals are not forced to design and construct their own, and they do not need to document their structure or valid values. Support from an information technology professional will still likely be needed to help develop tools to simplify the data entry process, and to bridge any gaps in understanding regarding the structure of the database system. Costs associated to this effort must be weighed against the cumulative savings of the effort of every data collector attempting to develop his/her own data storage system. The major advantage to storing data in a well-documented database is that other researchers will be able to use the database (and know how to use it appropriately) without needing to communicate directly with the original researcher.

2.2 Databases of Interest

Some databases have been designed and developed with the goal of facilitating the sharing of data among agencies and environmental professionals. These include three databases, primarily used for water related data: the United States Environmental Protection Agency (USEPA) STOrage and RETrieval Data Warehouse (STORET), the United States Geological Survey (USGS) National Water Information System (NWIS), and the CUAHSI Hydrologic Information System Observations Data Model (ODM). The pros and cons of adopting one of these databases for use with the WatIS are summarized in **Table 2-1**.

Table 2-1. Databases Considered for Inclusion in the WatIS.

Database	***Pros***	***Cons***
EPA STOrage and RETrieval Data Warehouse (STORET) (USEPA 1989b)	Encourages data submission. Offers some support.	See table note 1.
USGS National Water Information System (NWIS) (NWIS 2011a; NWIS 2011b)	Quality of data allowed into the database is controlled.	Does not encourage or support submissions from outside USGS. Ongoing support would be required. See table note 1.
CUAHSI Hydrologic Information System Observations Data Model (ODM) (CUAHSI 2011c)	Encourages and supports publication of data. Offers ongoing support to users.	May not be as tightly controlled as needed to assure entry of high quality data. See table note 1.

[1] For all three databases, data structures would need to be modified to accommodate different types of data including socioeconomic data, quality assurance/quality control data, modeled data, and engineered systems data.

The three databases listed in the table all pertain to natural water systems. Water utility data have historically been managed through proprietary supervisory control and data acquisition (SCADA) systems.

Each of the three databases listed in **Table 2-1** are described in more detail in the following sections.

2.2.1 USGS National Water Information System (NWIS)

The USGS National Water Information System (NWIS) was designed as a repository for the stream flow, groundwater level, and water quality data collected as part of the USGS's extensive monitoring network (NWIS 2011a; NWIS 2011b). The USGS has a web service that allows for the retrieval of data from NWIS, but does not offer the capability for data collectors outside of NWIS to submit data to the database. There are two significant advantages to restricting entry into the database: (1) it makes the design of the database simpler by limiting the types of data the system must accommodate, and (2) the degree of credibility associated with the data in the database is better known.

While the data stored in NWIS is useful, and is often the best source of data available for calibrating models of natural water systems, since it does not allow for ongoing expansion to meet the needs of a variety of users, the NWIS database is not a good candidate for the water information system data repository.

2.2.2 EPA STOrage and RETrieval Data Warehouse (STORET)

The USEPA's STORET was design to be a repository for data collected by the USEPA and by other agencies required to report their data to the USEPA. Prior to the 1960s, little thought was given to the notion that data collected as part of routine monitoring might have other uses. In part, this was because there was no good way to store and share information. Computers were just beginning to be used in the workplace, and electronic communication and the internet had yet to be invented. Consequently, sharing data typically meant copying and mailing hard copy results, and/or converting data into a useful electronic format. As computers evolved into being commonplace, the concept of STORET, to establish a single structure for water quality data, took shape and was initially implemented in 1964 on a Public Health Service Honeywell computer in Cincinnati (USEPA 1989b). STORET is still actively used, and the USEPA supports the use of STORET by providing some tools for working with STORET data.

STORET can accept data submitted from many recognized partners of the STORET program. The USEPA website indicates that data can be received from states, tribes, citizen science groups, federal agencies, and universities. Data are transmitted to STORET via an Extensible Markup Language (XML) protocol. USEPA's transfer protocol is called Water Quality Exchange (WQX), and follows the terminology described by the National Environmental Information Exchange Network (EN) (2001). STORET has a back-end set of database tables that hold submitted data. STORET also offers some web services; including some tools to download data from STORET, and tools aimed at facilitating the generation of WQX files to transfer data to the STORET warehouse.

It is possible that the STORET data structure could be used as a starting point for the database management piece of the WatIS. This would require an ongoing partnership between USEPA and multiple water data generators as changes will be required to the STORET structure to accommodate additional types of data (for example, results from a modeling tool). Ongoing maintenance of the system would be needed. For example, additional users will need to be added, ongoing support will be needed for data validation, and the controlled vocabulary would need to be expanded and maintained.

2.2.3 CUAHSI Hydrologic Information System Observations Data Model (ODM)

Currently, of the three water databases listed in **Table 2-1.** The strongest candidate for incorporation into the WatIS is the Observations Data Model being developed by the Consortium of Universities for the Advancement of Hydrologic Sciences (CUAHSI). CUAHSI is funded by the National Science Foundation for the purpose of providing services and developing infrastructure to support the advancement of hydrologic science. CUAHSI has developed the Hydrologic Information System (HIS) with the intent of making water data universally accessible (CUAHSI 2011c). Some of the background and conceptual organization of the HIS database have been published (Horshburgh et al. 2008; Horshburgh et al. 2009).

The CUAHSI-HIS is conceptualized as a triangle of data discovery, data publication, and data access. Data discovery includes data storage and searching capabilities; data publication includes organizing and posting data so they can be harvested by other users; and data access includes allowing others to have the ability to use published data (CUAHSI 2011c). The back-end structure of the CUAHSI-HIS is called the Observations Data Model (ODM) (CUAHSI 2011a). While the CUAHSI data system is most closely aligned with the needs of the WatIS, significant adaptations and enhancements would be required for the CUAHSI ODM to meet the needs of the WatIS. Specifically, the ODM does not currently handle socioeconomic data, nor is there a standardized schema for tracking input and/or output data from water system models. Furthermore, since the goal of the CUAHSI data system is open information exchange, the quality of the data contained within the database could be widely variable and may not be well documented. The CUAHSI ODM will also need to be modified to facilitate integration with SCADA systems designed to assist in optimizing the water treatment process by allowing the operator to monitor and control equipment and processes in real-time (Lahlou 2002).

Data are transmitted to CUAHSI via an Extensible Markup Language (XML) protocol. CUAHSI's transfer protocol is called Water Markup Language (WaterML) (CUAHSI 2011b). CUAHSI has also developed *WaterOneFlow* (Beran et al. 2006), a web service tool that provides access to data stored in the CUAHSI ODM repository and a few additional non-CUAHSI databases. CUAHSI HydroDesktop is another tool that can be used to access data that have been published in the ODM repository. HydroDesktop is a geographic information system (GIS) application with a few specially programmed features. The main function of HydroDesktop is to allow the user to query the CUAHSI recognized/registered databases (via the internet) using a GIS query interface (the request to get data is generated from selecting a location on a map). After the query is run, the dataset is stored on the local computer. HydroDesktop also has some cursory graphing and analysis tools, but serious data users will likely export their data (a feature that is provided within HydroDesktop) and use an analysis/graphing package of their choosing.

2.3 Uncertainty and Variability in Data

Data collected as part of an environmental sampling program incorporate multiple sources of uncertainty and variability, some that are known, and some that are unknown. Some commonly recognized sources of uncertainty and variability include: (1) errors/problems/limitations associated with sample collection, (2) cross-contamination during sample collection, transport, and/or analysis, (3) errors/problems/limitations associated with sample analysis, and (4) errors/problems in reporting. In addition to error-introduced uncertainty, data variability is a natural feature of sampling and analysis methods. Databases that contain environmental data must incorporate methods to store data associated with sample duplicates, analytical duplicates and quality assurance, quality control samples that are collected and analyzed as part of field sampling and laboratory work.

Another type of uncertainty that must be addressed in the WatIS is the uncertainty that is associated with data that are generated during modeling. The WatIS will need to track and store: (1) data that are directly measured, such as the concentration of dissolved oxygen in a watershed tributary, *and* (2) data that are modeled, such as the concentration of dissolved oxygen at the entrance to the drinking water treatment plant that has been simulated using calibrated models based on the concentration of dissolved oxygen measured at other locations in the system. Traditionally, only observed data are stored in shared databases, but as the use of models increases, simulated data (along with their associated uncertainty) will also need to be stored for multi-user access. This is especially important when simulations are time consuming or based on sampling from input data distributions and thus represent an investment that would be costly to repeat every time that simulated result was needed for another decision. The ability to propagate and track uncertainty throughout the WatIS will be important to decision makers when trying to set policies to most effectively allocate resources, as well as to scientists trying to understand where to focus sampling and/or analytical efforts to reduce the uncertainty in the model predictions.

2.4 Data Needed for Modeling

To model multiple water systems, the WatIS will be required to contain extensive information on the watershed, receiving waterbody, and drinking water systems being studied. **Figure 2-2.** presents the conceptual watershed to drinking water system flow, listing some of the key information needed for modeling.

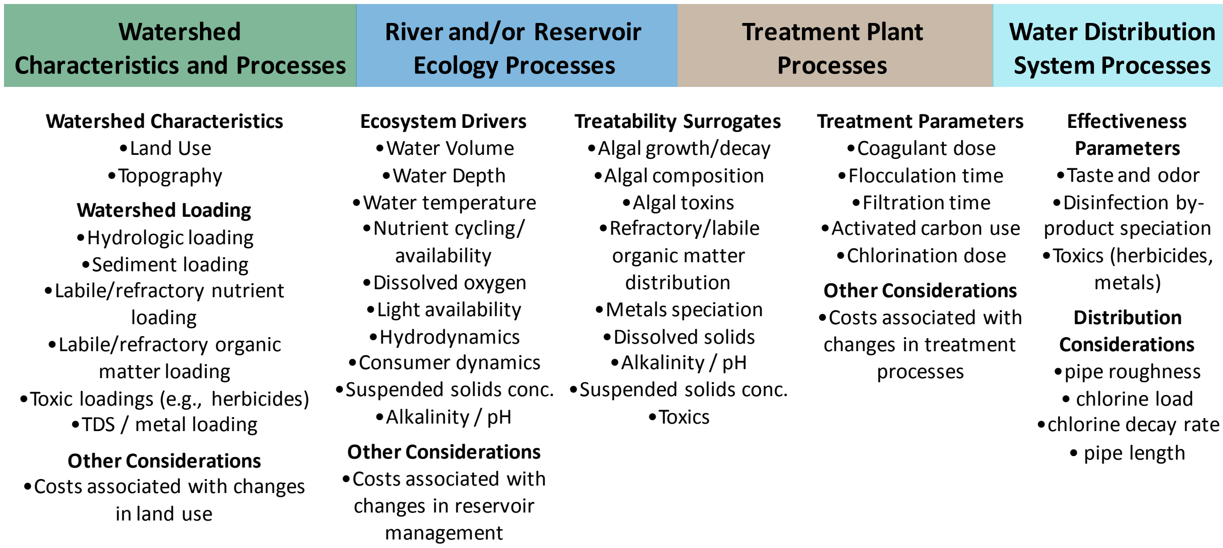

Figure 2-1. Schematic Relating Water Processes to the Information Needed for Modeling. Modeling water as it travels from the watershed to the distribution system requires a significant amount of information. For example, watershed models require hydrologic and sediment loading characteristics in order to predict the flow of water and sediment into the reservoir, and drinking water treatment plants use coagulant dose and flocculation retention time to predict the amount of suspended solids that will be removed during settling and filtration.

Modeling, even using an individual model, is an iterative process. Consider, for example, a modeler sets out to estimate the concentration of total ammonia (NH_3 + NH_4^+) in Lake Harsha at the point where water is extracted from the lake for use in the drinking water treatment plant. The modeler would go through the process of gathering data required for modeling, and formatting these data as needed. The modeler will then likely run the model to confirm that all the pieces of data are present and formatted correctly for use. An understanding of the uncertainly associated with these data may be well characterized, but often the modeler will be using data from a variety of sources without details of accuracy and associated uncertainty. Once data are added and the model is functioning, the modeler must calibrate the model. During this process, the modeler will compare simulated loadings with observed loadings, and adjust model parameters (such as the scour potency factor or the soil detachment coefficient) to achieve the best agreement between the two. Selecting the parameters to adjust, and how to adjust them requires professional judgment and an understanding of the dominant processes in the watershed being modeled. Parameter values have an associated uncertainty, and this uncertainty is generally not well understood.

The above example describes the ideal modeling scenario; however, what often happens in modeling is that data are missing, or contain gaps, and the modeler must decide how to appropriately fill the gaps so the model will run. In other cases, there may be data available to the modeler, but in a format that makes incorporation into the model unfeasible (either due to lack of resources to devote to data management, or because the uncertainty associated with the data is not well understood). In these situations, where measured data are not available, the modeler may need to synthesize data or simulate data to fill the gaps. For clarity, the types of data used for this project are described in **Table 2-2**.

Table 2-2. Description of Data Types Used for this Project.

Type	***Description***
synthetic	Data that are generated (not observed) or data observed at a different location used to demonstrate model functionality. Not suitable for predicting actual conditions at the site.
measured	Observed in the real world, formatted, and stored for model use. Could include sensor data as well as data from grab samples or relevant laboratory experiments. Could also include geophysical characteristics of the systems being studied.
simulated	Data obtained from executing : (1) *calibrated* models using *synthetic* data, or (2) *uncalibrated* models using *measured* data, or (3) *uncalibrated* models using *synthetic* data (for this project, none of the models were calibrated, thus, the results presented in this report should all be considered *simulated).*
modeled	Data obtained from running *calibrated* models using *measured* data (although, the uncertainty of the results may still be difficult to characterize).

Gathering and formatting data from the EFW were beyond the scope of this project, but the EFW is an ideal candidate for a case study focusing on the impact of changes in land use upstream of the drinking water treatment plant intake on the treatability of the water for human use due to the volume of samples collected in the watershed. Further, an extensive sensor network has been deployed in the EFW, and samples are collected regularly from within the watershed and from Harsha Lake to support researchers, the U.S. Army Corps of Engineers (USACE), and water professionals operating the county's drinking water treatment plant. More information on the specific data used for this project is provided in the modeling section of this report (**Section 3.2**).

3 Modeling in the Water Information System

Some of the features needed to model the surface water components of the water cycle are already available in commonly used models. Some of the features needed to model the drinking water treatment plant operations are also available, although these models are less frequently used for prediction.

The number and type of models that will be needed in the WatIS depend on the goal of the modeling and on the capabilities of the models. If land use changes are to be considered (e.g., implementation of best management practices (BMPs)), a watershed model is needed. If reservoir behavior is to be considered, the watershed model must be supplemented with a surface water quality model. To integrate the prediction of finished water quality into the WatIS, a treatment plant model is needed. To simulate the changes that occur in finished water quality as it travels to the consumers' taps, a water distribution system model would also be needed.

Since the East Fork Watershed system includes a reservoir, and the objective is to understand how changes in watershed use affect finished water at the drinking water plant, three models were investigated: a watershed model, a reservoir model, and a drinking water treatment plant/water distribution model.

3.1 Overview of Model Selection

There are many choices for modeling water systems (WERF 2001; Borah and Bera 2004; Borah et al. 2006; Migliaccio and Srivastava 2007; Park et al. 2008; Booty and Benoy 2009). Some models focus on water quantity, and some on water quality. Some models focus on the watershed, and some on reservoirs. Some models are freely available, and some are expensive. Choosing from the numerous models can be a daunting task. To help water quality managers and others interested in using mathematical models to evaluate the effectiveness of changing watershed management strategies, the Water Environment Research Foundation (WERF) published *Water Quality Models: A Survey and Assessment* in 2001. The authors of the WERF report evaluated approximately 150 models, segregating the models into model classes by their function. Model classes are described as follows (WERF 2001):

- *hydraulic or hydrodynamic models* – determine the circulation, transport, stratification, and depositional processes within a receiving water,
- *rural and urban pollutant runoff or loading models* – determine runoff quantity and quality of pollutants,
- *receiving water models* – determine the fate and transport of pollutants in surface waters,
- *chemical fate and transport models* – a special subclass of receiving water models designed to evaluate toxic chemicals, and
- *groundwater models* – determine the fate and transport of pollutants in subsurface soils and porous media and underground aquifers.

The WERF Report is the most extensive model assessment document identified, but other researchers have published model comparisons on a more limited scale (Imhoff et al. 2003;

Borah and Bera 2004; Borah et al. 2006; Migliaccio and Srivastava 2007; Park et al. 2008). Additional models (e.g., AQUATOX (Park et al. 2008)) have been released since the WERF report. The WERF report did not consider drinking water treatment plant models, however, these were reviewed by researchers with TECHNEAU (Dudley et al. 2008). **Table 3-1** lists some of the commonly used water models, both by the acronym for which they are commonly known, and, if applicable, their full name.

Table 3-1. Prominent Water System Models.

Model Acronym	***Model Full Name***
AGNPS	Agricultural Nonpoint Source
AnnAGNPS	Annualized Agricultural Nonpoint Source
ANSWERS	Areal Nonpoint Source Watershed Environment Response Simulation
APEX	Agricultural Policy/Environmental eXtender Model
AQUATOX	-
BATHTUB	-
CE-QUAL-W2	Two-dimensional, vertical-longitudinal, hydrodynamic and water quality model
EFDC	Environmental Fluid Dynamics Code
EPANET	-
WAM/ GLEAMS	Watershed Assessment Model / Groundwater Loading Effects of Agricultural Management Systems
GWLF	Generalized Watershed Loading Functions
HSPF	Hydrological Simulation Program - Fortran
KINEROS	KINematic runoff and EROSion model
MIKE SHE	MIKE Système Hydrologique Européen (Mike 11 integrated w / ground water model)
PRMS	Precipitation-Runoff Modeling System
QUAL2K/	River and Stream Water Quality Model
Stimela (TU Delft)	-
SWAT	Soil and Water Assessment Tool
SWMM	Stormwater Management Model
WARMF	Decision Support System for Watershed Management
WASP	Water Quality Analysis Simulation Program
WEPP	Water Erosion Prediction Project

The primary focus of this research was on surface water systems, thus groundwater models were not included in the list. The processes that can be simulated using the models listed in **Table 3-1** are shown in **Figure 3-1**.

Watershed Characteristics and Processes	
AGNPS	1, 2, 4
AnnAGNPS	2, 3, 4
ANSWERS	1, 2, 3, 4
APEX	1, 2
WAM/ GLEAMS	1, 3, 4
GWLF	1, 2
HSPF	1, 2, 3, 4
KINEROS	1, 2
MIKE SHE	1, 2, 4
PRMS	1, 2
SWAT	1, 2, 3, 4
SWMM	1, 2
WARMF	1, 4
WEPP	1, 2, 3, 4

River and/or Reservoir Ecology Processes	
AGNPS	2
AnnAGNPS	2
AQUATOX	2, 5[a]
BATHTUB	1, 2
CE-QUAL-W2	1[b], 2
EFDC	1[b], 5
HSPF	1[a], 2
KINEROS	2
MIKE SHE	2
QUAL2K/QUAL2	1, 2, 5
SWAT	2
WASP	1[a], 2, 5

[a]also models fate and transport
[b]also a hydrodynamic model

Treatment Plant Processes	
EPANET	7
Stimela	6, 7

Water Distribution System Processes	
EPANET	7

[1] WERF 2001
[2] Borah et al. 2006
[3] Migliaccio and Srivastava 2007
[4] Booty and Benoy 2009
[5] Park et al. 2008
[6] Dudley et al. 2008
[7] Worm et al. 2010

Figure 3-1. Waters System Models and the Processes They Simulate.
Water system models shown with their corresponding system processes. References to literature in which the models are reviewed are also provided.

All the models in **Table 3-1** have features conducive to specific applications, the present work focuses on the *integration across models* in a functioning WatIS, and thus details of the specific models are not extensively reviewed; only models that were: (1) free, and (2) supported either by a vibrant user community or by an agency contracted to provide user support were considered for further evaluation. Models that required a significant outlay of financial resources, either to purchase the model, or to purchase support for the model were not considered for the present work but might be appropriate for other applications or other users.

3.1.1 A Watershed Model

Watershed models that focus on water quality are often used in the development of a Total Maximum Daily Load (TMDL) (USEPA 2011d). The TMDL attempts to quantify the ability of a water system to assimilate certain pollutants by estimating the amounts of pollutants that can be delivered into a water system (both point and non-point sources) and still maintain an established in-stream water quality standard. Watershed models are commonly used in developing TMDLs and in developing an understanding of how changes in watershed use (e.g., urbanization, the implementation of best management practices (BMPs)) may impact the achievability of water quality goals.

In 2004, Borah and Bera published a review of eleven watershed models including: AGNPS, AnnAGNPS, ANSWERS, ANSWERS-Continuous (an update to ANSWERS), CASC2D, Dynamic Watershed Simulation Model (DWSM), HSPF, KINEROS, MIKE SHE, PRMS, and SWAT (Borah and Bera 2004). Of these, all but CASC2D, and DWSM were also addressed in the WERF report. According to Julien et al., CASC2D simulates surface water runoff, not water quality (Julien et al. 1995). According to a 2004 conference proceeding, DWSM was being developed by the authors to simulate surface and subsurface storm water runoff, propagation of flood waves, soil erosion, and transport of sediment and agricultural chemicals in agricultural and rural watersheds (Xia et al. 2001). In 2006, Borah et al. published a follow up study focused on models used for developing TMDLs (Borah et al. 2006); this work added a consideration of the Loading Simulation Program in C++ (LSPC) model.

In 2007, Migliaccio and Srivastava reviewed agricultural watershed models, including: AnnAGNPS, ANSWERS-2000, HSPF, SWAT, WAM, and WEPP (Migliaccio and Srivastava 2007). Of these, all but WAM were previously discussed in the WERF report. The WAM website indicates that the WAM model uses GLEAMS and Everglades Agricultural Area Model (EAAMod) (USEPA 2011c).

There was no compelling evidence in the literature to suggest that a watershed model that was not discussed in the 2001 WERF report should be considered for inclusion in this assessment project; furthermore, since a model with a vibrant user community was a selection criteria, all but the models listed as being "prominent" in the 2007 Migliaccio and Srivastava review were eliminated from further consideration (Migliaccio and Srivastava 2007). The East Fork Watershed is mostly rural, and while an urban model was not necessary for this assessment project, modeling the Little Miami Watershed would require an urban land use model. According to Table 2-1 of the WERF report, all six of the models discussed in Migliaccio and Srivastava in their 2007 article can be used for rural watersheds, but according to Table 3-1 of the WERF report, only HSPF will also model urban watersheds. HSPF was also favored by researchers working on the project since it is part of the Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) group of software[1]. BASINS is essentially a geographic information system (GIS) interface that provides tools that assist the user in populating data tables needed to use models that can be accessed through the BASINS interface (USEPA 2001; Kittle et al. 2006).

HSPF was selected for further evaluation due to its popularity, its BASINS interface, and its capability to model both rural and urban watershed systems.

3.1.2 A Reservoir Model

The 2001 WERF report classified a category of models as *receiving water models*, and further subdivided the models into those that model conventional pollutants (such as pathogens, biochemical oxygen demand, dissolved oxygen, nutrients) and those that model toxic pollutants. To model Harsha Lake, a reservoir model, simulating the fate and transport of conventional pollutants is needed (with the option to model toxic pollutants). For this project, researchers were interested in exploring the links built into the BASINS software, and were also interested in working with AQUATOX. AQUATOX was not included in the WERF report, likely due to its release date (first released in 2000 (Park et al. 2009)). AQUATOX is part of the BASINS software bundle, and it can be used to model the effects of conventional and toxic pollutants. It will model flow, but it assumes that each defined segment in the waterbody is well mixed

[1] BASINS provides links to several other models. The nature of the links varies with the software. BASINS places the links under two different menu tabs (Plug-ins and Models). Models include: PLOAD, SWMM, WASP, HSPF, and AQUATOX. Plug-ins include: SWAT and WCS (BASINS 4 menu system). Additional details about the BASINS interface are provided later in this chapter.

(Clough 2009) and thus, were it reviewed by the WERF report, it would likely have been classified as a *Receiving Water Model* and as a *Chemical Fate and Transport* model. In addition to AQUATOX, WASP and HSPF can model both conventional pollutants and toxic pollutants, but only AQUATOX has the ability to model a complete aquatic system, incorporating multiple biological agents (Park et al. 2008). Park and Clough (2008) describe 13 applications of the AQUATOX model, and note that there are likely more studies underway.

AQUATOX was selected for further evaluation due to its BASINS interface, its capability to model multiple biological components, and the significant prior experience of the lead USEPA researcher on the team (Christopher Nietch, personal communication, 2010).

3.1.3 A Water Treatment Plant Model

Models have rarely been applied to the dynamic simulation of *source water quality* as it pertains to drinking water treatability. Historically, treatment plant engineers presume source water impairment and focus on in-plant operational changes or upgrades to control finished water quality using bulk approaches that target broad impairments (e.g., removing all suspended solids in order to capture microbial contaminants along with solids) rather than removal of specific contaminants. When contemplating process changes (e.g., to enhance removal of disinfection by-product (DBP) precursors), many treatment plants rely on one-time pilot plant tests to refine their procedures. These strategies, while adequate for removal of constituents of common concern, like microorganisms and suspended particles, have significant limitations when source waters contain more complex constituents that vary over time (e.g., algal taste and odor (T&O) precursors, herbicides like atrazine). These much more challenging problems, particularly as each source water has a unique set of these complexities, require a tighter coupling of source water characteristics with operational choices in the plant to produce the optimal quality finished drinking water.

Modeling these complexities requires mechanistic models of treatment plant unit operations (e.g., settling, filtration, disinfection), and integrated systems models of the complete plant to predict water quality outcomes possible under dynamic operational conditions. Models exist for specific applications in drinking water, for example, prediction of disinfection by-product speciation based on source water characteristics (Harrington et al. 1992; Williams et al. 1997; Simpson and Hayes 1998; Weinberg et al. 2002; Obolensky and Singer 2005; Obolensky et al. 2007; Obolensky and Singer 2008; Van Leeuwen et al. 2005; Rosario-Ortiz et al. 2007; Francis et al. 2009; Francis et al. 2010). Models also exist targeting specific unit operations, for example, coagulation (Edwards 1997; Tseng and Edwards 1999; Stanley et al. 2000; Volk et al. 2000; Fisher et al. 2004), and targeting specific chemical reactions, for example, those focused on organic removal for DBP precursors, taste and odor reduction, or toxicant control. Depending upon the complexity of the DWTP and the parameter being targeted, models can focus on individual unit operations or can link multiple unit operations to simulate the entire DWTP. While these individual process models are available, alternatives for modeling the drinking water treatment plant process as a whole are fairly limited. In 2008, the TECHNEAU group reviewed five water treatment plant models: OTTER, Stimela, Metrex, WTP, and WatPro. After describing each model, the authors concluded that the use of these models has been limited due to the quantity of data needed to calibrate the models and the poor performance of the models when applied outside the range of calibration (Dudley et al. 2008). In the past decade, a few new simulators have been developed, but have not been widely used.

In 2010, researchers from Delft University of Technology in the Netherlands, developed and announced the completion of a functioning simulator, named Waterspot (Worm et al. 2010). Waterspot is a drinking water treatment plant operator training tool with a SCADA-like graphical user interface; it incorporates EPANET as a functional component. Developed as a research tool for learning about the fate and transport of drinking water constituents, EPANET, and its extension EPANET-MSX (Multi-Species), are specifically designed for water distribution piping system modeling (Rossman 2000; Shang et al. 2008. In 2009, the team that developed Waterspot established an EPANET library defining elements needed to hydraulically model the DWTP (Worm et al. 2009), opening the possibility of using EPANET to model water flow, and with further development, the water quality throughout the DWTP. This would be a significant improvement to the current ad-hoc work-flow modeling method of combining mechanistic models of treatment plant unit operations (e.g., settling, filtration, disinfection) to predict water quality outcomes possible under dynamic operational conditions.

EPANET is free for download from the USEPA website, and comes with training materials, and there is an active list serve group where users can ask and respond to questions related to EPANET. Furthermore, since it appears likely that EPANET will eventually be expanded to include 'in plant' modeling capabilities, it was selected for further evaluation.

3.1.4 A Water distribution System Model

EPANET is the primary tool used for modeling water distribution systems. It enables modeling of water age, and performs trace analysis and constituent analysis, which allows various types of reaction coefficients to be used as input to the model (ASCE 2004). The first application of EPANET was in 1994 as a model to predict chlorine decay in a water distribution network in a portion of the South Central Connecticut Regional Water Authority's service area. Good agreement was achieved between the modeled results and observed chlorine levels at locations where the system hydraulics were well characterized (Rossman et al. 1994). Subsequently, EPANET has been widely used and forms the basis of a number of commercial water distribution system modeling packages (e.g., H2OMAP (Salomons 2005), PipelineNet (Samuels et al. 2003), and WaterCAD (Bentley Systems Incorporated 2009)). Version 2 of EPANET was released in 2000 (Rossman 2000), followed by an updated version, EPANET-MSX in 2006; this expanded version includes the capability of modeling more than one chemical species at a time, including bulk and surface species reactions (Shang et al. 2008), which will be particularly important for in-plant operational simulations.

The dominant use of water distribution system models is to predict hydraulic conditions in the system. Water quality prediction is infrequent due to the need for significant calibration of chemical reaction parameters in the system, but some researchers have been exploring this application. One group of researchers has suggested that water quality models can be integrated with the real-time data available through a SCADA system to more accurately predict current and future behavior of the system, and to enable interpolation of values between sparsely distributed SCADA remote terminal units (Joshi et al. 2004). Other researchers have proposed using real-time data with models: (1) to identify the location and extent of damage in a network (Shinozuka et al. 2005), (2) to confirm system design, develop operational scenarios, and train operators (Schulte and Malm 1993), and (3) to improve operational control and emergency preparedness (Joshi et al. 2004; Schulte and Malm 1993; Shinozuka et al. 2005; Tiburce et al. 1999).

In the current work, modeling the distribution system was not specifically explored; however, EPANET was evaluated for potential inclusion into the WatIS as a DWTP simulator.

3.2 Experience Working with the Models

As described above, HSPF, AQUATOX, and EPANET were evaluated for potential use in the WatIS. Each of these models is discussed in detail below, followed by a discussion on model integration methods.

3.2.1 Modeling with HSPF

HSPF was used to model changes in the watershed. HSPF is a set of computer codes designed to simulate hydrologic systems, including water quality. It is specifically intended to allow consideration of impervious surfaces (e.g., urban landscape features like parking lots), pervious surfaces (e.g., rural features like fields), and well-mixed water bodies (Bicknell et al. 2001). The Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) system provides the user with a graphical interface for working with watersheds and watershed data. BASINS 4.0 is built as an extension to MapWindow, an open-source, non-proprietary GIS (AQUA TERRA Consultants 2011). BASINS is often used as a front-end for new HSPF projects (USEPA 2001; Johnson 2005). BASINS provides the user with tools that help populate the data tables/files needed to use models that can be accessed through the BASINS menu system; these include: PLOAD, WASP, SWMM, HSPF, SWAT, and AQUATOX (Duda et al. 2011)). With the tools that are in BASINS, several kinds of data can be accessed from sources on the internet and from sources that come packaged with the BASINS software. Additionally, there are a number of standard GIS features available in BASINS that can be used to add information from, and export information to other GIS applications. BASINS also provides techniques for analyzing landscape information and displaying geographic relationships. It is a very useful tool for the models it draws upon. Many of BASINS features were used in this assessment project for calculating physical parameters of the watershed (watershed boundaries, land slopes, etc.), and as an interface for working with HSPF. The sources of data used in this project are summarized in **Table 3-2**.

Table 3-2. Data Requirements for HSPF Watershed Model.

Model	*Data*	*How Loaded Into the HSPF Model*	*Data Type*
BASINS	*GIS Background (map)* The background map consists of political boundaries (states), hydrologic unit codes (HUC) 8, and a stream layer. When BASINS is not used as the interface, the HUC data can be downloaded from the USGS[2], and state and stream layers can be obtained from Environmental Systems Research Institute (ESRI)[3].	Automatically loaded when the user generates a new project	measured
BASINS	*Land Use data* There are multiple sources of land use data. BASINS uses the USEPA Geographic Information Retrieval and Analysis System (GIRAS) land use/land cover spatial data[4]. An alternative data source is the USGS National Land Cover Database (NLCD)[5]. The United States Department of Agriculture, National Agricultural Statistics Service, Research and Development Division, Geospatial Information Branch, Spatial Analysis Research Section also makes available a Cropland Data Layer[6]. Local land use data may also be available from those working in the region.	Used BASINS tools: File->Data Download	measured
BASINS	*National Hydrography Dataset* (locations of lakes, ponds, streams, rivers, canals, dams and stream gages) BASINS uses the USGS dataset[7], but other sources of this information are likely available (for example, from ESRI or from local sources).	Used BASINS tools: File->Data Download	measured
BASINS	*Census Data (zip codes, counties, etc.)* Can be downloaded from a variety of locations, including the United States Census Bureau[8].	Used BASINS tools: File->Data Download	measured

[2] See http://water.usgs.gov/GIS/huc.html or http://datagateway.nrcs.usda.gov/GDGOrder.aspx?order=QuickState for download information.
[3] See http://www.arcgis.com/home/group.html?owner=esri&title=ESRI%20Maps%20and%20Data for more information on layers ESRI.
[4] See http://water.epa.gov/scitech/datait/models/basins/metadata_giras.cfm for download information.
[5] See http://datagateway.nrcs.usda.gov/GDGOrder.aspx?order=QuickState for download information.
[6] See http://datagateway.nrcs.usda.gov/GDGOrder.aspx?order=QuickState for download information.
[7] See http://nhd.usgs.gov/ for download information.
[8] See http://www.census.gov/ for download information.

Model	Data	How Loaded Into the HSPF Model	Data Type
BASINS	*Digital Elevation Model (DEM) Grid Data* There are DEM files and National Elevation Database (NED) files. Both are from the USGS but are processed a bit differently. There are several sites were DEM data can be downloaded[9]. The files are available in different resolutions (usually 3, 10, and/or 30 meter).	Used BASINS tools: File->Data Download	measured
BASINS	*Meteorological Data (precipitation, temperature, potential evaporation)* There are a variety of sources were these data (precipitation and temperature) can be obtained. The quality of the data may not be well understood so care must be taken when downloading. For this project, data from NOAA stations were used[10].	Used BASINS tools: File->Data Download	measured
BASINS	*NWIS Daily Discharge Stations (flow measuring stations, daily discharge)* Can be downloaded from the USGS[11].	Used BASINS tools: File->Data Download	measured
BASINS	*Ohio HUC12 Boundaries (for display)* Can be downloaded from a variety of locations[12].	Used BASINS tools: View->Add Layer	measured
BASINS	*East Fork Watershed Boundary (to select target study area)* Obtained by dissolving borders of selected HUC12 watersheds.	Used BASINS tools: View->Add Layer	measured
HSPF	*Sediment Parameters and Sediment Loadings (to model sediment loads)* These data must come from a local researcher.	Used WinHSPF entry screens	synthetic
HSPF	*Stream Geometry* These data came from BASINS.	Can be modified using WinHSPF entry screens	synthetic

[9] See http://data.geocomm.com/dem/demdownload.html, http://datagateway.nrcs.usda.gov/GDGOrder.aspx?order=QuickState, http://gis1.oit.ohio.gov/geodatadownload/osip.aspx for download information. Also see http://www.petroleumgeographics.com/faq.shtml#seven for information.

[10] See http://lwf.ncdc.noaa.gov/oa/climate/climatedata.html#notes for download information. May also see http://ars.usda.gov/Research/docs.htm?docid=19388 or http://www.ncdc.noaa.gov/oa/ncdc.html (and find a station) for more information. NEXRAD data should also be considered.

[11] See http://waterdata.usgs.gov/nwis for download information.

[12] For this project, http://www.oh.nrcs.usda.gov/technical/12-digit/download.html was used for data download.

Model	*Data*	*How Loaded Into the HSPF Model*	*Data Type*
HSPF	*Atmospheric Data (solar radiation, cloud cover, wind, dew point temperature)* A source for these data has not yet been identified. National Climatic Data Center (NCDC) or National Oceanic and Atmospheric Administration (NOAA) may have information.	Used WinHSPF tool that uses scripts to import from text files	synthetic
HSPF	*Nutrient Loadings* These data must come from a local researcher.	Used WinHSPF entry screens	synthetic
HSPF	*Various Coefficients and Parameters* These data must come from a local researcher.	Used WinHSPF entry screens	synthetic
HSPF	*Point Source Loading (synthetic sediment point source data added to explore the point source feature of the model)* These data must come from a local researcher. Some relevant data may be available from Envirofacts[13].	Used WinHSPF entry screens	synthetic
-	*Soils Data* While not used explicitly in HSPF, other watershed models require soil geospatial data. Usually either Natural Resources Conservation Service, United States Department of Agriculture - Soil Survey Geographic (SSURGO) Data or Natural Resources Conservation Service, United States Department of Agriculture - U.S. General Soil Map (STATSGO2) data[14].		measured

Chapter 3 of a report published by the United States Department of Energy provides an extensive list of data available for hydrologic modeling (Whelan et al. 2009). This reference includes many of the data sources listed in **Table 3-2**, plus many others. **Table 3-2** describes the data needed for HSPF and indicates how these data are added to the HSPF model (by using either the BASINS interface and/or the WinHSPF tool). It should be noted that states, counties, and other local agencies sometimes distribute data for their area of interest. For example, Pennsylvania maintains the Pennsylvania Spatial Data Access (PASDA) website, which serves as the public access geospatial information clearinghouse for the Commonwealth of Pennsylvania (http://www.pasda.psu.edu/).

The *Data Type* for this project, as explained in **Section 2.4** is also shown in the table. In a fully functional model, all data would need to be measured or modeled. More details regarding the requirements for HSPF modeling are provided in the HSPF User's Manual (Bicknell et al. 2001); a summary is provided below.

For this project, first the BASINS program was launched and a new project was built. The Little Miami Watershed in Ohio was chosen (hydrologic unit code (HUC) 8: 05090202), and reference

[13]See the Envirofacts page http://www.epa.gov/enviro/html/pcs/adhoc.html for download information.
[14] See http://soildatamart.nrcs.usda.gov/ or http://datagateway.nrcs.usda.gov/GDGOrder.aspx?order=QuickState for download information.

spatial zones were selected. Then, six data sources were added using the *File->Download* feature (as indicated in the table above), and two boundary layers were added using the 'add layer' feature (layers shown in the table above). The map displayed in the BASINS GIS interface after the data have been added is shown in **Figure 3-2**.

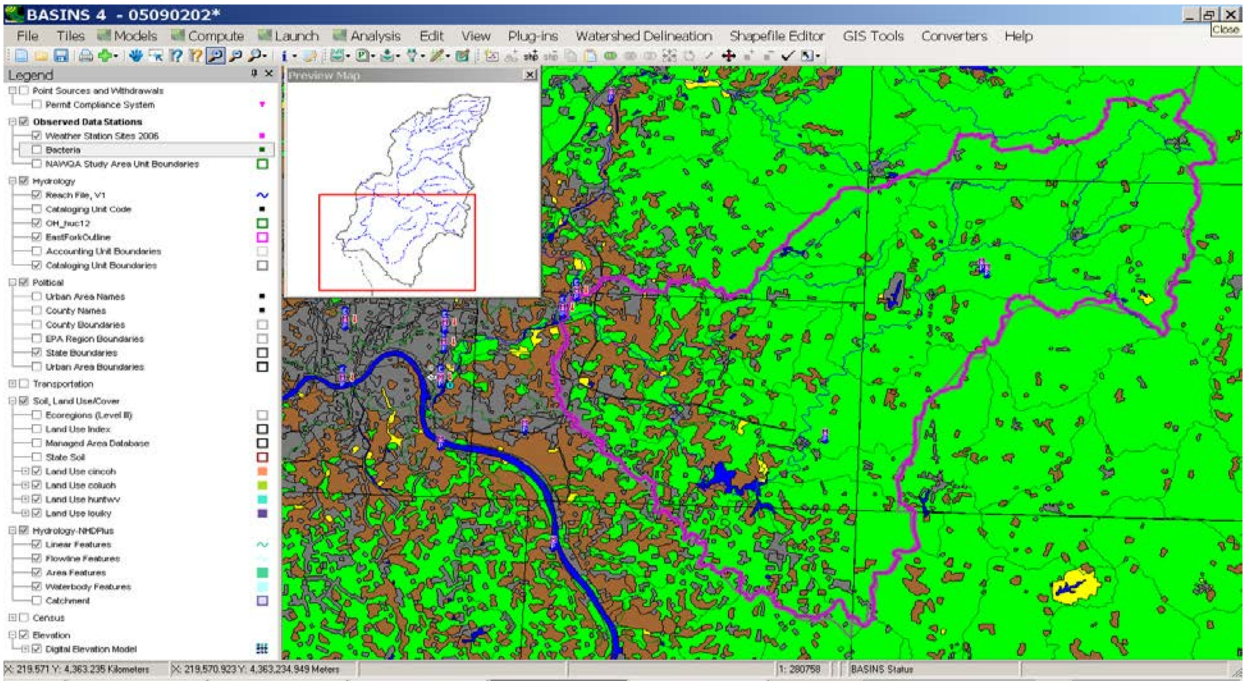

Figure 3-2. BASINS GIS Map of the East Fork Watershed with Physiogeographic Data. The *Preview Map* window (in the upper left corner of the larger/main map window) shows, using a red rectangle, the portion of the Little Miami Watershed that is displayed in the main map window. In the main map window, the pink outline is the East Fork Watershed. The HUC12 watershed outlines are narrow dark green lines. The Ohio River is the dark blue line in the lower left hand corner of the image. Harsha Lake is the blue patch near the bottom center of the map.

Three additional layers are required in BASINS before an HSPF project can be generated: a *subbasins* layer, a *streams* layer, and an *outlets* layer. These layers can be generated manually or automatically using the BASINS watershed delineation tool. **Figure 3-3** shows the BASINS screen after the EFW has been divided into subwatersheds using the BASINS automatic watershed delineation tool.

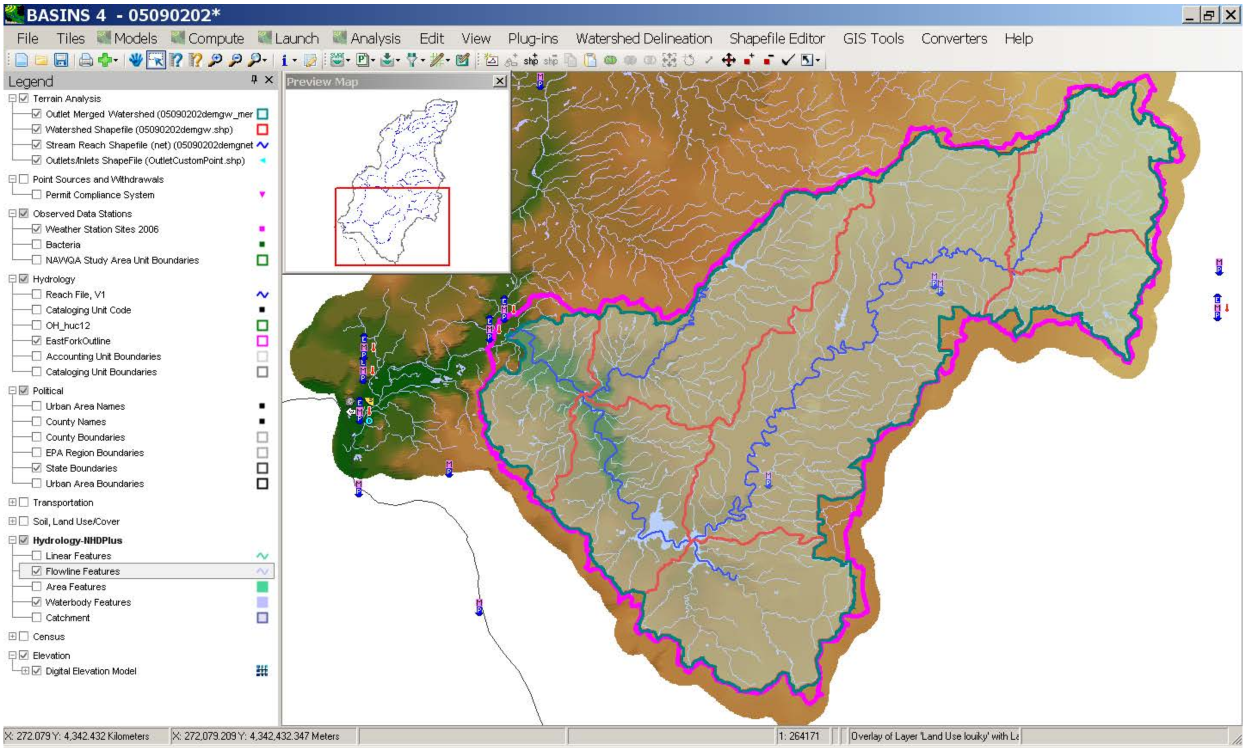

Figure 3-3. Subbasins Generated Using Automatic Watershed Delineation in BASINS. The seven subwatersheds are delineated with solid red borders. The river reaches are shown in blue in the interior of the subwatersheds.

The automatic delineation tool provides the option to select a threshold for the area of each subwatershed; decreasing the size of the threshold area will increase the number of subwatersheds that are automatically generated. The BASINS tool uses the DEM information to determine the boundaries of the subwatersheds. If more control over the watershed delineation is needed, subwatersheds can be manually delineated. This allows construction of subwatersheds in the model such that their outflow locations match existing field monitoring locations (needed for calibration), or match locations that are points of transition between a stream and a reservoir (needed for modeling in AQUATOX).

BASINS stores project information in four main files (Duda et al. 2001). When an HSPF project is created, the BASINS interface transfers information into HSPF and then the BASINS files are no longer needed. When a new HSPF project is opened, an HSPF User Control Input (*uci*) file is generated using values estimated from the information contained in the corresponding BASINS project. The *uci* file is a text file, and can be viewed or edited with a simple text editor. A *wdm* file is also created. The *wdm* file is not a text file, it is a binary direct-access file. The file is used to hold time series data (both for storing the point source inputs and time series outputs of the HSPF simulation).

For this project, HSPF was used to model loadings from the watershed into streams and rivers, and those loadings were then transferred into AQUATOX as input into riverine segments that drain into Harsha Lake. When HSPF is called from BASINS, it opens WinHSPF, a graphical user interface for HSPF. The main window of WinHSPF, including the schematic of the watershed that is automatically generated by calling HSPF from BASINS, is shown on the right side of **Figure 3-4**.

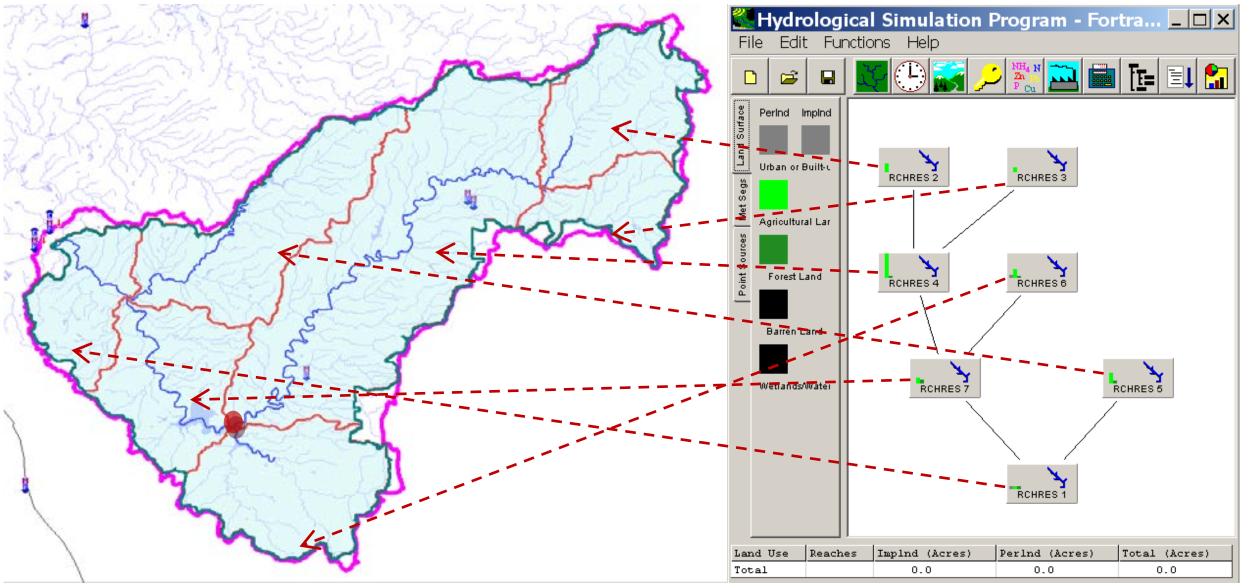

Figure 3-4. Schematic of the East Fork in BASINS and Corresponding Reaches in HSPF. The East Fork Watershed with BASINS subwatersheds is shown on the left. The WinHSPF main window with modeled river reaches is shown on the right.

The left side of the figure shows the subwatersheds of the EFW as defined in BASINS. The relationships between the subwatersheds and the model segments are shown with red dashed lines. As mentioned above, the shape, size, and outlet locations of the subwatersheds can be controlled using the BASINS manual watershed delineation tools.

Once the HSPF project has been generated and a time period for a simulation selected, the model is ready to execute. The results of the simulation of flow at the outlet of Reach 4 (watershed outlet shown in **Figure 3-4** as a large red dot) are shown in **Figure 3-5**.

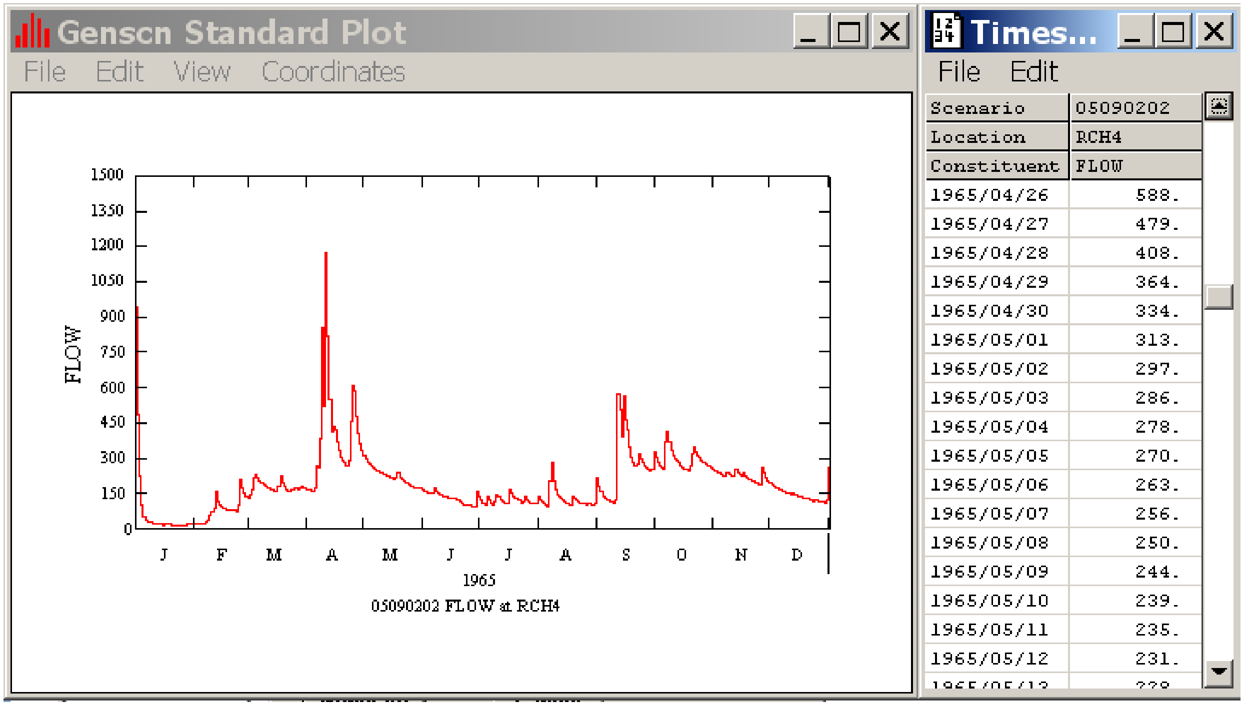

Figure 3-5. Results of the Flow Simulation at the Outlet of Reach 4 Shown Using GenScn. The reporting feature in HSPF, GenScn, will show data graphically (left) and/or as a time series (right). Simulation results show two peak flow events in 1965, one in April and the other in September; flow is graphed in ft^3/sec.

Figure 3-5 shows the predicted flow at the subwatershed outlet for the watershed conditions selected. A full year is simulated in this example, with two high flow events predicted (in April and in September). To run simulations other than flow, additional input is required. For this project, most of these loadings were entered manually, using the WinHSPF menu screens. HSPF also provides a means of loading time series data a batch at a time using import scripts. This method was used when loading atmospheric data.

Sediment data were added to allow for the transport of nutrients and other chemicals with the suspended sediment. The HSPF modules that simulate sediment erosion and delivery from the landscape and in-stream transport require the input of several coefficients and parameters, along with the initial distribution of silt, sand, and clay in both the water column and the sediment bed; synthetic data were added to the HSPF model to allow simulation of in-stream transport. Water quality parameters are referred to by HSPF as “pollutants”. To simulate pollutants, HSPF requires that atmospheric data (solar radiation, cloud cover, wind, dew point temperature) be included in HSPF; synthetic atmospheric data were added to the HSPF model.

For this project, the following “pollutant” terms were added to the HSPF model: ammonia, nitrate, orthophosphate, biological oxygen demand, and dissolved oxygen. An example of the output of the simulation, for ammonia and nitrate at Reach 4 is shown in **Figure 3-6**.

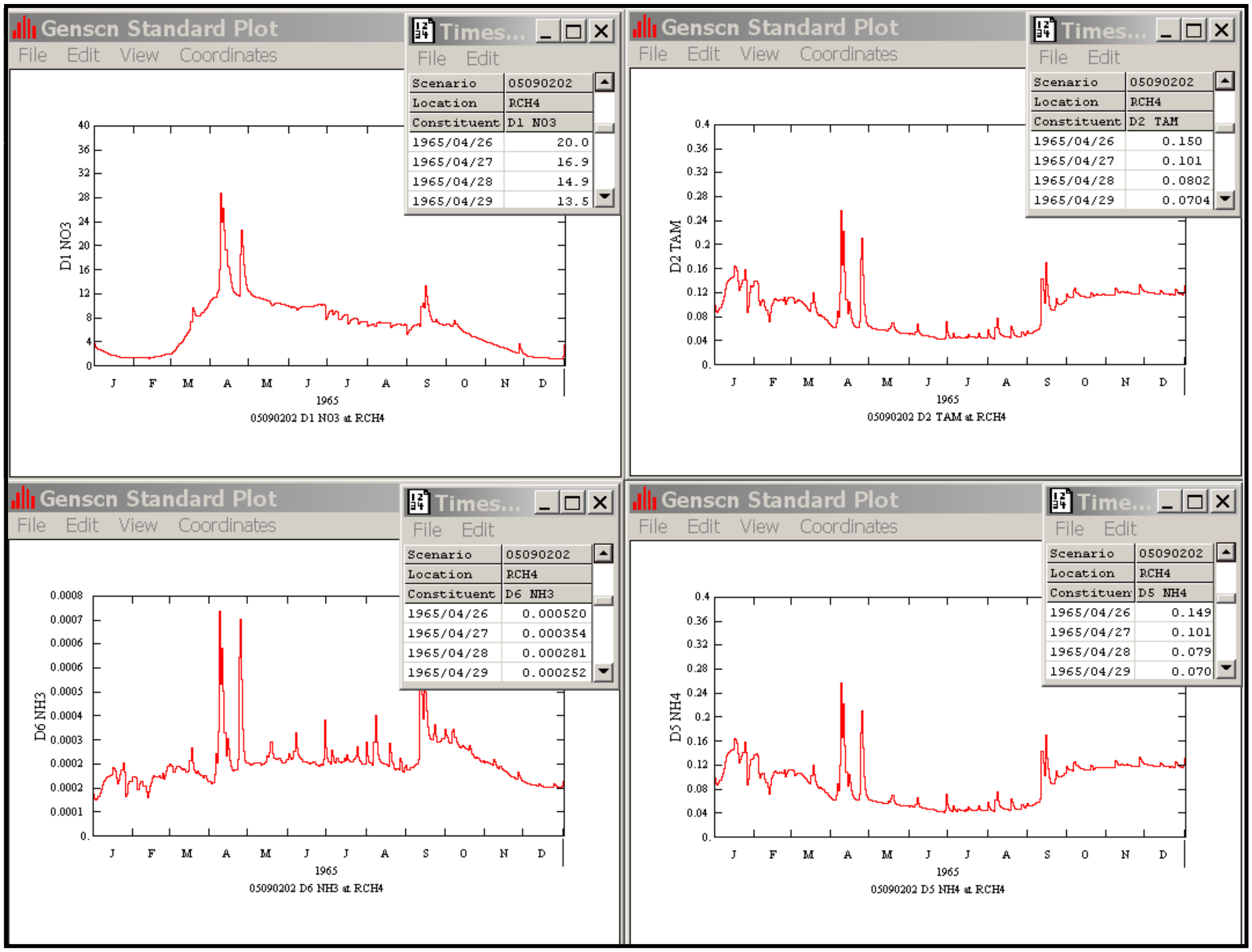

Figure 3-6. Simulated forms of Nitrogen Modeled using HSPF at the Outflow of Reach 4. Time series results of an HSPF simulation. Nitrate is shown in the top left box, total ammonia in the top right. The components of total nitrogen (NH3 and $NH4^{+}$) are shown in the bottom of the figure (left and right respectively). TAM = total ammonia concentration in mg N/L; sum of $NH4^{+}$ and NH3. This simulation was performed using both measured and synthetic data.

The figure shows that in 1965 there were a few spikes of nitrogen in April, then another in September. It should be noted that point sources of pollutants (including sediment sources) can be added into the HSPF model. For this project, synthetic sediment point source data were added. Effluent from a wastewater treatment plant would be considered a point source, and could be added into the HSPF model if desired.

The loadings at the outlet of the reaches simulated using HSPF can be exported to text files. In addition, the WinHSPF interface provides a means of exporting some simulated results for a river reach directly into an AQUATOX segment. This feature was explored as part of this project. More of the details on linking HSPF and AQUATOX are provided in the User's Manual (Clough 2005) and in **Section 3.4** below.

3.2.2 Modeling with AQUATOX

AQUATOX models the fate of organic chemicals, nutrients, and other pollutants in an aquatic ecosystem (Park et al. 2008; Clough 2009) and was used in this project to simulate changes in the water quantity and water quality in a reservoir as result of changes in the upstream watershed. To provide the framework for discussing the modeling process, Harsha Lake was used as a demonstration site. A top view schematic of Harsha Lake is shown in **Figure 3-7.**

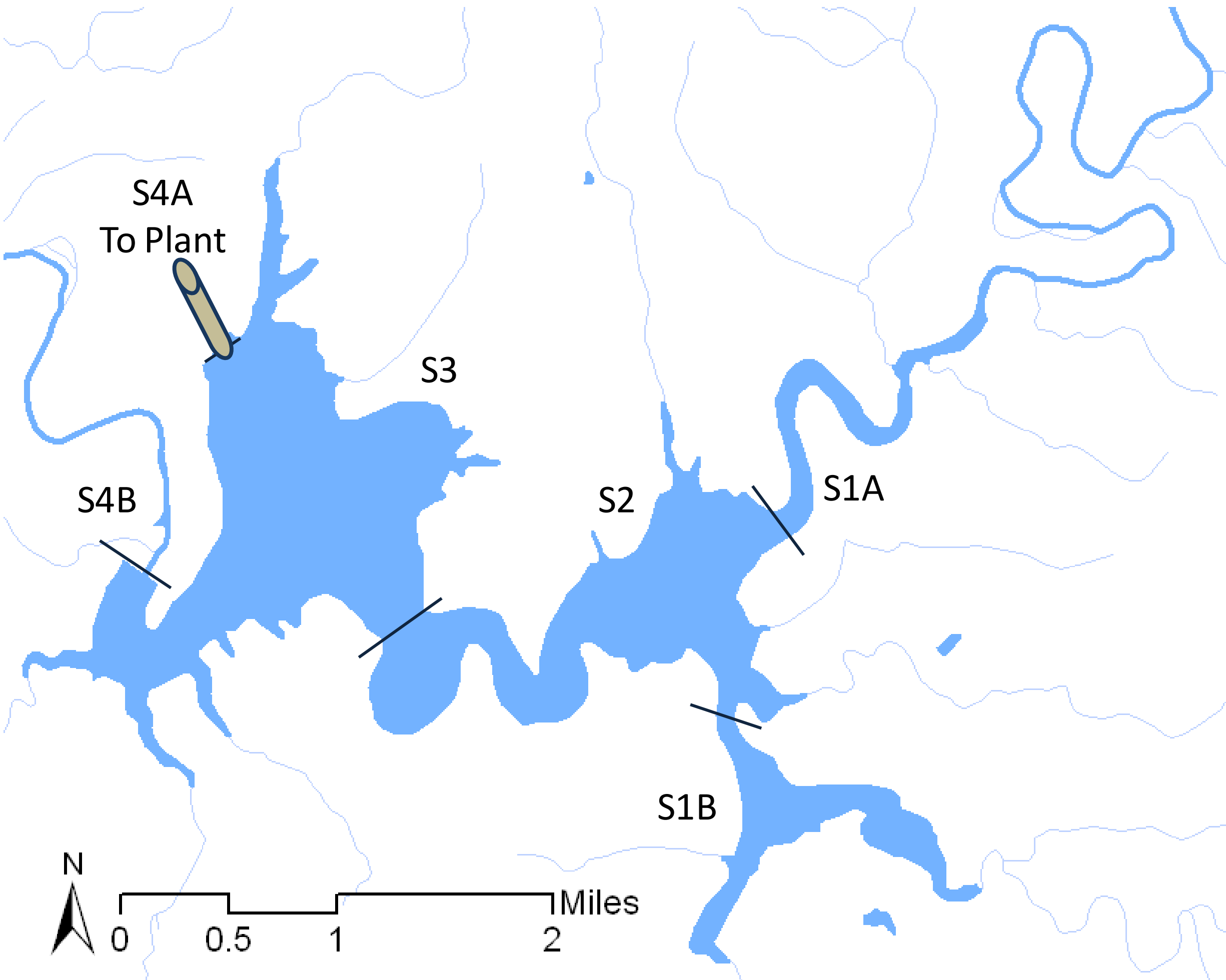

Figure 3-7. Harsha Lake with the Sections Used in AQUATOX Modeling Indicated. The lake is divided into six sections, with one section (S4A) representing the outflow to the drinking water treatment plant.

Using AQUATOX, a waterbody can be analyzed as a whole, or as a network of linked segments. When modeling a lake, segments can be linked into vertically stratified pairs, with an upper and a lower segment, simulating the epilimnion and hypolimnion. For this project, Harsha Lake was modeled using ten segments, with two representing hypolimnion segments, as indicated on the map in **Figure 3-7,** and in the schematic representation in **Figure 3-8**.

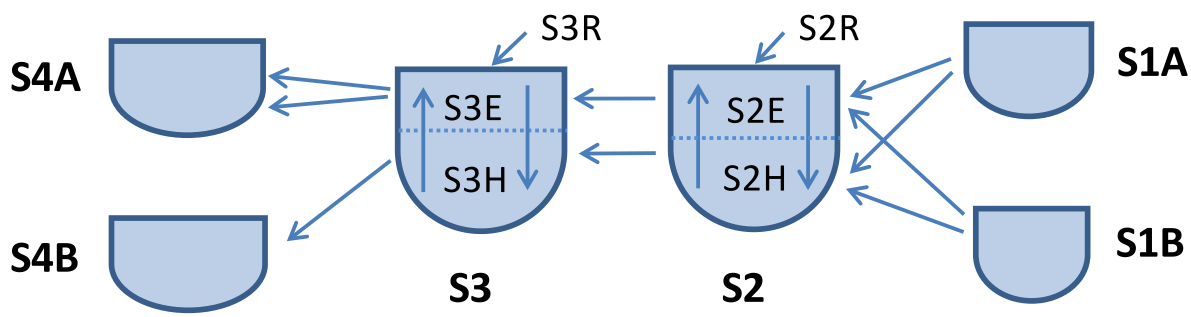

Figure 3-8. Schematic of Harsha Lake Labeled with Segments of the AQUATOX Model. The six sections are modeled using ten segments (two of the sections are stratified pairs). One section (S4A) represents the outflow to the drinking water treatment plant. Two segments, S3R and S2R represent runoff directly into the lake.

Harsha Lake was modeled as having two inflow riverine segments (S1A and S1B), two lake pools (2 and 3), each with an epilimnion (S2E and S3E) and a hypolimnion (S2H and S3H) segment, and two outflow segments (4A and 4B). Surface water segments of pools 2 and 3 (S2E and S3E) allow for watershed runoff directly into the lake (S2R, and S3R). The Harsha Lake segments were designed such that S1A corresponds to the HSPF Reach 4, and S1B corresponds to the HSPF Reach 6. This segment pattern was chosen due to the shape of the reservoir, and also due to the distribution of the sample collection locations from within the watershed.

There are 15 linking relationships in the model; these are shown with arrows indicating the direction of flow in **Figure 3-8**. The main AQUATOX window, showing the segments included in the model is shown in **Figure 3-9**.

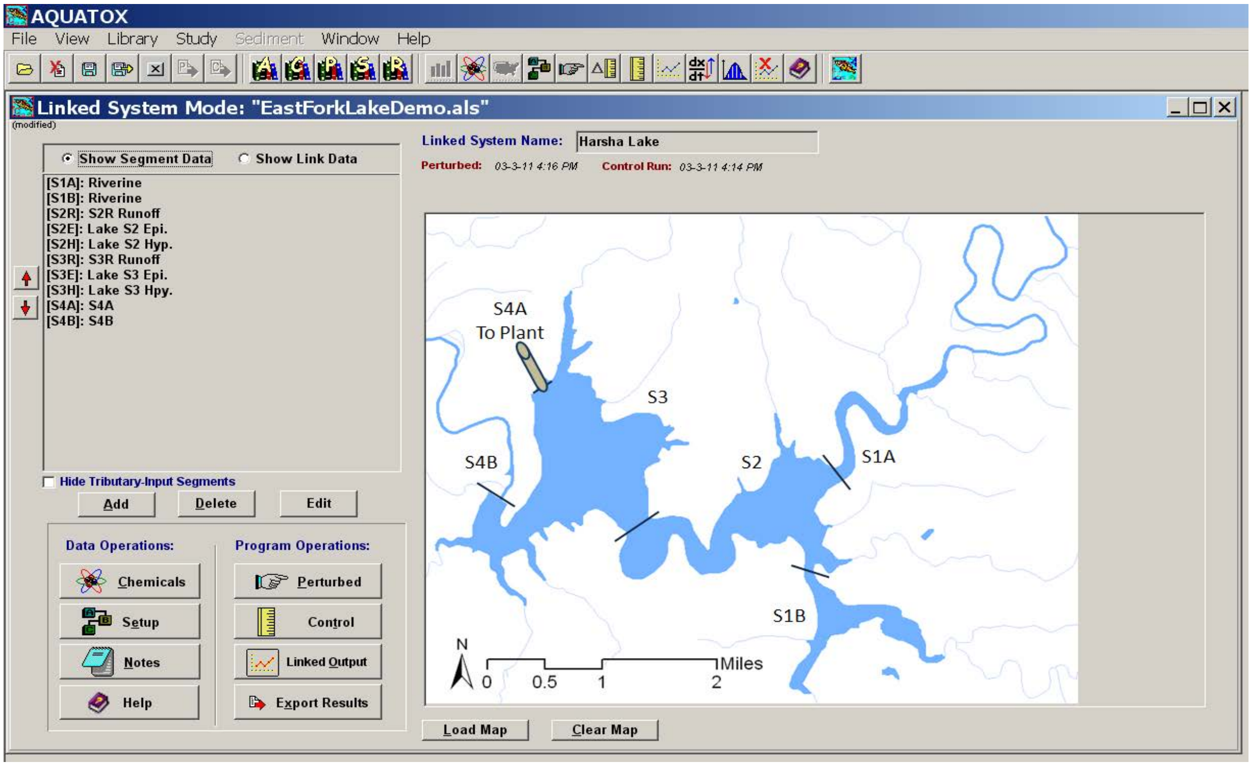

Figure 3-9. The Main AQUATOX Window Showing Segment List and Lake Schematic. The left side of the AQUATOX screen shows the segments used in the AQUATOX model. Note that there are ten; two represent direct runoff into the lake.

There are three primary types of data that must be entered into AQUATOX to run a simulation: site information, initial conditions, and loadings. When running AQUATOX on the waterbody as a whole, only one set of these three types of data is required. When modeling using linked segments, a set of these three types of data is required for each segment. These data requirements are summarized in **Table 3-3**.

Table 3-1. Data Requirements for AQUATOX Waterbody Model.

Model	***Data***	***How Loaded Into Model***	***Data Type***
AQUATOX	*Waterbody Physiogeographic Information* The surface area of waterbodies can be obtained from the National Hydrography Dataset, but the depth (and thus the volume) must be obtained from local researchers (via a bathymetry survey).	For each segment, click the Site button from the main window	synthetic
AQUATOX	*Initial Conditions for all State/Driving Variables* This information must be measured or estimated by local researchers.	For each segment, click the Initial Conditions button from the main window	synthetic

Model	Data	How Loaded Into Model	Data Type
AQUATOX	*Loadings associated to State/Driving Variables* This information must be measured or estimated by local researchers.	Double click on the State/Driving variables in the list – loadings can be uploaded from an Excel file by clicking “change” in the loading screen	synthetic
AQUATOX	*Linking Relationships (flow loadings between segments)* This information must be measured or estimated by local researchers.	From the main window, click on Show Link Data and double click the segment of interest	synthetic
AQUATOX	*Sediment bed data (when needed)* This information must be measure or estimated by local researchers.	Entered by clicking on the Sediment Layer(s) button inside the Segment menu	synthetic

The table describes data needed for AQUATOX modeling and also explains how these data are added to the model. The *Data Type* for this project, as explained in **Section 2.4** is also shown in the table. In a fully functional model, all data would need to be measured or modeled. More details regarding the requirements for AQUATOX modeling are provided in the AQUATOX User’s Manual (Clough 2009).

The site information required includes the volume, length, surface area, and depths (maximum and mean) of the water, temperature ranges for the air and water, light and wind data, and various coefficients. When site specific coefficients are not available, defaults provided in the model can be used; however, this will reduce the applicability of the resulting predictions. More specific information on entering site data is provided in the AQUATOX User’s Manual, along with information on the defaults that can be used when site specific information is not available (Clough 2009).

For each segment, an initial condition for each State/Driving variable included in the study must be provided. The State/Driving variables included in this AQUATOX model are shown on the right side of **Figure 3-10**.

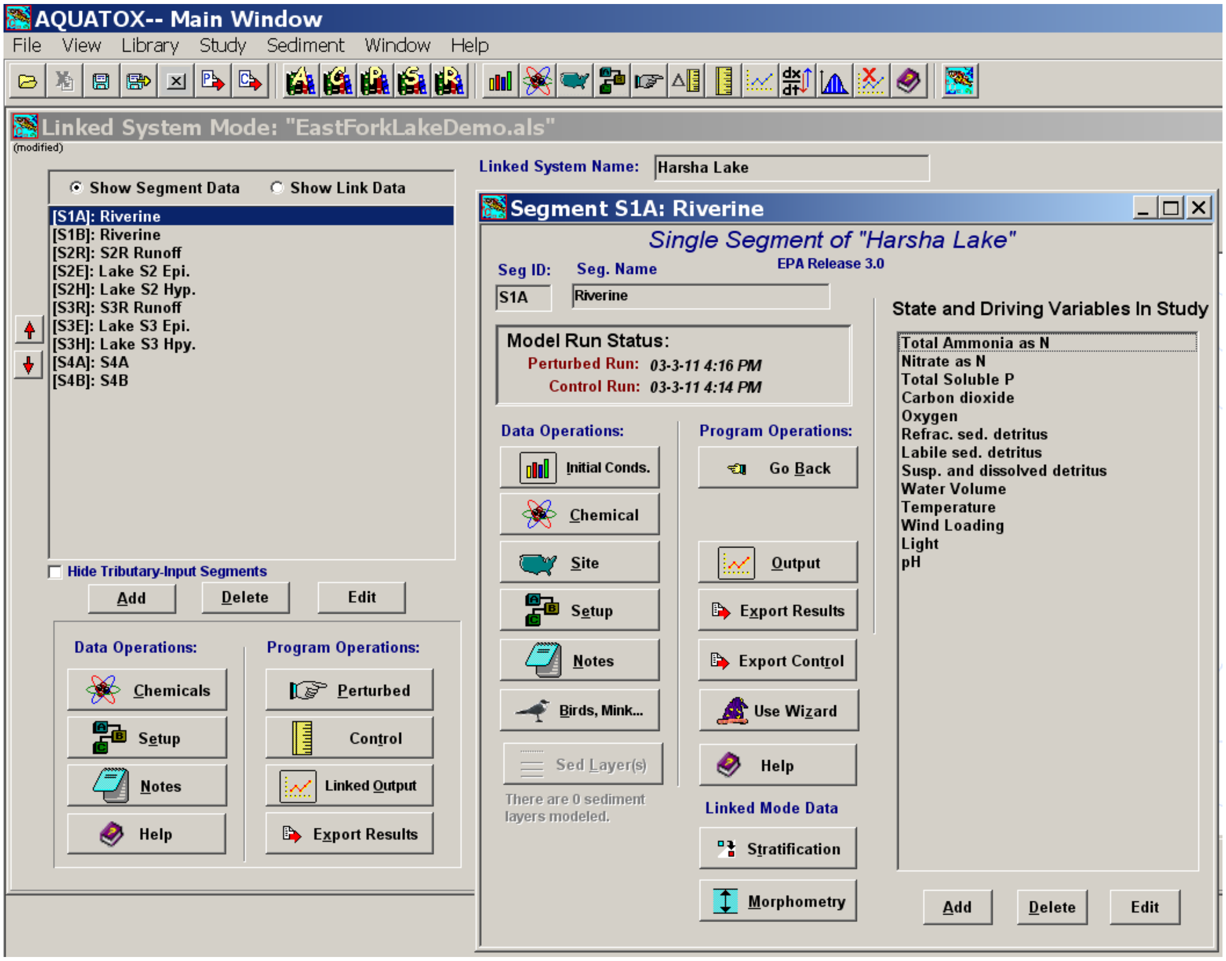

Figure 3-10. AQUATOX Screen Showing the State/Driving Variables Used in the Study. State and Driving Variables are shown on the right. This model uses 13 variables; more can be added as needed.

State/Driving variables can be added and removed from the model to meet the needs of the simulation (e.g., Chlorophyll A, toxicants) (Clough 2009). When used in the model, loadings for each of the State/Driving variables must be added. More specific guidance on adding initial conditions and loadings is provided in the AQUATOX User's Manual (Clough 2009).

When a linked segment model is used, in addition to site information, initial conditions, and loadings must be provided for each individual segment, and the relationship between the links (the exchange of water flow) must also be defined. AQUATOX provides a menu screen where the flow between the segments can be entered. Due to the relationship between the segments in a linked model, the state and driving variables included in each segment of a linked model must be the same.

In the Harsha Lake system, it is hypothesized that the lake bottom is a seasonal store for nutrients. When the sediment is to be used as a sink and/or source for pollutants, the sediment diagenesis feature of AQUATOX must be used. To use this feature, sediment bed data are required.

Chemicals, such as pesticides can be modeled in AQUATOX provided that they are listed as a State/Driving Variables, and the relevant parameters, initial concentrations, and loadings have been added. More on the specifics regarding the requirements for AQUATOX modeling can be found in the User's Manual (Clough 2009).

The developers of AQUATOX provided multiple ways of entering data into the model. In addition to manually entering data using the menu system, AQUATOX has the ability to accept some input directly from specifically formatted Excel files. Examples of the input files are provided with the download of the AQUATOX model. These files are discussed in more detail in the User's Manual (Clough 2009). AQUATOX will also accept data directly from WinHSPF. WinHSPF can export the information for a riverine reach out of WinHSPF and into an AQUATOX segment. It is a one-to-one transfer, and data defining the individual segment must then be transferred into the multi-segment AQUATOX model.

The output from an AQUATOX simulation includes time series flow and loadings to/from the defined segments. The developers provided two ways of exporting data from AQUATOX. For documentation purposes, the user can download a complete record of the model simulation to a text file. Results can also be exported to an Excel file. This file contains the time series loadings needed for input into the next model in the WatIS. It should be noted that, while AQUATOX can import data from Excel files, and export data to Excel files, the formats of the import and export files are not the same; this complicates sending data into and out of AQUATOX for communication with upstream and downstream models.

There is some degree of flexibility regarding the time step used in the AQUATOX model. The time step can be an hour, a day, or fractions of either. For this project, a daily time step was selected and a 24 day simulation was performed. More on the details of working with AQUATOX can be found in the User's Manual (Clough 2009).

3.2.3 Modeling with EPANET

EPANET was evaluated for inclusion into the WatIS; however, extensive modeling with EPANET was not undertaken. EPANET is intended to simulate water hydraulic behavior and water quality in a pressurized pipe water distribution system network. EPANET comes with a library of components that are found in a pipe network, including pipes, pumps, storage tanks, nodes (pipe junctions), valves, and reservoirs. Using the EPANET user interface or the Programmers' Workbench, these components can be added to an EPANET project. When the EPANET simulation is performed, the software predicts water flow in each pipe, water pressure at each node, the height of the water in each tank, the age of the water in the system, and, if the water quality parameter is included in the simulation, the concentration of a chemical species (USEPA 2011a). EPANET can predict the behavior of a non-reactive tracer over time as it travels through the pipe network, or it can track the fate of a reactive material as it grows or decays over time, provided reaction kinetic terms are included in the input file (Rossman 2000).

EPANET was used as a place holder for an actual model of the water treatment plant for two reasons: (1) no more viable alternative could be identified, and (2) EPANET may be expanded to model the drinking water treatment plant at some point in the future (Worm et al. 2009). The processes used in the treatment train of the Bob McEwen Water Treatment Plant, which draws its source water from Harsha Lake, are shown in **Figure 3-11**.

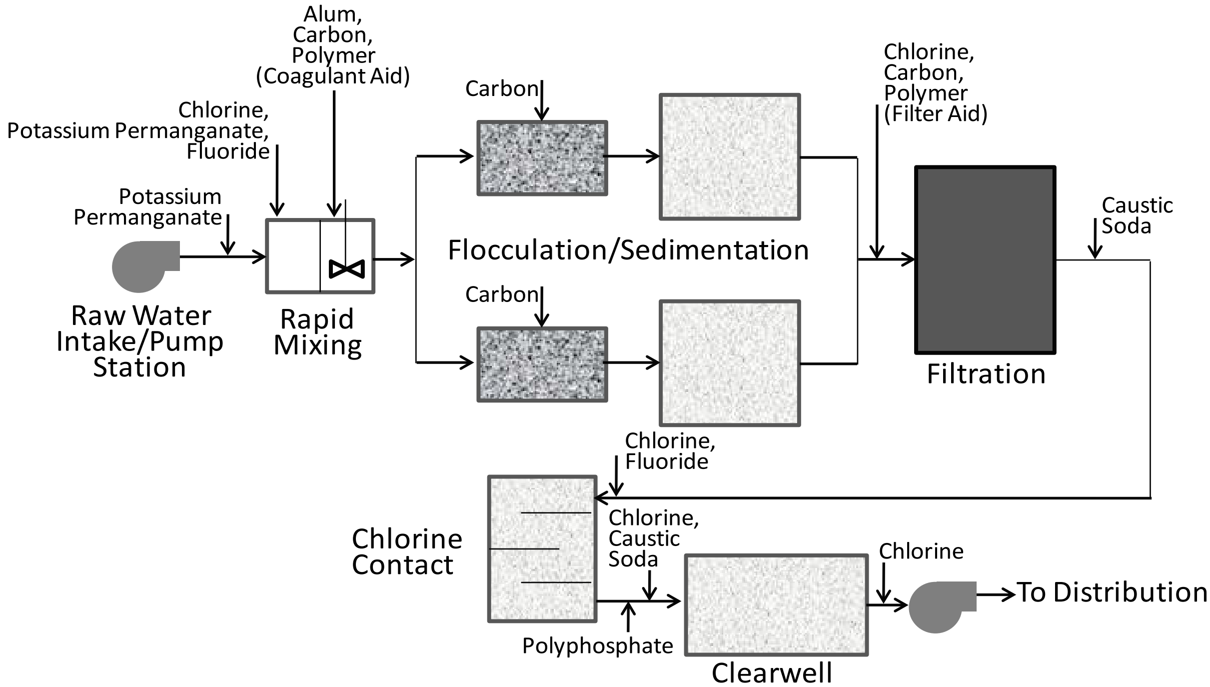

Figure 3-11. Treatment Process Schematic of Bob McEwen Water Treatment Plant. Figure shows treatment processes, from intake to distribution. Unit operations at the plant include: pre-oxidation (potassium permanganate and chlorine addition), coagulation and rapid mix, flocculation and sedimentation, carbon filtration, sand filtration, primary disinfection and secondary disinfection.

Raw water quality is characterized by measuring total organic carbon (TOC), dissolved organic carbon (DOC), natural organic matter (NOM), bromide concentration, total dissolved solids (TDS), pH, conductivity, alkalinity, manganese and iron concentrations, and atrazine concentrations.

For this project, with the exception of the pumps, the EPANET library feature termed a *reservoir* was used to stand in for the actual treatment unit processes. The EPANET network representing the treatment plant is shown in the *Network Map* window on the left of **Figure 3-12**.

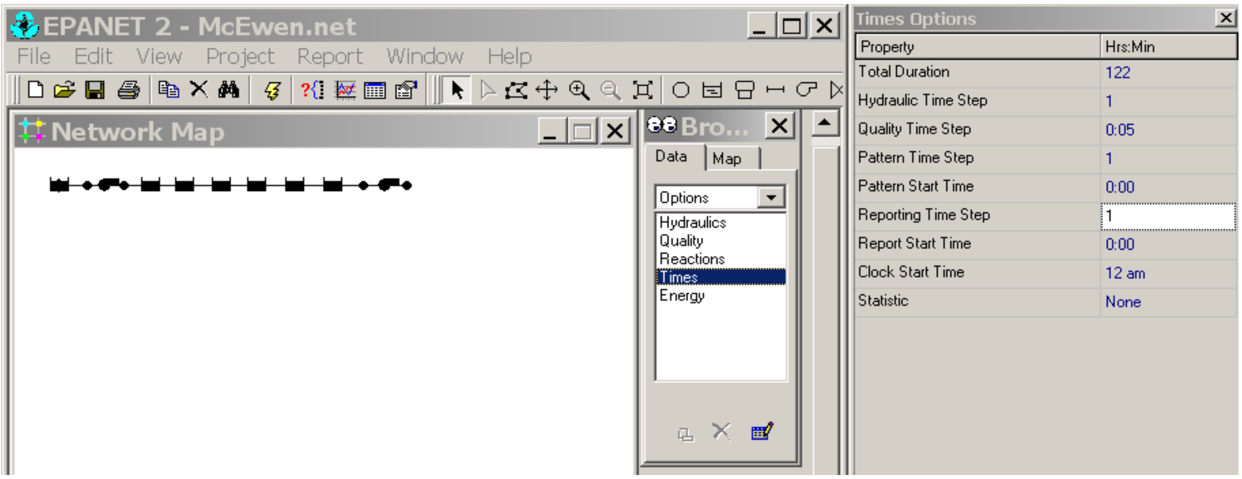

Figure 3-12. EPANET Network Representing the Drinking Water Treatment Plant. The network is shown on the left. The time step options window is shown on the right. In this simulation, a pattern, hydraulic, and reporting time step of 1 is selected.

The right side of the figure shows the menu screen where the time step used in the simulation can be adjusted. When the number of *Time Periods* selected in the *Pattern Editor* is set to 24, a *Pattern Time Step* of *1* means that the pattern will be applied for each hour of a 24 hour day. EPANET does not accept time series data as it is exported from HSPF and AQUATOX. For loading output from these models to be used as input into EPANET, it will need to be converted into an average and a pattern. This data processing step highlights the differences in the approach to data management and model development in natural and engineered systems.

In addition to being able to generate a project and execute the model using the user interface and menu system, EPANET can be executed directly from DOS. To use this feature, the network input data must be stored in a specifically formatted text file. Results from the model will be sent directly to a text file. The EPANET User's Manual indicates that the text file exported from EPANET can be read back into EPANET. This is a very useful feature that allows the user to make changes to model parameters, run the model, save the results, then make changes directly in the text file, then import the file back into EPANET and rerun the revised project; it can also be used to run multiple simulations in series.

While EPANET was used only as a stand-in for a DWTP model, evaluating the model for use in the WatIS highlighted some of the challenges that must be addressed when attempting to bridge the gap between natural and engineered water systems.

3.3 Overview of Model Integration

While the specifics of the data structures and models described above are critically essential, they are insufficient to enable decision-making across the full space from watershed to DWTP. Rather, it is necessary to develop methods to share information between models. In Droppo et al. (2010) four approaches to model coupling are described; three external coupling methods and one internal coupling method. The three external coupling methods are: (1) modify the source code of existing models to pass data from model to model, (2) write code to create "model wrappers" that handle data exchange without modifying the source code of the models, and (3) in specific cases were the data formats are well defined, use "data-parsing" and "data mapping" to send data from one model to the next. Droppo et al. 2010) also presents an internal coupling method, OpenMI. Using OpenMI, data are exchanged directly between models according to a standardized exchange protocol. The OpenMI defines the protocol for models to exchange data at runtime, allowing models to be run in parallel and share information at each time-step (Gaber et al. 2008).

Droppo et al. (2010) explored all four of these approaches and found that all had pros and cons; these are summarized in the article. While Droppo et al. (2010) and Gaber et al. (2008) offer insight into the direction integrated modeling may be heading, researchers currently needing to work in the decision space requiring multiple models have limited options. Since implementing procedures for internal model coupling generally needs to be performed by the model developers (Droppo et al. 2010), researchers typically use some method of external coupling, cascading data from one model to the next.

3.3.1 Cascading Data – Downward Data Flow

Cascading data is the current state-of-the-practice in integrated modeling (USEPA 2001; Kittle et al. 2006; Finholt and VanBriesen 2007; Cuddy and Fitch 2010). This approach has been adopted by the Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) user interface for natural water systems. BASINS is a geographic information system (GIS) interface that provides tools that assist the user in populating data tables needed in working with some of the commonly used watershed models (USEPA 2001; Kittle et al. 2006; Johnson 2005), and has been expanded to include some receiving water/reservoir models. Using Hydrological Simulation Program - Fortran (HSPF) (Bicknell et al. 2001), AQUATOX (Clough 2009), and EPANET (Rossman 2000), **Figure 3-13** demonstrates the conceptual flow of information in, out, within, and outside the BASINS interface.

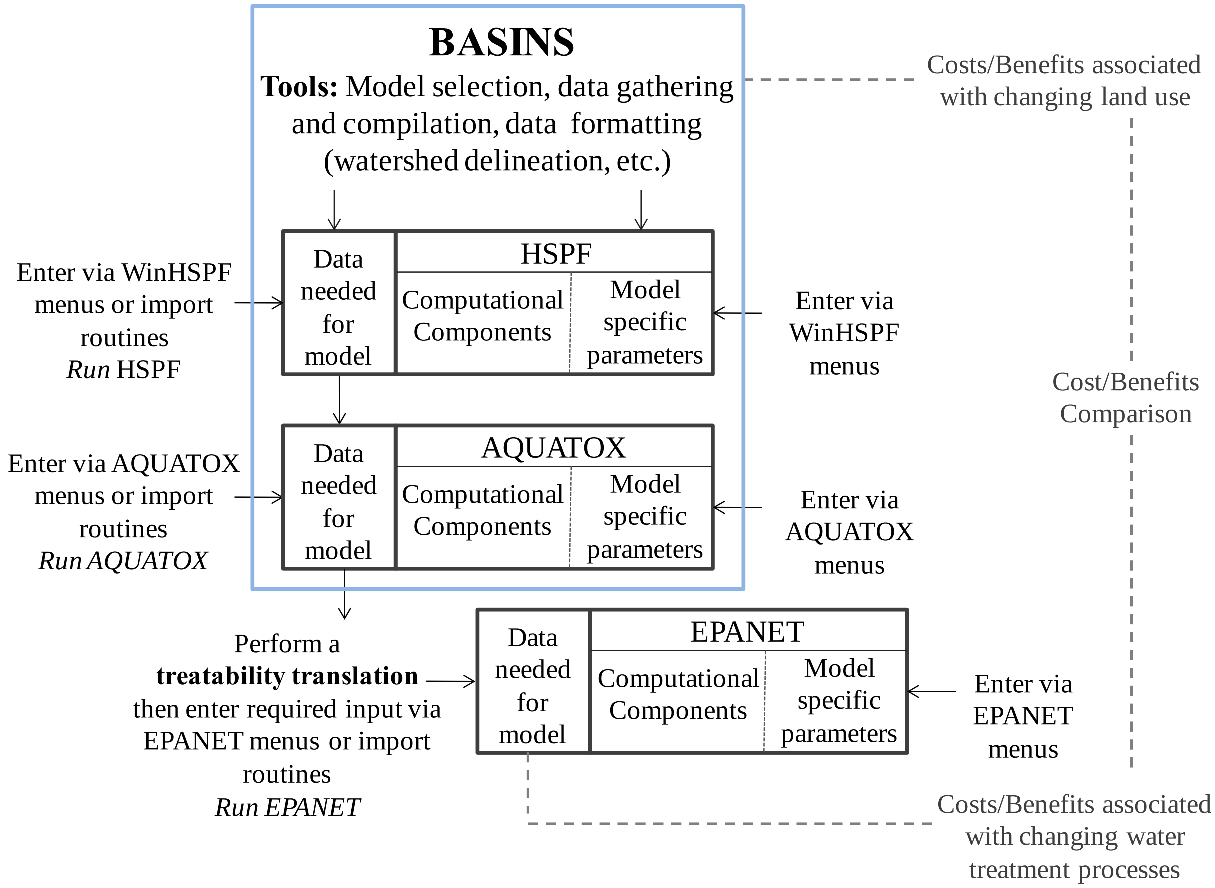

Figure 3-13. Schematic of the Cascading Data Flow Approach to Model Integration. BASINS boundaries shown with a blue box. Data shown cascading from model to model in series. EPANET and costs/benefits comparisons shown outside the BASINS interface.

In **Figure 3-13,** it is shown that the BASINS interface can be used for gathering and compiling some of the data needed for modeling with HSPF and AQUATOX (note that EPANET is outside of the BASINS interface), and is then used to indicate that the modeling will be performed using HSPF. The BASINS tools will populate some of the information required for modeling with HSPF, but, as shown in the figure, other data and model parameters must still be entered using the HSPF specific interface. The HSPF model can then be executed using the HSPF interface, and output from HSPF passed to AQUATOX (using the HSPF specific tools), where a similar process is followed. As shown in **Figure 3-13,** using the cascading data approach, data needed for a specific model are stored within that model, and customized tools are used to export and import data from one model to the next. These tools can be provided by the model developers, or programmed by individual modelers or by a third party.

Figure 3-13 demonstrates the concept of external model coupling. External coupling is demonstrated in the integration between AQUATOX and EPANET; the models are linked offline. While there is a preprogrammed link between HSPF and AQUATOX, it too is an external coupling.

3.3.2 Shared Data – Multidirectional Data Flow

The method of cascading data from model to model is a viable option for the deployment of a functional water information system. Data structures can be expanded and additional models and tools could be brought into a robust user interface. But there are limitations to this approach, most significantly, the data only flow in one direction. For example, changes to the HSPF control files, made using the HSPF menu tools do not propagate back to the BASINS tables from which they came, nor do changes made from within the AQUATOX menu system propagate back to HSPF. This downward flow of information limits the modeler's ability to fully document the parameters used in an integrated model simulation and also makes it difficult to avoid confusion over model parameterization.

To combat this problem, a multidirectional data flow approach via a master data repository was shown previously, in **Figure 1-2**. A comparison of the two approaches to data flow, depicted using abbreviated images of **Figure 1-2** and **Figure 3-13** are shown side-by-side in **Figure 3-14**.

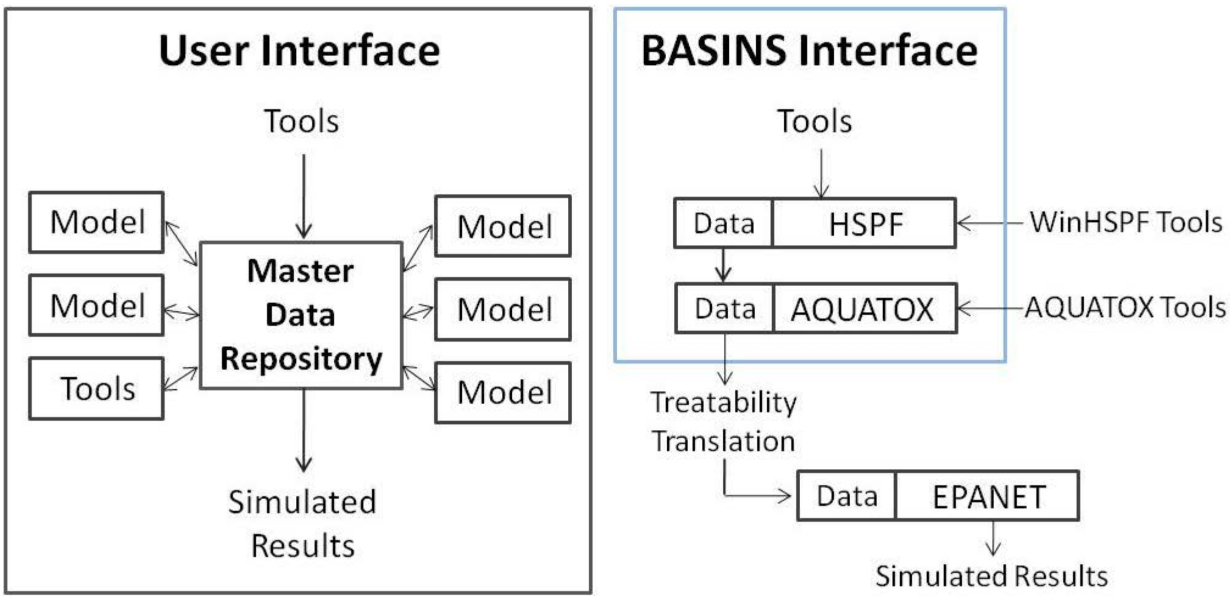

Figure 3-14. A Side-by-Side Comparison of Data Flow and Data Transfer Methods. The left image shows multidirectional data flow via a master data repository. The right image shows the more traditional method of cascading data downward through the models.

The benefits to the WatIS shown on the left in **Figure 3-14** are: (1) data can be accessed from all models in the system and data flow is multidirectional, (2) the data structure is common to all the models, encouraging data standardization among researchers, (3) the results of the model simulation and all the associated metadata can be stored in the data repository, and (4) the structure allows for a 'plug and play' nature for model inclusion. This approach would eventually reduce work for those using the models, but would require an extensive effort on the part of the model developers, at least initially, to transition their existing data structures to a new format, and would require a large, upfront investment, with a long-term commitment to maintaining the master data repository and the user interface.

The cascading data approach offers the advantage of being able to be developed in pieces, one model at a time, but cannot overcome the limitation of unidirectional data flow. Multidirectional flows are necessary to integrate real-time applications and facilitate adaptive management. Multidirectional flows could also be achieved by allowing data exchange between all models used in a multi-model simulation at runtime. Unfortunately, without a master data repository, modelers will still be required to learn data storage schemas for all the models they want to use, and will have no way of tracking or documenting the inputs and outputs of an integrated model simulation. Thus, the ideal solution would be to standardize both data storage and data transfer protocols.

3.3.3 Translating Data

Regardless of whether data are being cascaded or shared via a repository, the data may need to be translated. Conceptually, the process of moving data from HSPF to AQUATOX is fairly straight forward; HSPF generates a daily time series loading, and AQUATOX accepts a daily time series loading. Some complexity is introduced when transferring time series information that is not correlated one-to-one between the two models (Droppo et al. 2010). For example, the output loadings of biological oxygen demand and organic carbon are summed to estimate the organic matter loading in AQUATOX.

While some data processing is required to cascade data from the watershed model to the receiving water/reservoir model, it is minor compared with the challenges of translating data from tools that model natural systems to tools that model built/engineered systems. These transfers sit right at the interface between the natural and the built environments, and, as a result of the conceptual isolation of the two systems, the key parameters of the two types of models are often significantly different. To highlight this point, a few examples are provided. One example relates to taste and odor problems in drinking water; a watershed model may predict temperature, nutrient levels, and the algal biomass concentration in the source water, while the drinking water plant requires the concentrations of taste and odor (T&O) precursors like geosmin and 2-methylisoborneol (2-MIB). Another example focuses on disinfection by-products (DBPs) in drinking water; naturally-occurring organic matter (NOM) that is present in raw source water can react with chemical oxidants used for disinfection and form DBPs in finished water. Laboratory work has been extensive to characterize NOM to better understand DBP precursors and water treatability (e.g., Richardson and Ternes 2005; Reckhow et al. 1990; Owen 1995; Nikolaou et al. 2004; Rosario-Ortiz et al. 2007; Archer and Singer 2006), but NOM, as an explicitly defined parameter, is not simulated in the watershed/reservoir models.

The prediction of T&O and/or DBPs problems now relies on detailed information about the characteristics of the source water, but there is no direct, easily incorporated translation from the watershed parameters to the drinking water intake parameters. Thus, to integrate these system models, as shown in both **Figure 1-2** and **Figure 3-13**, a treatability translation is needed. A treatability translation is a complex conversion of data, using algorithms that incorporate expert knowledge regarding the controlling physical, chemical, and biological processes that lead from the state of the water system at the DWTP intake to the state of the water system as engineered processes are initiated. The accuracy of the algorithms, and thus of the treatability translation, hinges on access to water quality monitoring in the watershed and on water quality monitoring at the DWTP intake in order to inform the relationship among the different water quality terms.

3.3.4 Integrating SCADA Data into the Water Information System

One of the goals of the water information system (WatIS) is to link prediction of the quality of finished drinking water at a future point in time with information received from sensors in the watershed upstream of the drinking water treatment plant. To accomplish this, the plant's supervisory control and data acquisition (SCADA) system will need to be integrated into the WatIS. Data from water utilities have historically been managed through the use of SCADA systems. Although the SCADA system has traditionally been limited to the engineered environment (i.e., the drinking water distribution system), there is no reason why it cannot be expanded to reach outside the walls of the built environment, and into the natural environment. Such an expansion has previously been suggested as part of a source water protection plan focused on spill detection and response (Grayman et al. 2001), but there is scant literature to suggest that such an integration has been deployed for real-time operational change in response to less urgent upstream events.

The purpose of a SCADA system is to allow operators to monitor and control equipment and processes from a central processing center and in real-time (Lahlou 2002) to ensure regulatory compliance (Joshi et al. 2004). In addition to water treatment systems, SCADA systems are used to control various utilities such as power generation systems, electrical distribution systems, and hazardous waste treatment facilities (USEPA 2011b). The capabilities of current SCADA systems generally include collection, storage, and management of a variety of historical and sensor data (Joshi et al. 2004; Lahlou 2002; USEPA 2011b). These data, integrated with the SCADA system, can be used to detect operational anomalies, trigger alarms, and automate operations such as chlorine addition or pump activation (Doyle and Fayyad 1991; Walski et al. 2001; Joshi et al. 2004; USEPA 2011b).

Several teams have explored the potential for integrating system models with real-time data. Joshi et al. (2004) have suggested that water quality models can be integrated with the real-time data available through a SCADA system to develop a model that is able to more accurately predict current and future behavior of the system. Shinozuka et al. (2005) studied the use of real-time data with models to identify the location and extent of damage in a network. Schulte and Malm (Schulte and Malm 1993) considered a system in Illinois used to confirm system design, develop operational scenarios, and train operators. Tiburce et al. (1999) and Joshi et al. (2004) report improved operational control and emergency preparedness for systems with integrated modeling and sensing.

Integrating the WatIS and the SCADA will require an interdisciplinary, cooperative effort between water professionals, watershed specialists, and information technology professionals (Computing Community Consortium 2011). Once completed, the WatIS would interface directly with the plant, and treatment processes could be altered in real-time based on measured and modeled data and information, maximizing the quality of the finished drinking water while minimizing treatment costs. While the development of a WatIS could be accomplished independent from the SCADA, since many of the functions of the SCADA are needed in the

WatIS, it would be best to integrate these efforts. This will also enable alignment with SCADA upgrades that many water plants have in their plans for the coming decade to comply with increasingly frequent data requests from regulatory and policy-setting agencies as well as to exert greater operational control over their distribution systems (Shinozuka and Dong 2005; Shastri and Diwekar 2006).

Experience with Model Integration

As discussed in **Section 3.3**, three model linkages were explored: BASINS to HSPF, HSPF to AQUATOX, and AQUATOX to EPANET. These linkages are shown in the schematic in **Figure 3-13** and represent a cascading data approach to integrated water modeling.

BASINS to HSPF

The transition from BASINS to HSPF is very smooth. From the information entered into BASINS, an HSPF project can be created and populated with enough information to perform a flow simulation. The relationship between BASINS and HSPF is not dynamic; once an HSPF project has been created, additional changes to the HSPF files are typically performed using WinHSPF rather than in BASINS. The flow of data from BASINS to HSPF is discussed in **Section 3.2.1.** The process is summarized in **Figure 3-15**.

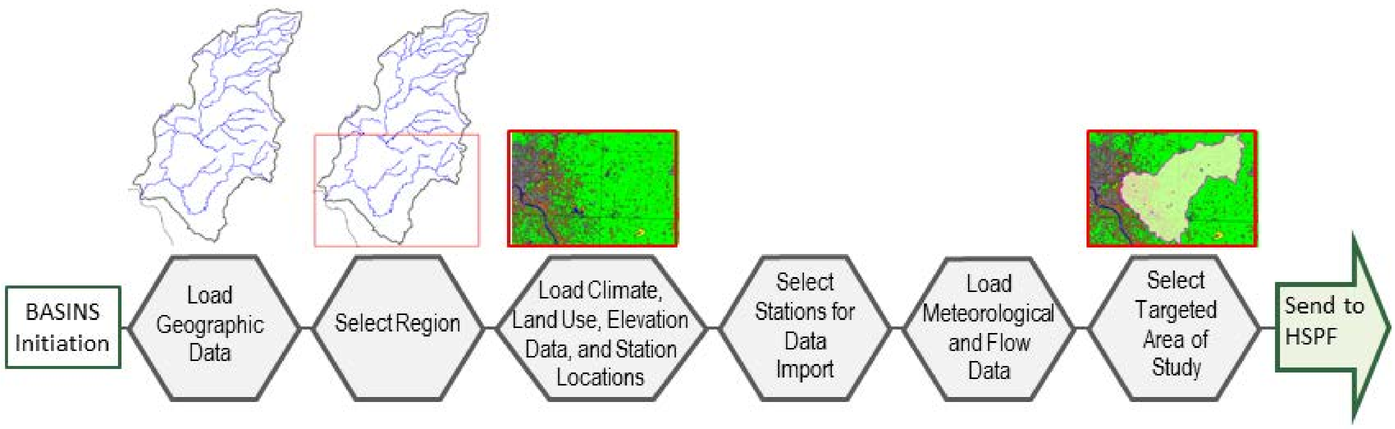

Figure 3-15. The Process of Selecting Data for Watershed Modeling Using BASINS. Physiogeographic information narrowed three times during BASINS selection process; from the world, to the Little Miami Watershed, to the East Fork Watershed. BASINS tools are then used to send data to HSPF.

Notice in **Figure 3-15** that there are three times when the user can narrow the area of the study. It would be too resource intensive to pull all the available information into a BASINS project, so BASINS allows the user to narrow the study area geographically. In some cases, it may be necessary to use data gathered from outside the boundaries of the targeted area of study. For example, if no precipitation monitoring stations are located within the targeted study area, data may be needed from a location close by. BASINS allows the user to keep the neighboring locations, without having to keep all data associated to a whole region.

Once all the required information is loaded into BASINS, the built-in feature to generate an HSPF project is used. This feature sends data required for running an HSPF project into the appropriate files, and automatically opens the HSPF files using WinHSPF.

HSPF to AQUATOX

The process of using WinHSPF to model changes in the watershed (e.g., implementing BMPs) and send the model results to AQUATOX is summarized in **Figure 3-16**.

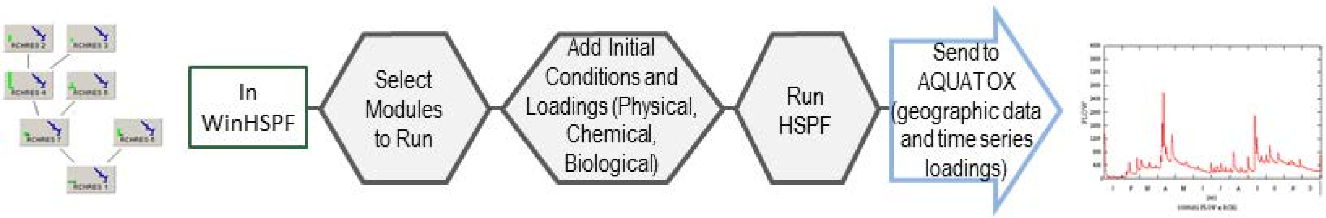

Figure 3-16. The Process of Running a Simulation in HSPF Using WinHSPF.
In WinHSPF, model parameters and data not coming from BASINS are loaded. After the simulation is run, time series loadings at the outlets of the river reaches are passed into AQUATOX.

The details of the process shown in **Figure 3-16** are provided in **Section 3.2.1.** HSPF allows the user to select which modules to use in modeling the watershed (e.g., modeling a pollutant requires the use of a specific pollutant module). Based on the modules selected, the required information matching those modules must be included in the input file. Running the HSPF model yields estimated loadings of modeled parameters at the outflow of the modeled river reaches. These time series loadings can be sent to AQUATOX. There is a feature built into the WinHSPF interface that transfers data from HSPF to AQUATOX (Clough 2005); this feature worked well, but it only transfers information for riverine reaches, and each reach must be transferred individually.

AQUATOX to EPANET

Once in AQUATOX, the complete structure of the model is constructed, and site information and loadings for all the other segments of the model must be entered. This process is shown in the left half of **Figure 3-17**.

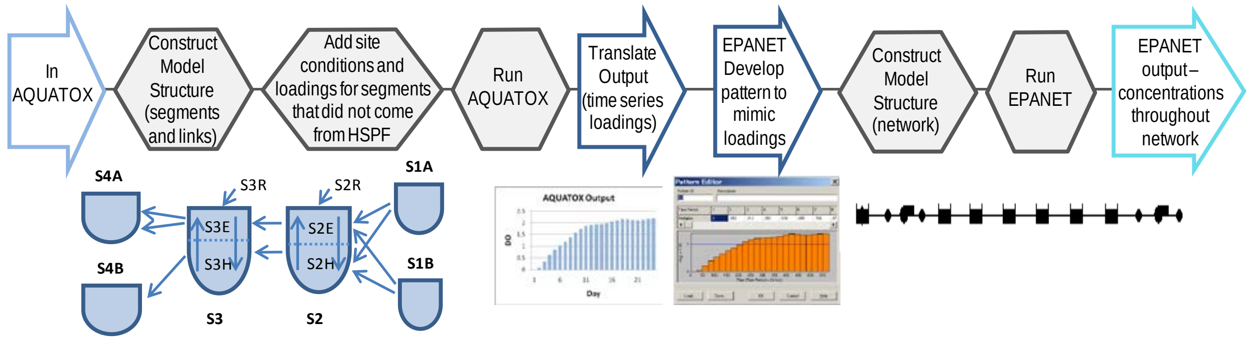

Figure 3-17. Process of Modeling with AQUATOX and Transferring Data to EPANET. Data not sent over from WinHSFP must be loaded. The model is executed, and the time series concentrations in the segments are exported. Time series loadings needed for EPANET are converted to an average and a pattern and entered into EPANET. EPANET is executed and concentrations of a parameter of interest are generated.

3.4.4 Multiple Models and Time Steps

With regard to time steps of the models, HSPF documentation indicates that it runs on an hourly time step, but the output can be changed to display in a variety of formats, including a daily sum, average, maximum, minimum, or one of a few other formats. AQUATOX is fairly flexible in its time step; the user can choose to model daily, hourly, or fractions of either. The number of time periods in the EPANET pattern editor can be set to the desired duration, and the time step can be adjusted so that those units can be one hour to 24 hours. In the case where the time step is a day, the pattern will last the number of days that corresponds to the number of time periods set in the pattern editor. For this project, the HSPF output was exported on a daily basis, AQUATOX was modeled with a daily time step for 24 days, and EPANET was modeled with a pattern of a daily time step for 24 days.

4 Moving Towards a Fully Functional Water Information System

There are significant challenges to designing and developing a fully functional water information system; these are discussed below followed by a section summarizing the project findings and recommendations for next steps.

4.1 Challenges

Challenges identified as part of this assessment project are grouped below as data challenges, model challenges, and challenges with integrating the models.

4.1.1 Data Challenges

The importance of organizing data is discussed in detail in Section 2 of this report. Good data management takes commitment and attention to detail. Forward thinking and a good data management plan in advance of a sampling project can significantly reduce the overall cost of organizing data, and a target data structure for compiling and archiving data should be in place prior to the start of sample collection. The issue regarding data that must be addressed in preparing to develop a fully functional WatIS is whether to compile data and format it for the specific models to be used in the very near future, or to take a more holistic approach and attempt to develop and work with a common data structure that could be used as a master repository (and that could be used to interface with the SCADA). Clearly, the holistic approach will require a higher level of effort in the near term, but the long-term gains could be significant.

A limiting factor in modeling is often the availability of organized, documented data. In many cases data are collected, but, because they are not formatted consistently and/or stored with metadata they are difficult to share with the broader scientific community. Analyzing water systems across sites and times will be essential to transform the study of water from local case studies to managing water as a global resource (Horshburgh et al. 2008; Horshburgh et al. 2009). This will require data from multiple research projects to be aggregated across environmental sampling programs. The only reasonable way to enable data sharing is for researchers to format their data in a standardized, documented data structure. Thus, water professionals should move in the direction of storing data from sampling programs and research projects in a data structure that could eventually be used for the master data repository of the WatIS. High value data from previous sampling programs should be organized, and reformatted to fit into the same data structure. The data management system selected should be robust enough to manage existing data and to easily incorporate additional data that becomes available from laboratory and field experiments as well as real-time sensors deployed in water systems (ASW 2011).

Further, in order to inform decisions across the watershed to drinking water space, cost and benefit information for changing land use in the watershed (e.g., implementing BMPs) as well as costs and benefits of altering treatment processes and/or making capital improvements to the drinking water treatment plant must be compiled and integrated into the database so that they are available to drive economic cost-benefit models.

4.1.2 Model Challenges

Depending on the simulation capabilities needed, models must be identified and explored. To model the source water through drinking water treatment plant interface, a watershed model and a drinking water treatment plant model are required. A reservoir model may also be needed, and a water distribution network model could be included depending on the focus of the study. The models used for this assessment project were selected to meet a set of specific goals: to assess the difficulty in developing a WatIS and to identify the challenges that must be overcome. This focus led to the selection of HSPF, AQUATOX, and EPANET. As a result of working with the models, the project researchers are considering the possible need to include additional models to characterize the linked natural and built water system. For example, while AQUATOX will simulate changes in populations of biotic life, it is not a complex hydrodynamic model; thus, to get to a robust WatIS, a biotic model such as AQUATOX will need to be coupled with a hydrodynamic model such as CE-QUAL-W2 (Cerco and Cole 1995). The determination of which models will be incorporated into the WatIS will depend, in part, on the model developer's willingness and ability to adapt their models to fit the WatIS framework.

In addition to watershed, reservoir, treatment plant, and water distribution models, there are several other kinds of models and modeling tools that will need to be added to the WatIS; these include treatability translation tools, socioeconomic and cost/benefit analysis tools, and/or data analysis tools. To achieve the goal of being able to alter treatment processes in real-time as a result of changes in the upstream watershed, the drinking water treatment plants' SCADA systems will need to be integrated into the WatIS. To move in this direction, the number of different SCADA systems, and the proprietary nature of each system will need to be assessed; thus, SCADA developers will need to be included in the interdisciplinary team working on WatIS and will have to agree to either transform SCADA so that it will interface internally with the WatIS, or will need to develop algorithms for reading and writing to the WatIS.

4.1.3 Model Integration Challenges

In moving to a fully functional water information system, some strategic decisions must be made. Two approaches to integrating models have been discussed in this report. These approaches are shown in **Figure 1-2** and **Figure 3-13** and are summarized side-by-side in **Figure 3-14**. The significant difference between the two approaches is the master data repository. In the multidirectional flow WatIS (shown in **Figure 1-2**), whenever possible, data are stored in the master data repository. Using the cascading data approach (see **Figure 3-13**), there is no data repository, rather, data are passed from model to model, in some cases using specialized data transfer routines. Implementation of these model integration methods can take many forms, as described in **Section 3.3**.

Unfortunately, developing the multidirectional data flow WatIS would require a large, upfront investment with a commitment to maintaining the master data repository. In contrast, using the BASINS cascading data flow approach, the interface can be developed in pieces, one model at a time. The problem with this approach is that the modeler must learn the intricacies of all data storage structures for each model, and since there is no master data repository, the metadata associated to a simulation performed using multiple models may be lost. In the short-term, particularly since no data structure has emerged as the leading candidate for inclusion into the WatIS, it is likely that the BASINS approach will continue to dominate in model integration work. Perhaps the most significant problem with cascading data is that, since there is not a unifying data structure, there is no additional motivation for researchers to standardize their data, and thus, multiple formats for data collected on water-based research projects will remain the norm.

For this project, data were transferred using the BASINS tools or were transferred from model to model manually. Several of the models have built-in tools to import from Excel or text files, and export to Excel or text files; these were also used. BASINS does a good job of helping the modeler move data from one model to the next, but still falls short of allowing data to move in real-time and does not incorporate the built water environment systems. In a functioning WatIS, models will have to share a common data structure or robust automated translation algorithms will need to be developed and implemented to allow data and simulation results to move among models in real-time.

4.2 Findings and Recommendations

The goal of the present work was to evaluate the complexities of integrating multiple models from the watershed through the drinking water plant and to identify knowledge and information gaps that make working in this decision space a continuing challenge. Issues identified include those associated with data, models, and model integration.

4.2.1 The Data

Modeling requires an extensive dataset be available, and it is critical that these data be used appropriately. While much watershed data exists, it is often stored in a non-standardized, undocumented format, making direct dialogue with the primary researcher necessary to understand how data can be restructured for use in modeling. This situation is a typical one, and limits the ability of modelers to maximize data and fully characterize water systems. Water professionals should move in the direction of storing data collected as part of field sampling programs and research projects in a data structure that could eventually be used for the master data repository. High value data from previous sampling programs should be organized, and reformatted to fit into the same data framework.

4.2.2 The Models

In some cases, there is no ideal model option to perform the necessary task (e.g., a drinking water treatment plant model). These models will need to be developed. All models to be integrated into the multidirectional flow WatIS must be restructured to work with the master data repository. New models could be built directly using the data structure and older models could either be completely overhauled to work directly with the master data structure (the preferred approach), or could be equipped with data processing scripts to read and write directly to the master data repository. All models considered for incorporation into the WatIS should be evaluated in terms of their relevance in a multisystem model framework; models selected for inclusion will need to be adapted for use in both natural and built/engineered systems. To assist the user in selecting the right model for a specific task, a model selection guidance tool should be incorporated into the WatIS user interface. To make WatIS a reality, a consortium of modeling experts, data collectors, and information technology professionals will be needed to make sure that all the intricacies associated with each model are understood and tuned correctly.

4.2.3 The Model Integration

The decision of whether to adopt the master data repository approach to the WatIS, to pursue the cascading data approach, or to develop some sort of a hybrid approach needs to be considered carefully. There are significant challenges to deploying a fully functional multidirectional data flow WatIS as describe above, including a significant upfront investment to design the master data repository, adapt a core group of models for use within the WatIS framework, and develop the user interface. In the near term, specific models can be selected and linked through customized tools using external coupling (e.g., expanding the BASINS approach) or integrated using internal coupling (e.g., using OpenMI), but at some point, it may become more desirable to upgrade all models so they work with a standardized common data structure. If the cascading data approach is used, priority should be given to the development of automated data handoff tools to avoid the laborious process of regenerating the data movement for each new watershed studied. Furthermore, ongoing research is needed to inform the development of the treatability translation and cost/benefit models and tools. Ultimately, for use in real-time, the drinking water treatment plant supervisory control and data acquisition (SCADA) systems must be incorporated into the WatIS to provide operational control and management of drinking water systems.

Adopting multidirectional flow and the unifying structure of the WatIS master data repository will expedite a ‘plug and play’ nature for model inclusion, facilitating the development of data analysis tools (and/or links to commonly used tools), the inclusion of treatability translation tools, and the incorporation of models for comparing costs/benefits, leading to a robust, integrated water information system for managing the quantity and quality of water across the natural and built environments.

There are several directions for “next steps" in the research that will contribute to the overall goal of developing a WatIS to enable improved management of water systems. These include: (1) determine essential data elements that need to be stored in the master data repository, (2) improve data management systems by further defining the structure of the master data repository, (3) determine the best way to exchange data within the WatIS, (4) investigate fundamental relationships that will lead to the development of treatability translation modules that are critical to linking natural systems with engineered systems, (5) assess the types of tools that will be needed in the WatIS User Interface, (6) determine other types of models that will be needed in the WatIS, and (7) perform a case study at a specific location using one model to better understand what other data and/or modeling capabilities are needed in the WatIS. Working to accomplish these tasks will help inform the longer-term effort to develop a multidirectional data flow WatIS.

5 References Cited

AFCEE (2009). ERPIMS 2008 Data Loading Handbook Version 5.1.1222; http://www.afcee.lackland.af.mil/erpims/DLhandbook/html/index.html?home.htm.

AQUA TERRA Consultants (2011). "BASINS 4.0 Development." Retrieved August 18, 2011; http://www.aquaterra.com/projects/descriptions/basins40.php.

Archer, A. D. and P. C. Singer (2006). "An Evaluation of the Relationship between SUVA and NOM Coagulation Using the ICR Database." Journal American Water Works Association **98**(7): 110-123.

ASCE (2004). Interim Voluntary Guidelines for Designing an Online Contaminant Monitoring System, American Water Works Association Water Environment Federation; http://www.michigan.gov/documents/deq/deq-wb-wws-asceocms_265136_7.pdf.

ASW (2011). "Aquatic Sensor Workgroup Methods and Data Comparability Board." Retrieved April 27, 2011; http://www.watersensors.org/.

Bentley Systems Incorporated (2009). "WaterCAD Water Distribution Modeling and Management." ftp://ftp2.bentley.com/dist/collateral/docs/watercad/watercad_product_data_sheet.pdf.

Beran, B., J. Goodall, T. Min, D. Tarboton, E. To, D. Valentine and T. Whiteaker (2006). "CUAHSI HIS Web Services Workbook." Retrieved July 16, 2011; http://www.cuahsi.org/his/docs/HIS-workbook-20061130.pdf.

Bicknell, B. R., J. C. Imhoff, J. L. Kittle, Jr., T. H. Jobes and A. S. Donigian, Jr. (2001). Hydrological Simulation Program - Fortran (HSPF). User's Manual for Release Version 12. Athens, GA, U.S. EPA National Exposure Research Laboratory in cooperation with U.S. Geological Survey, Water Resources Discipline, Reston, VA.

Booty, W. and G. Benoy (2009). "Multicriteria Review of Nonpoint Source Water Quality Models for Nutrients, Sediments, and Pathogens." Water Quality Research Journal of Canada **44**(4): 365-377.

Borah, D. K. and M. Bera (2004). "Watershed-Scale Hydrologic and Nonpoint-Source Pollution Models: Review of Applications." Transactions of the ASAE **47**(3): 789-803.

Borah, D. K., G. Yagow, A. Saleh, P. L. Barnes, W. Rosenthal, E. C. Krug and L. M. Hauck (2006). "Sediment and Nutrient Modeling for TMDL Development and Implementation." Transactions of the ASABE **49**(4): 967-986.

Cerco, C. F. and T. Cole (1995). User's Guide to the CE-QUAL-ICM Three-Dimensional Eutrophication Model, Release Version 1.0, Technical Report El-95-15. Vicksburg, MS, US Army Engineer Waterways Experiment Station.

Clough, J. S. (2005). AQUATOX (Release 2.1) Modeling Environmental Fate and Ecological Effects in Aquatic Ecosystems Volume 3: User's Manual for the BASINS (Version 3.1) Extension to AQUATOX Release 2.1. Washington DC, USEPA.

Clough, J. S. (2009). AQUATOX (Release 3) Modeling Environmental Fate and Ecological Effects in Aquatic Ecosystems, Volume 1: User's Manual. Office of Water. Washington DC, USEPA. **Vol 1**.

Computing Community Consortium (2011). Science, Engineering, and Education of Sustainability: The Role of Information Sciences and Engineering; http://cra.org/ccc/docs/RISES_Workshop_Final_Report-5-10-2011.pdf.

CUAHSI (2011a). "ODM Relational Tables." Retrieved March 1, 2011; http://his.cuahsi.org/images/ODM1_1SchemaDiagram_md.jpg.

CUAHSI (2011b). "WaterOneFlow Web Services & WaterML." Retrieved April 11, 2011; http://his.cuahsi.org/wofws.html.

CUAHSI (2011c). "What Is CUAHSI?" Retrieved February 26, 2011; http://www.cuahsi.org/docs/What-is-CUAHSI.pdf.

Cuddy, S. M. and P. Fitch (2010). Hydrologists Workbench - a Hydrological Domain Workflow Toolkit. 2010 International Congress on Environmental Modelling and Software Modelling for Environment's Sake, Fifth Biennial Meeting, Ottawa, Canada, International Environmental Modelling and Software Society (iEMSs).

Doyle, R. J. and U. M. Fayyad (1991). Sensor Selection Techniques in Device Monitoring. Second Annual Conference on AI, Simulation and Planning in High Autonomy Systems, IEEE Computer Society Press.

Dozier, J., J. B. Braden, R. P. Hooper, B. S. Minsker and J. L. Schnoor (2009). Living in the Water Environment: The WATERS Network Science Plan; http://www.watersnet.org/docs/WATERS_Network_SciencePlan_2009May15.pdf.

Droppo, J. G., G. Whelan, M. E. Tryby, M. A. Pelton, R. Y. Taira and K. E. Dorow (2010). Methods to Register Models and Input/Output Parameters for Integrated Modeling. International Environmental Modelling and Software Society (iEMSs) 2010 International Congress on Environmental Modelling and Software Modelling for Environment's Sake, Fifth Biennial Meeting, Ottawa, Canada.

Duda, P., J. Kittle, Jr., M. Gray, P. Hummel and R. Dusenbury (2001). WinHSPF Version 2.0: An Interactive Windows Interface to HSPF (WinHSPF) Users' Manual, US Environmental Protection Agency.

Duda, P. B., D. P. Ames and J. N. Carleton (2011). "BASINS 4.0: Overview and Recent Developments." Retrieved February 28, 2011; http://www.aquaterra.com/about/news20100329a.php.

Dudley, J., G. Dillon and L. C. Rietveld (2008). "Water Treatment Simulators." Journal of Water Supply Research and Technology-Aqua **57**(1): 13-21.

Edwards, M. (1997). "Predicting DOC Removal During Enhanced Coagulation." Journal American Water Works Association **89**(5): 78-89.

Environmental Information Exchange Network (2001). "Data Standards List." Retrieved February 26, 2011; http://www.exchangenetwork.net/standards/listing.htm.

ESRI (2011). "What Is ESRI's History? Where Are ESRI and GIS Headed in the Future?" Retrieved June 24, 2011; http://events.esri.com/uc/QandA/index.cfm?fuseaction=answer&conferenceId=2A8E2713-1422-2418-7F20BB7C186B5B83&questionId=2550.

Finholt, T. and J. VanBriesen (2007). WATERS Network Cyberinfrastructure Plan, The WATERS Network Project Office Cyberinfrastructure Committee

Fisher, I., G. Kastl, A. Sathasivan, P. Chen, J. van Leeuwen and R. Daly (2004). "Tuning the Enhanced Coagulation Process to Obtain Best Chlorine and THM Profiles in the Distribution System." Water Science and Technology: Water Supply **4**(4): 235-243.

Francis, R. A., M. J. Small and J. M. VanBriesen (2009). "Multivariate Distributions of Disinfection by-Products in Chlorinated Drinking Water." Water Research **43**(14): 3453-3468.

Francis, R. A., J. M. VanBriesen and M. J. Small (2010). "Bayesian Statistical Modeling of Disinfection Byproduct (DBP) Bromine Incorporation in the ICR Database." Environmental Science and Technology **44**(4): 1232-1239.

Gaber, N., G. Laniak and L. Linker (2008). Integrated Modeling for Integrated Environmental Decision Making; http://www.epa.gov/crem/library/IM4IEDM_White_Paper_Final_(EPA100R08010).pdf.

Grayman, W. M., R. A. Deininger and R. M. Males (2001). Design of Early Warning and Predictive Source-Water Monitoring Systems, American Water Works Association Research Foundation.

Harrington, G. W., Z. K. Chowdhury and D. M. Owen (1992). "Developing a Computer-Model to Simulate DBP Formation During Water-Treatment." Journal American Water Works Association **84**(11): 78-87.

Hedeen, S. (2010). "The Little Miami Wild & Scenic River Ecology & History." Retrieved April 1, 2010; http://littlemiami.com/LITTLE%20MIAMI%20RIVER%20ECOLOGY%20AND%20HISTORY.pdf.

Horshburgh, J. S., D. G. Tarboton, D. R. Maidment and I. Zaslavsky (2008). "A Relational Model for Environmental and Water Resources Data." Water Resources Research **44**(W05406): 1-12.

Horshburgh, J. S., D. G. Tarboton, M. Piasecki, D. R. Maidment, I. Zaslavsky, D. Valentine and T. Whitenack (2009). "An Integrated System for Publishing Environmental Observations Data." Environmental Modelling and Software **24**: 879-888.

Imhoff, J. C., A. Stoddard and E. M. Buchak (2003). Evaluation of Contaminated Sediment Fate and Transport Models Final Report. Athens, Georgia, National Exporsure Research Laboratory.

Jakeman, A. J., R. A. Letcher and J. P. Norton (2006). "Ten Iterative Steps in Development and Evaluation of Environmental Models." Environmental Modelling and Software **21**(5): 602-614.

Johnson, N. W. (2005). ArcGIS and HSPF Model Development. Masters, The University of Texas.

Joshi, P., T. M. Walski, S. Gandhi, J. A. Andrews and C. F. Newswanger (2004). "Case Study: Linking Bristol Babcock's SCADA Systems to WaterCAD, a Water Distribution Modeling Tool." American Water Works Association DSS Conference.

Julien, P. Y., B. Saghafian and F. L. Ogden (1995). "Raster-Based Hydrologic Modeling of Spatially-Varied Surface Runoff." Water Resources Bulletin **31**(3).

Kittle, J. L., Jr., P. B. Duda, D. P. Ames and R. S. Kinerson (2006). Geographic Information Systems and Water Resources IV AWRA Spring Specialty Conference Houston, Texas, AWRA.

Lahlou, Z. M. (2002). Tech Brief - a National Drinking Water Clearinghouse Fact Sheet; System Control and Data Acquisition (SCADA), National Drinking Water Clearinghouse; http://www.nesc.wvu.edu/pdf/dw/publications/ontap/2009_tb/system_control_SCADA_DWFSOM20.pdf.

Liu, Y., H. Gupta, E. Springer and T. Wagener (2008). "Linking Science with Environmental Decision Making: Experiences from an Integrated Modeling Approach to Supporting Sustainable Water Resources Management." Environmental Modelling and Software **23**(7): 846-858.

Migliaccio, K. W. and P. Srivastava (2007). "Hydrologic Components of Watershed-Scale Models." Transactions of the ASABE **50**(5): 1695-1703.

National Ecological Observatory Network (2011). "Welcome to NEON." Retrieved June 24, 2011; http://www.neoninc.org/.

Nikolaou, A. D., S. K. Golfinopoulos, G. B. Arhonditsis, V. Kolovoyiannis and T. D. Lekkas (2004). "Modeling the Formation of Chlorination by-Products in River Waters with Different Quality." Chemosphere **55**(3): 409-420.

NSF (2007). Cyberinfrastructure Vision for 21st Century Discovery. National Science Foundation Cyberinfrastructure Council, National Science Foundation http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf.

NSF (2011). "Data Management & Sharing Frequently Asked Questions (FAQs) - Update November 30, 2010." Retrieved March 23, 2011; http://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp#3.

NWIS (2011a). "NWIS Getting Data." Retrieved February 26, 2011; http://river.sdsc.edu/NWISTS/NWIS.asmx.

NWIS (2011b). "USGS Water Data for the Nation." Retrieved February 26, 2011; http://waterdata.usgs.gov/nwis.

Obolensky, A. and P. C. Singer (2005). "Halogen Substitution Patterns among Disinfection Byproducts in the Information Collection Rule Database." Environmental Science and Technology **39**(8): 2719-2730.

Obolensky, A. and P. C. Singer (2008). "Development and Interpretation of Disinfection Byproduct Formation Models Using the Information Collection Rule Database." Environmental Science and Technology **42**(15): 5654-5660.

Obolensky, A., P. C. Singer and H. M. Shukairy (2007). "Information Collection Rule Data Evaluation and Analysis to Support Impacts on Disinfection by-Product Formation." Journal of Environmental Engineering **133**(1): 53-63.

Ocean Research Interactive Observatory Networks (2011). "What Is the OOI?" Retrieved June 24, 2011; http://www.orionprogram.org/OOI/default.html.

Owen, D. M. (1995). "NOM Characterization and Treatability." Journal American Water Works Association **87**(4): 148-148.

PADEP (2011). "Drinking Water Reporting System." Retrieved March 31, 2011; http://www.drinkingwater.state.pa.us/dwrs/HTM/Welcome.html.

Park, R. A., J. S. Clough and M. Coombs-Wellman (2009). Modeling Fate and Effects of Pollutants with AQUATOX Release 3. Short Course Program "Human-Environment Interactions: Understanding Change in Dynamic Systems." Society of Environmental Toxicology and Chemistry. New Orleans, LA.

Park, R. A., J. S. Clough and M. C. Wellman (2008). "AQUATOX: Modeling Environmental Fate and Ecological Effects in Aquatic Ecosystems." Ecological Modelling **213**(1): 1-15.

Reckhow, D. A., P. C. Singer and R. L. Malcolm (1990). "Chlorination of Humic Materials - by-Product Formation and Chemical Interactions." Environmental Science and Technology **24**(11): 1655-1664.

Richardson, S. D. and T. A. Ternes (2005). "Water Analysis: Emerging Contaminants and Current Issues." Analytical Chemistry **77**(12): 3807-3838.

Rosario-Ortiz, F. L., S. A. Snyder and I. H. Suffet (2007). "Characterization of Dissolved Organic Matter in Drinking Water Sources Impacted by Multiple Tributaries." Water Research **41**(18): 4115-4128.

Rossman, L. A. (2000). EPANET 2 User's Manual. Cincinnati, OH 45268, National Risk Management Research Laboratory Office of Research And Development USEPA.

Rossman, L. A., R. M. Clark and W. M. Grayman (1994). "Modeling Chlorine Residuals in Drinking-Water Distribution-Systems." Journal of Environmental Engineering **120**(4): 803-820.

Salomons, E. (2005). "Water Simulation." Retrieved July 16, 2011; http://www.water-simulation.com/wsp/2005/05/11/h2omap-water/.

Samuels, W. B., R. Bahadur, D. Amstutz and J. Pickus (2003). Pipelinenet: An Extended Period Simulation Hydraulic Model for Distribution System Emergency Response, American Water Works Association; http://eh2o.saic.com/SectionProjects/Transport/DistributeSys/PipelineNet3X/Papers/20545.pdf.

Schulte, A. M. and A. P. Malm (1993). "Integrating Hydraulic Modeling and SCADA Systems for System-Planning and Control." Journal American Water Works Association **85**(7): 62-66.

Shang, F., J. G. Uber and L. A. Rossman (2008). EPANET Multi-Species Extension User's Manual.

Shastri, Y. and U. Diwekar (2006). "Sensor Placement in Water Networks: A Stochastic Programming Approach." Journal of Water Resources Planning and Management-ASCE **132**(3): 192-203.

Shinozuka, M. and X. Dong (2005). Monitoring and Management of Water Supply Systems. 4th KISTEC International Seminar on Safety of Infrastructures.

Shinozuka, M., J. Liang and M. Q. Feng (2005). "Use of Supervisory Control and Data Acquisition for Damage Location of Water Delivery Systems." Journal of Engineering Mechanics: 225-230.

Simpson, K. L. and K. P. Hayes (1998). "Drinking Water Disinfection by-Products: An Australian Perspective." Water Research **32**(5): 1522-1528.

Small, M. J. (1997). "Show Me the Data." Journal of Industrial Ecology **1**(4): 9-12.

Stanley, S. J., C. W. Baxter, Q. Zhang and R. Shariff (2000). Process Modeling and Control of Enhanced Coagulation, American Water Works Association.

Tiburce, V., J. L. Gagnon, J. L. Hamon, P. Chopard and P. Feugier (1999). "Linking SCADA and Hydraulic and Water Quality Simulator at the Centre Des Mouvements De L'eau (Cme) in Paris (France)." Houille Blanche-Revue Internationale De L Eau **54**(2): 85-89.

Tseng, T. and M. Edwards (1999). "Predicting Full-Scale TOC Removal." Journal American Water Works Association **91**(4): 159-170.

USEPA (1989a). Resolution on the Use of Mathematical Models by EPA for Regulatory Assessment and Decision-Making Science Advisory Board (SAB). Washington, D.C., U.S. Environmental Protection Agency (EPA) EPA-SAB-EEC-89-012.

USEPA (1989b). STORET and the Water Quality Enterprise: An Initial Assessment for STORET 1995. C. Corporation. Washington, D.C. , Systems Development Center Office of Information Resources Management

USEPA (2001). BASINS 3.0 User's Manual. Washington, D.C., USEPA Office of Science and Technology; http://www.epa.gov/waterscience/basins/.

USEPA (2011a). "Drinking Water Research - EPANET." Retrieved March 3, 2011; http://www.epa.gov/nrmrl/wswrd/dw/epanet.html.

USEPA (2011b). "Security Product Guides." http://cfpub.epa.gov/safewater/watersecurity/guide/.

USEPA (2011c). "Watershed Assessment Model." Retrieved March 7, 2011; http://www.epa.gov/athens/wwqtsc/WAMView.pdf.

USEPA (2011d). "What Is a TMDL?" Retrieved March 7, 2011; http://www.epa.gov/owow/tmdl/overviewoftmdl.html.
USGS (2010). "USGS Boundary Descriptions and Names of Regions, Subregions, Accounting Units and Cataloging Units." Retrieved April 8, 2010; http://water.usgs.gov/GIS/huc_name.html#Region05.
Van Leeuwen, J., R. Daly and A. Holmes (2005). "Modeling the Treatment of Drinking Water to Maximize Dissolved Organic Matter Removal and Minimize Disinfection by-Product Formation." Desalination **176**(1-3): 81-89.
Volk, C., K. Bell, E. Ibrahim, D. Verges, G. Amy and M. Lechevallier (2000). "Impact of Enhanced and Optimized Coagulation on Removal of Organic Matter and Its Biodegradable Fraction in Drinking Water." Water Research **34**(12): 3247-3257.
Walski, T. M., D. Kaufman and W. Malos (2001). Establishing a System Submetering Project. American Water Works Association 2001 Annual Conference Proceedings.
Weinberg, H. S., S. W. Krasner, S. D. Richardson and A. D. Thruston (2002). The Occurrence of Disinfection by-Products (DBPs) of Health Concern in Drinking Water: Results of a Nationwide DBP Occurrence Study. Athens, GA. **EPA/600/R-02/068** 1-460.
WERF (2001). Water Quality Models: A Survey and Assessment. Alexandria, WERF.
Whelan, G., N. D. Tenney, M. A. Pelton, A. M. Coleman, D. L. Ward, J. G. Droppo, Jr., P. D. Meyer, K. E. Dorow and R. Y. Tairan (2009). "Techniques to Access Databases and Integrate Data for Hydrologic Modeling." Retrieved April 28, 2011; http://www.pnl.gov/main/publications/external/technical_reports/PNNL-18244.pdf.
Williams, D. T., G. L. LeBel and F. M. Benoit (1997). "Disinfection by-Products in Canadian Drinking Water." Chemosphere **34**(2): 299-316.
Worm, G. I. M., G. A. M. Mesman, K. M. van Schagen, K. J. Borger and L. C. Rietveld (2009). "Hydraulic Modelling of Drinking Water Treatment Plant Operations." Drinking Water Engineering and Science **2**: 15-20.
Worm, G. I. M., A. W. C. van der Helm, T. Lapikas, K. M. van Schagen and L. C. Rietveld (2010). "Integration of Models, Data Management, Interfaces and Training Support in a Drinking Water Treatment Plant Simulator." Environmental Modelling and Software **25**: 677-683.
Xia, R., D. Borah and M. Bera (2001). Modeling Agricultural Chemical Transport in Watersheds. Bridging the Gap: Meeting the World's Water and Environmental Resources Challenges, Illinois State Water Survey, 2204 Griffith Drive, Champaign, IL 61820, ASCE.