2009-02-27

Stochastic Spatio-Temporal Uncertainty in GIS-Based Water Quality Modeling of the Land Water Interface

Ahmad Mohamad Salah
Brigham Young University - Provo

STOCHASTIC SPATIO-TEMPORAL UNCERTAINTY IN

GIS-BASED WATER QUALITY MODELING

OF THE LAND WATER INTERFACE

By

Ahmad M. Salah

A dissertation submitted to the faculty of

Engineering and Technology

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Civil and Environmental Engineering

Brigham Young University

April 2009

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a dissertation submitted by

Ahmad M. Salah

This dissertation has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date	E. James Nelson, Chair
Date	A. Woodruff Miller
Date	Norman L. Jones
Date	Alan K. Zundel
Date	Gustavious P. Williams

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the dissertation of Ahmad M. Salah in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

E. James Nelson
Chair, Graduate Committee

Accepted for the Department

Steve E. Benzley
Department Chair

Accepted for the College

Alan R. Parkinson
Dean, Ira A. Fulton College of Engineering and Technology

ABSTRACT

STOCHASTIC SPATIO-TEMPORAL UNCERTAINTY IN

GIS-BASED WATER QUALITY MODELING

OF THE LAND WATER INTERFACE

Ahmad M. Salah

Department of Civil and Environmental Engineering

Doctor of Philosophy

Integrated water resources management has been used for decades in various formats. The limited resources and the ever growing population keep imposing pressure on decision makers to better-, and reliably, manage the available waters. On the other hand, the continuous development in computing and modeling power has helped modelers and decision makers considerably. To use these models, assumptions have to be made to fill in the gaps of missing data and to approximate the current conditions. The type and amount of information available can also be used to help select the best model from the currently available models. Advances in data collection have not kept up to the pace of advances in model development and the need for more and reliable input parameter values. Hence, uncertainty in model input parameters also needs to be quantified and addressed.

This research effort develops a spatially-based modeling framework to model watersheds from both water quantity and quality standpoints. In this research, Gridded Surface Sub-Surface Hydrologic Analysis (GSSHA) and CE-QUAL-W2 models are linked within the Watershed Modeling System (WMS); a GIS interface for hydrologic and hydraulic models, to better handle both models pre and post processing. In addition, stochastic analysis routines are developed and used to examine and address the uncertainty inherent in the modeling process of the interface between land and water in the designated watershed.

The linkage routines are developed in WMS using C++. The two models are linked spatially and temporally with the general direction of data flow from GSSHA to CE-QUAL-W2. Pre-processing of the CE-QUAL-W2 model is performed first. Then stochastic parameters and their associated distributions are defined for stochastic analysis in GSSHA before a batch run is performed. GSSHA output is then aggregated by CE-QUAL-W2 segments to generate multiple CE-QUAL-W2 runs. WMS then reads the stochastic CE-QUAL-W2 runs upon successful completion for data analysis. Modelers need to generate a WMS Gage for each location where they want to examine the stochastic output. A Gage is defined by a segment and a layer in the CE-QUAl-W2 model. Once defined, modelers are able to view a computed credible interval with lower, upper bounds in addition to the mean time series of a pre-selected constituent.

Decision makers can utilize this output to better manage watersheds by understanding and incorporating the spatio-temporal uncertainty for the land-water interface.

ACKNOWLEDGMENTS

بسم الله الرحمن الرحيم

"قَالُوا سُبْحَانَكَ لَا عِلْمَ لَنَا إِلَّا مَا عَلَّمْتَنَا إِنَّكَ أَنْتَ الْعَلِيمُ الْحَكِيمُ"

صدق الله العظيم

البَقَرَة 32

"They say: Glory to Thee: of knowledge we have none, save what Thou has taught us: in truth it is Thou Who are perfect in knowledge and wisdom."
(2:32)

It has been a great educational and learning experience. Thanks GOD for gifting me with such exceptional family presence, superb educational guidance, and eventful social life.

Dedication: To my wonderful mum.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

Water quality is a fundamental component in today's holistic approach to water resources management. Water scarcity is a worldwide crisis even though 29% of the total planet surface is covered with water (Figure 1-1). Only 3% of this water is considered "fresh", of which only 0.3% is available in rivers/lakes (Figure 1-2) (WWC, 2007). The primary reason that 97% of the planet water is not available for human use is its unsuitability from a water quality perspective (WHO, 2004). Thus, even though there might be an abundance of water, it is not readily available for human uses and water quality becomes crucial.

In brief, the water shortage crisis has five main causes (WHO/UNICEF, 2005) & (Shiklomanov, 1999):

1. Shortage of the resources (3% of fresh water).
2. Drastic population growth which exerts a continuously increasing demand.
3. Imbalanced distribution of water shares where water-rich regions have high per capita water shares (low population relative to high, readily-available water resources) and water-poor regions have low water shares (high population relative to low, readily-available water resources).
4. Industrialization and urbanization constituting increased pollution potential.

5. Regional water use imbalances. For example, water that is suitable for irrigation may not be suitable for domestic use, or industrial cooling water may not be used as process water.

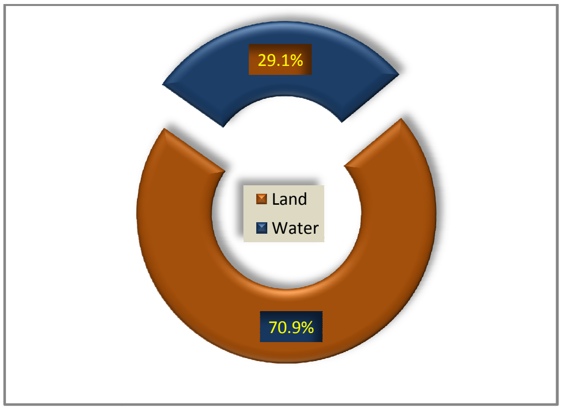


Figure 1-1: Percentage of Land/Water Coverage of Earth.

One of the earliest proposed solutions to the water quality problem was dilution and was referenced by the common phrase: "The Solution to Pollution is Dilution". This concept worked with low population densities. But, with the current requirements and regulations, it might not appear to be a technical resolution; in fact, it might not even be a valid criterion, as the question is if there is enough fresh water to dilute polluted water in all polluted regions in the world. Moreover, adequate water quality data are needed to see if the available water is "good enough" for dilution.

Sufficient and reliable water quantity/quality data are necessary for any comprehensive water resources management project. However, insufficient data has always been an issue in pursuing successful water quantity/quality simulations (Karamouz, et al., 2003). Nevertheless, decisions still need to be made based on the best available technologies and information at the time of decision making.

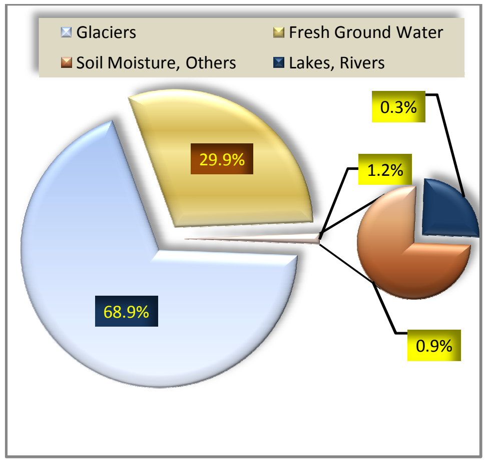


Figure 1-2: Fresh Water Distribution.

In October 1997, President Clinton announced the "Clean Water Action Plan" to clean up the polluted water bodies of the United States. In his memo, the President emphasized the "watershed approach", which combines the analysis of both point and non-point source pollutants over an entire region, as opposed to concentrating on just direct discharges to an impaired water body (WEF, 1998). Even though the "watershed approach" can be modeled with an integrated water resources modeling framework, most models do not include robust uncertainty tools to account for the inherited underlying uncertainty in the hydrologic arena.

1.1 Research Drive

Society is concerned with maintaining a good quality water resources locally, nationally and regionally. A water body's quality can be continuously examined on two scales; temporal and spatial. On spatial scales, concentration maps and/or grids of a certain water quality parameters can be generated for the area/profile of interest and the associated water bodies at a specific point of time (Figure 1-3).

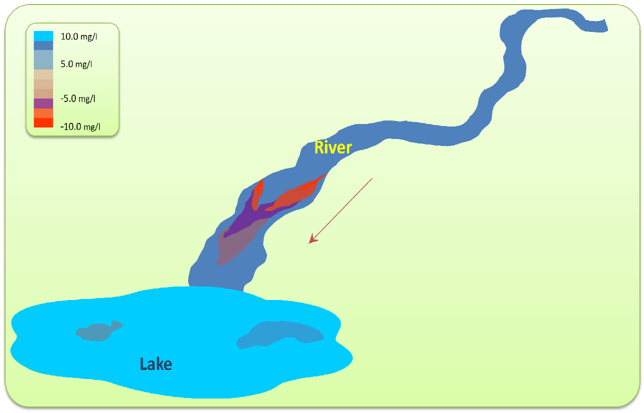


Figure 1-3: Hypothetical Spatial Concentration of a Pollutant.

Temporal analysis can be made for a time period of interest at a specific location. For example, in many studies, researchers are interested in examining pollutant concentrations over a specific time frame at a specific location (Figure 1-4).

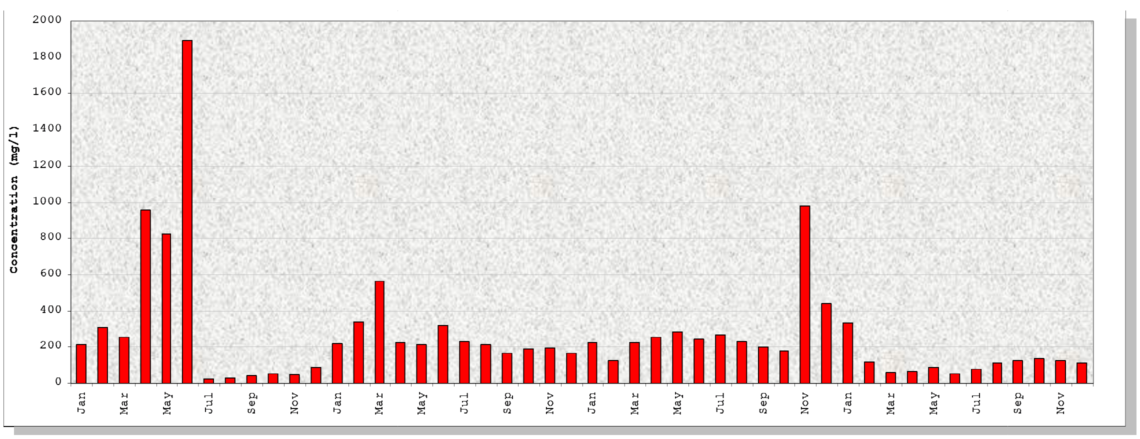


Figure 1-4: Hypothetical Temporal Concentration of a Pollutant.

In some cases both dimensions; spatial and temporal, are combined together to form presentations called "change detection maps" which depict changes over a given time period of pollution levels, for a certain pollutant, in an area of interest (Salah, 1999).

For these three cases, model outputs are typically a single value for a location (in the case of a map) or a single point in time (for a time series). The most important advantage of this deterministic approach is that it provides an easy and, relatively, quick capture of information. Yet, it does not address uncertainty nor does it provide a probable range of values on spatial or temporal dimensions that may be useful in making decisions, especially in a multi-criteria decision making process.

The contributing sources to pollution are either point or non-point. Their impacts must be modeled as accurately as possible to estimate the overall conditions of any water body under investigation. Generally, there are not enough data or there are uncertainties involved in estimating pollution loads. These uncertainties should be addressed and adequately incorporated in the modeling process to improve decision making.

Current practices (chapter 2) and this research, indicate a need to integrate a distributed land water quantity/quality model with a complex hydrodynamic and water quality model. This integration could provide a base for a holistic integrated basin-wide management scheme. Since most of the distributed land models do not have comprehensive hydrodynamic/water quality sub-routines (chapter 2), it is almost always advisable to link them to a more complex river/lake model to provide the detailed information required for basin-wide management decisions.

1.2 Land-Water Interface

The land-water interface is where point and non-point pollution sources from man-made or natural activities on the land come into contact with surface water (i.e. lakes, reservoirs, streams,…etc) either directly through surface runoff or indirectly through groundwater. Linking a land model to a water body model at this boundary line is a major step in developing an integrated water quantity/quality modeling process and poses some technical problems.

This research utilizes the Gridded Surface Subsurface Hydrologic Analysis (GSSHA) (Downer, et al., 2006), as the land model, and CE-QUAL-W2 (Cole, et al., 2007), as the water body model, to quantitatively model the land water interface. The Watershed Modeling System (WMS) (Nelson, 2008) developed at Aquaveo, L.L.C., formerly the Environmental Modeling Research Laboratory (EMRL) at Brigham Young University (BYU), incorporates these two models as water resources/quality tools and was used as the integration platform in this research.

To achieve this integration, GSSHA; two dimensional hydrologic model, models the land-based activities throughout the watershed. GSSHA output provides boundary conditions to CE-QUAL-W2 to model the receiving water body. CE-QUAL-W2 output is used to estimate water quantity and quality parameters at specific watershed locations.

This approach can be used to develop information for managing the watershed and to examine the effects of different management practices. It can also be used for various basin management approaches such as, Best Management Practices (BMP) or Total Maximum Daily Loads (TMDL) studies (US-EPA, 2002).

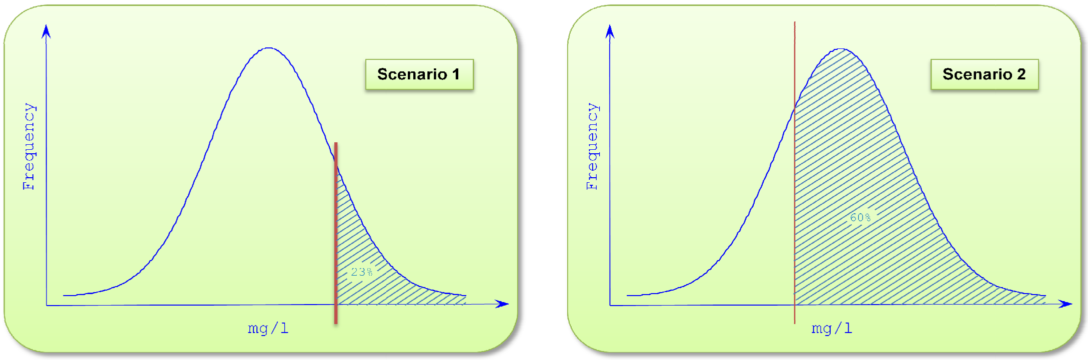
A TMDL is a regulatory standard that defines the maximum load of a pollutant that a water body can receive and still meet a given water quality standard. It requires an analysis of uncertainty. A TMDL, for a given water body is computed as the sum of the allowable loads of a single pollutant from all contributing point and non-point sources in the watershed for the chosen water body (US-EPA, 2005). The TMDL estimate must include a margin of safety (MOS), and must account for seasonal variations in order to ensure that the water body can be used for the purposes it is designated for, such as drinking water supply, contact recreation (swimming), and aquatic life support (fishing) (US-EPA, 1999-a).

1.3 Management Scenarios

In general, models are used to simulate real world situations and results are used for management decisions. To use these tools to incorporate uncertainty, multiple scenarios are established using different management practices or TMDL plans. Each scenario is run multiple times in the two models; i.e. land and water models, while varying certain parameters based on the stochastic nature of these parameters. The resulting data can be used to obtain a probability density function (PDF) of a desired constituent for every scenario. This PDF can be used to estimate the probability of having a concentration exceeding a certain value for the plan under investigation (Figure 1-5).

Figure 1-5 - A shows a PDF for the concentration of a constituent indicating 23% as a probability of exceeding a certain concentration threshold, while Figure 1-5 - B shows a 60% probability for the same constituent but with a different management scenario.

PDFs can be generated for water quantity parameters as well, i.e., volume of a reservoir, discharge of a river at a certain location and so forth. Multiple PDFs can be generated for the same scenario at different times. In this case, the PDFs can be used to generate a "Probability Density Surface" (PDS) which is a probability surface representing all PDFs in time or space for the parameter value. The height of this surface is the frequency of occurrence in combination of time and frequency. This generalization of the obtained PDFs to produce a PDS can be perceived as adding a third dimension to the PDF. This third dimension can be time, to obtain a “time-series PDF” or distance to obtain a PDF of a certain parameter along the center line of a river. (Figure 1-6).



(A) (B)

Figure 1-5: Probability Density Functions of a Constituent Concentration.

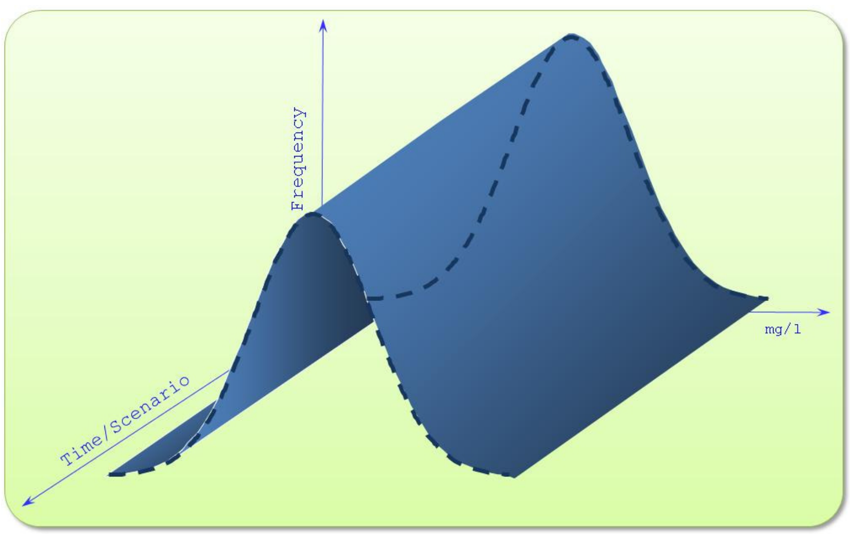


Figure 1-6: A Probability Density Surface (PDS).

1.4 Stochastic Modeling

Most of the data obtained and used as input to water quantity and quality models, are of a stochastic nature and should not be used in a deterministic modeling approach without addressing uncertainty. Addressing uncertainty for such operations may be done at various levels. One of which is to have models use their output to compute a "probable range", rather than a "single value" answer.

An important factor in determining the probability or the threshold value required for a decision is the level of accuracy desired (Ramsey, et al., 2002). Modelers may spend more time and effort for cases with high level of accuracy needed. Alternatively, with the same, or less, effort, time and money, modelers, with the aid of PDFs or PDSs, can analyze and incorporate multiple results into the decision process. For instance, results could infer that there is a 95% probability that the total nitrogen level at the outlet of a lake ranges between 5.7 and 11.1 mg/l, instead of a single value of 8.2 mg/l. The last value might not even be the mean of the obtained PDF and hence it would be an inaccurate representation of the population in a deterministic approach (Figure 1-7).

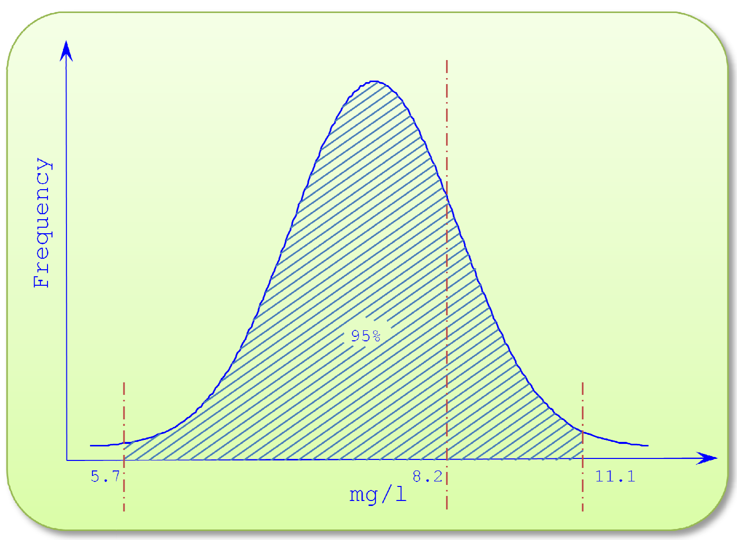


Figure 1-7: Probable Range of a PDF.

Results could also show a similar, but narrower, 90% credible interval and so forth. PDFs can also be used to determine the probability of exceeding a certain value such as a water quality standard or a spillway elevation. This allows management plans to be better analyzed.

Water resources decisions often depend on water quality information (Calder, 2005). However, water quality data are not readily available, and rarely are they available at the specific site where decisions need to be made (Takyi, 1995). An important need for water quality estimates exists but the data necessary to support this need is limited. Water quantity and quality should always be perceived as two faces for one coin; both are indispensable factors in integrated water resources management.

1.5 Objectives

The goal of this research is to develop methodologies and implement tools as a proof of concept, to incorporate uncertainty in an integrated modeling framework for managing water resources for both quantity and quality. Specifically, the research objectives can be listed as follows:

1. Use stochastic analysis to examine uncertainty in integrated water resources modeling on spatial and temporal scales and characterize the effect that insufficient data has on model output. This was done using a few selected water quantity/quality parameters.
2. Conceptualize an integrated spatially-based modeling technique to model watersheds from both a water quantity and quality standpoint. This approach will be able to incorporate existing data sets such as elevation, land use, soil that can be used interactively to develop a complete integrated water resources picture for a given watershed.

3. Link GSSHA to CE-QUAL-W2.
4. Develop uncertainty analysis sub-routines (within WMS) that enable modelers to compute and analyze time series credible intervals, rather than a single value for each time step, of pollutants associated with probability levels.

1.6 Dissertation Layout

The current research effort is organized in the following chapters:

1. Chapter 1 introduces the work, states the research need and lists the objectives of this research.
2. Chapter 2 illustrates current research effort that is related to the outlined problem.
3. Chapter 3 covers the methodology followed to reach the stated research objectives.
4. Chapter 4 presents the results obtained and discusses applicability.
5. Chapter 5 summarizes the conclusions and lists recommendations for further research.

2 Contemporary Research

Operation policies for water resources systems should not be implemented without forecasting the future state of the resources. For example, consider a reservoir that supplies water for multiple purposes; the amount of each scheduled release depends on the probable range of inflow to the reservoir. Because of the lack of adequate data to characterize physical processes in a hydrologic system, many investigators have expanded the application of statistical models to generate synthetic data for use in forecasting. Synthetic data also help by incorporating uncertainties and probable extreme events (Karamouz, et al., 2003).

2.1 Integrated Water Resources Modeling

Water resources management is mainly aimed at making resource allocation decisions which include mitigating or preventing the adverse effects of excessive runoff and water shortage. Hydrologic modeling serves as a valuable tool in water resources management (Calder, 2005). Simulating the hydrologic and water quality behavior of a watershed can be used to predict the impacts of proposed land use changes and to evaluate management strategies on both a short and long term basis. Water resources/quality models have improved considerably over the last decade which helps to improve the reliability of model output and gives water resources engineers a better understanding of real water systems (McBride, 2005).

The concept of a watershed (Figure 2-1) is the foundation for hydrologic planning and design. An understanding of watershed processes is a basic requirement for understanding freshwater resources all over the world. Sound water resource management must be undertaken at a whole-watershed level, rather than just a local point-source contamination level (Melching, et al., 2001). A watershed can be regarded as a closed system with well-defined boundaries and elements with clear relationships both structurally (in terms of morphology) and functionally (by virtue of the flow of matter and energy). The inputs and outputs across the watershed boundary can also be clearly distinguished (Karamouz, et al., 2003).

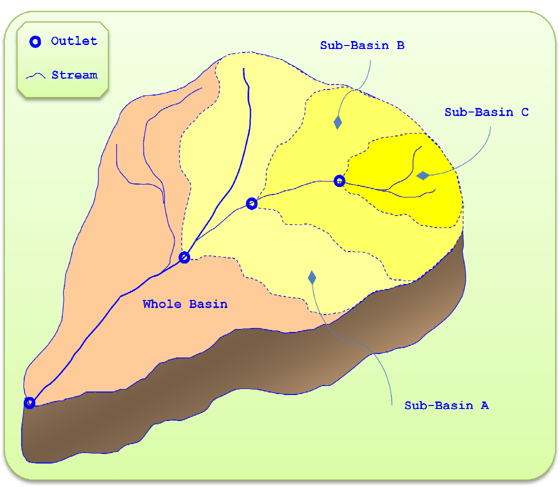


Figure 2-1: Watershed Basins and Sub-Basins Sizing Concept.

Qualitative characteristics of water are also important for water resources planning and management. Water quality is evaluated in terms of physical, chemical and microbiological properties. Quantitative measurement of these properties is necessary to determine water quality. Water resource engineers set water quality requirements to help manage these resources. Water quality requirements are determined based on the intended use of water; for example, water contaminated by chemicals might reduce crop yield but could be suitable as industrial cooling water. Therefore, the water quality requirements for each type of water use should be determined, along with assessment of raw water quality and selection of suitable water treatment processes, if necessary when making management decisions. (Karamouz, et al., 2003)

The quality of surface and ground water resources can significantly affect water use in many regions. In regions where pollutants from human activities have critically degraded water quality, the main issue in water quality management is to control pollution sources. Control level requirements depend on the water quality standards defined for the various water uses. Water quality management is different from water quantity management, which is the engineering of water resources systems so that enough water will be provided to all potential modelers within a region (Krenkel, et al., 1980). Water quality modeling, like other modeling efforts, involves unavoidable uncertainty which can contribute to inefficient, and in some cases inadequate, decisions affecting society environmentally and financially if the uncertainty is not included in the decision making process.

2.1.1 Watershed Concept

The concept of modeling the watershed as a whole, rather than locally, can be applied to any watershed or sub-watershed and is not dependant on the catchment size or any other geomorphologic aspect of the watershed. Figure 2-1 shows how a sub-catchment can still be modeled as a whole watershed, even while the whole watershed can be modeled as one unit. Selecting the outlet location defines the watershed boundaries. An outlet should be chosen based upon the management perspective of the study.

For Geographic Information System (GIS)-Integrated models, especially in water resources, watershed reporting is mainly the compilation and output of information on the selected watersheds. It involves modeler supplied data, tools and models that can answer the research questions of interest. Ultimately, the goal is to provide decision makers with accurate and reliable information that can help make water resources related decisions. The GIS approach of water resources modeling has existed for a substantial time. The nature of the water resources and quality modeling is aided by the use of GIS in one form or another. Conventional methods used various maps in paper format and used approximation techniques for the various geo-processing tools that are currently available in computer format. Using the digital, as opposed to the "manual" form of GIS entails many benefits, some of which are as follows:

1. Easier implementation.
2. Automatic computed parameters.
3. Conventional overlay operations.
4. Easier to understand.
5. Requires consistent raw data.
6. Models point and non-point source pollutions simultaneously.

The Better Assessment Science Integrating point and Non-point Sources (BASINS) software system is an example of the integrated watershed and water quality management software. BASINS (US-EPA, 2007-a) is a multipurpose environmental analysis system designed for use by regional, state, and local agencies in performing watershed and water quality-based studies. This system makes it possible to quickly assess large amounts of point source and non-point source data in a format that is easy to use and understand.

BASINS allows the modeler to assess water quality at selected stream sites or throughout an entire watershed. This tool integrates environmental data, analytical tools, and modeling programs to support development of cost-effective approaches to watershed management and environmental protection, including TMDLs.

BASINS includes pre-processing tools, custom databases and a set of standardized modeling tools including HSPF, SWAT, QUAL2K (formerly QUAL-2E), PLOAD, Aquatox and PEST among other models. However, the only model that has an interface in BASINS-4.0, currently, is HSPF. BASINS-4.0 utilizes the open source programmable GIS platform MapWindow. It is intended that BASINS 4.0 will interoperate with ArcView 3.x and ArcGIS 8.x. Access to data for BASINS is web-based, which makes it more efficient. Users do not have to store large volumes of data in local drives. Instead, data is accessible as needed.

2.1.2 Analysis versus Synthesis

Like most of the basic sciences, hydrology requires both analysis and synthesis to use the fundamental concepts in the solution of engineering problems (McCuen, 2005). The word *analysis* is derived from the Greek word *analusis* which means “a dissolving” or “to break apart”. Analysis can be compared with the word synthesis. The word synthesis comes from the Latin word *suntithenai*, which means “to put together”. Because of the complexity of most hydrologic engineering design problems, the fundamental elements of the hydrologic sciences cannot be used directly (McCuen, 2005). Instead, it is necessary to take measurements of the response of a hydrologic process and analyze them in an attempt to understand how the process functions. Quite frequently, a model is formulated on the basis of the physical concepts that underlie the process, and the fitting of the model to the measurements provides the basis for understanding how the physical process varies as the input to the process varies. After the measurements have been analyzed (taken apart) to fit the model, the model can be used to synthesize (put together) design rules.

2.1.3 Total Maximum Daily Load (TMDL)

Regulatory standards for watersheds in the U.S. are based in part on TMDL requirements. As a result, the focus of water quality management for nutrients like phosphorous has moved from end of the pipe or point source control to watershed-scale analyses that incorporate point and non-point source pollution assessment (Shoemaker, et al., 2003).

The TMDL process includes the following key steps (US-EPA, 1999-b):

1. Standards setting which involves specifying designated uses, and selecting appropriate water quality criteria with numeric targets for the water body;
2. Water body assessments for impairment listing;
3. Watershed assessment as linkage analyses that associate pollutant sources with water quality targets resulting in an estimate of the loading capacity of the water body-the TMDL;
4. Planning and allocating the loading capacity among point and non-point sources; and finally

5. Implementation of control actions to reduce pollutant loading to the water body such as adopting better technologies and operational rules.

One part of a TMDL study is the Use Attainability Analysis (UAA) which will be discussed in Section 2.1.4.

2.1.4 Use Attainability Analysis (UAA)

UAA is a structured scientific assessment of the factors affecting the use attainment for a water body, such as swimming, fishing and drinking. A UAA is the tool used to evaluate the potential to remove non-existing and non-attainable designated uses or to establish subcategories of uses (Washington, 2005).

UAA provides the means for setting new standards and revising or refining existing ones. However, like MOS, UAA is not widely used because of lack of technical guidance. For this reason, UAA is either arbitrarily employed or not used at all, leading to water bodies being falsely listed and efforts wasted in developing TMDLs on the basis of inappropriate water quality standards (Olufemi, et al., 2003).

2.2 Watershed Modeling System (WMS)

This research effort will use WMS as a pre and post processor GUI for the two models, CE-QUAL-W2 and GSSHA. The development of WMS is partly supported by the United States Army Corps of Engineers (US-ACE) through research funds to Aquaveo, L.L.C. (formerly EMRL of BYU). GSSHA, developed by the US-Army Corps of Engineers, and CE-QUAL-W2, developed by the US-Army Corps of Engineers and Portland State University, are used in this research. As indicated in Chapter 4, the underlying research would be applicable for other models as well.

WMS (Figure 2-2) is a comprehensive graphical modeling environment for various phases of watershed hydrology, hydraulics and water quality. WMS includes tools to automate modeling processes such as automated basin delineation, geometric parameter calculations, GIS overlay computations (land use, soil, rainfall depth, HSPF segments, etc.), and cross-section extraction from terrain data. WMS, version 8.0, supports hydrologic modeling with HEC-1 (HEC-HMS), TR-20, TR-55, Rational Method, NFF, MODRAT, and HSPF. Hydraulic models supported include HEC-RAS, SMPDBK, and CE-QUAL-W2 (Nelson, 2008). Two-dimensional integrated hydrology (including channel hydraulics and groundwater interaction) can now be modeled with GSSHA (Downer, et al., 2006). All of this in a GIS-based data processing framework makes the task of watershed modeling and mapping easier and reliable (EMS-i, 2006).

As indicated in Figure 2-2, in the pre-processing phase, modelers gather data including maps, databases and digital terrain models and other types of field-gathered data in an effort to try to extract the necessary model input. Inputs may be generically prepared for any model or be generated specifically for a model through the model interface within WMS. WMS has various tools to easily extract model-required input. As an example, there are tools in WMS to extract cross sectional data from an underlying digital elevation model (DEM) or triangulated irregular network (TIN). A major portion of the effort in this pre-processing phase is to make sure that all data gathered from different sources overlay properly, both horizontally and vertically.

WMS allows modelers to generate model input files and import pre-generated model input files. For some models, WMS enable modelers to verify that the input file will not generate any warnings or errors running the model. In other cases, models need more than one input file. Currently, WMS generates the spatial-related input files and does not generate the meteorological files for CE-QUAL-W2.

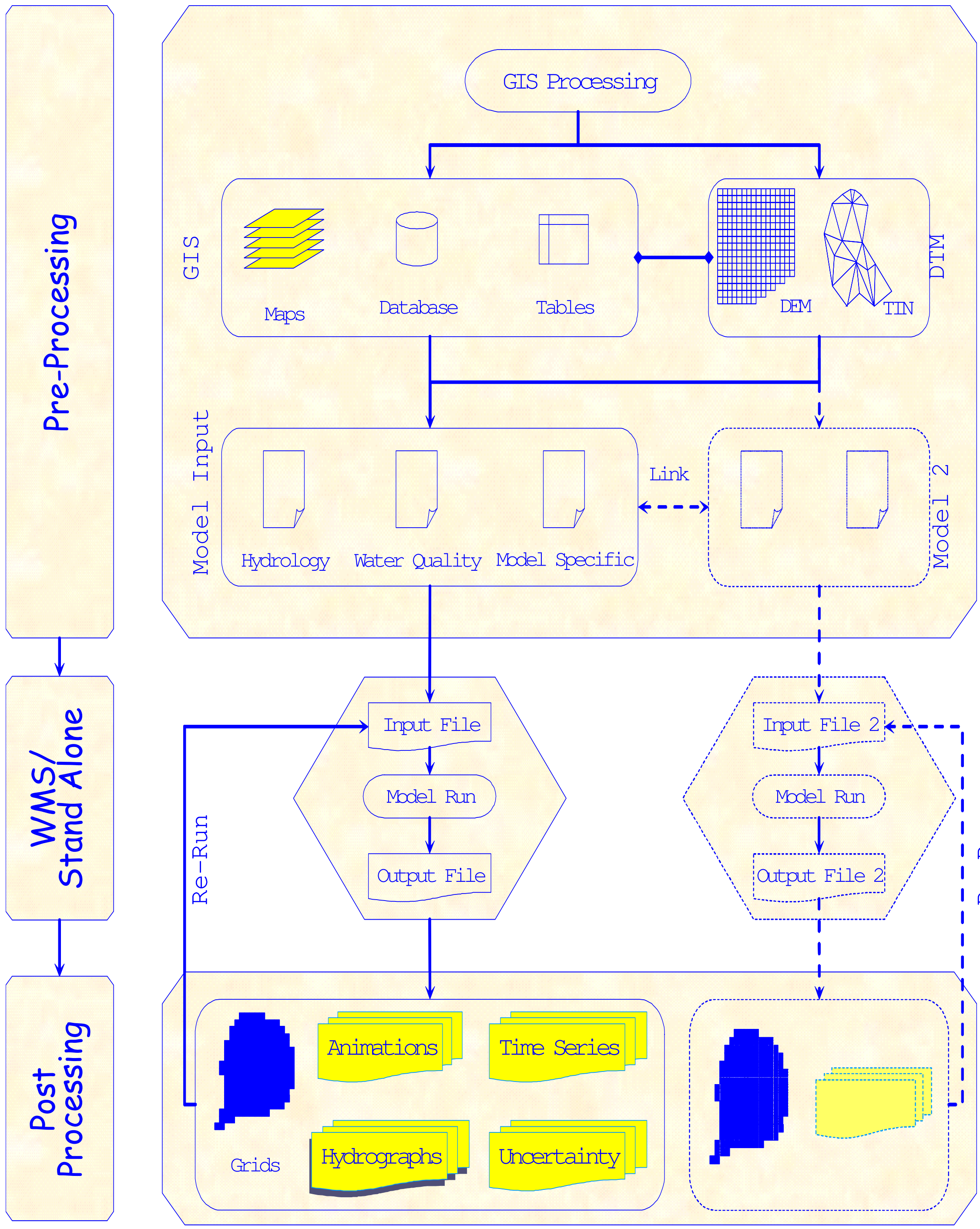


Figure 2-2 : Conceptual Representation of the Watershed Modeling System.

Models can be run from their interfaces within WMS, if applicable, or through the stand-alone version of the model. Once run, model output may be read by WMS post-processing tools to generate data sets, grids, film loops, time series graphs, and hydrographs. In some cases, models may be linked through a uni- or bi-directional data exchange. WMS links some models together providing an integrated water resources framework. There are two types of linkage:

1. Subsequent linkage (uni-directional): in which model "A" needs to run first before model "B" can start running.
2. Interactive linkage (bi-directional): in which data interchange between the two models happens within both runs.

An example of the subsequent linkage in WMS is the stochastic linkage between HEC-1 (US-ACE, 1998) and HEC-RAS (US-ACE, 2006). In this linkage, HEC-1 runs multiple times before the output is used in HEC-RAS, which, in turns, runs multiple times. This linkage is used to generate flood probability maps (Smemoe, 2004).

Similar to BASINS, WMS uses a collection of hydrologic, hydraulic and water quality models to simulate the integrated behavior of a given watershed. Nevertheless, WMS uses a different set of models to achieve the same goals. While BASINS uses PLOAD, HSPF, SWAT, QUAL2E and GENSCN (US-EPA, 2007-a), WMS uses HSPF, GSSHA, CE-QUAL-W2 and customized post-processing tools.

2.3 CE-QUAL-W2

CE-QUAL-W2 Version 3.5 (Cole, et al., 2007) is a two-dimensional water quality and hydrodynamic model capable of modeling water bodies with interconnected rivers, reservoirs and estuaries. A typical model domain is shown in Figure 2-3. The model is based on solving the two dimensional unsteady hydrodynamic and advective-diffusion equations.

Historically, starting with version 3.1, CE-QUAL-W2 allows modelers to include riverine branches in conjunction with reservoir and estuary branches. The current version also allows the modeler to insert hydraulic elements between branches (pipes, weirs, spillways and gates with dynamic gate openings), use up-to-date re-aeration (including spillway effects) and theoretical evaporation models. In addition, the modeler can view model results graphically as they are being computed, use a variety of turbulence closure schemes, insert internal weirs in the computational domain, use the updated numerical scheme ULTIMATE-QUICKEST for advective transport of mass/heat, add float-activated pumps, use a dynamic vegetative and topographic controlled shading algorithm, and include a user-defined number of algal, epiphyton/periphyton, carbonaceous biological oxygen demand, suspended solids, and generic model water quality constituents. (Wells, 2002)

CE-QUAL-W2 is jointly-developed by US-ACE, and Portland State University. It is one of very few two dimensional water quality and hydrodynamics models currently available. It is used by several federal, state, private and international agencies to perform hydrodynamics studies of dam operations, eutrophication, dissolved oxygen, other water quality issues as well as TMDL watershed processes (Wells, 2002).

The current version of CE-QUAL-W2 model runs in a stand-alone interface (Figure 2-4) that is distributed with the generic version of the model. This interface helps CE-QUAL-W2 modelers monitor the progress of their models as they run. WMS modelers generate the control and bathymetry files within WMS and then run CE-QUAL-W2 outside of WMS. Upon completion of a successful run, the output files are generated where they are specified in the CE-QUAL-W2 control file. WMS modelers can then use WMS to read the solution and perform further post-processing efforts.

2.3.1 CE-QUAL-W2 vs. EPA Models

EPA, through the Watershed and Water Quality Modeling Technical Support Center (US-EPA, 2007-c), sponsors the development of models to address similar hydrodynamics and water quality issues. Their watershed models include WCS, SWMM, WARMF while the water quality line of models include WASP, QUAL2K , Aquatox and their hydrodynamic ones include EFDC and EPD-RIV1.

The Environmental Fluid Dynamics Code (EFDC) is a 3-dimensional hydrodynamic model that uses Cartesian, curvilinear or orthogonal horizontal coordinates to represent the characteristics of a water body (US-EPA, 2007-b). EFDC is not designed to perform water quality modeling. The Water Quality Analysis Simulation Program (WASP) utilizes the EFDC hydrodynamic output for water quality modeling in the receiving water body.

QUAL2K is a 1-dimensional river and stream water quality model that is a modernized version of the QUAL2E model. It uses MS-Excel as a graphical user interface (US-EPA, 2007-c).

Watershed Analysis Risk Management Framework (WARMF) is a decision support system designed to facilitate TMDL analysis and watershed planning. WARMF is compatible with the data extraction and watershed delineation tools of BASINS.

WARMF is organized into five (5) linked modules under one, GIS-based graphical user interface (US-EPA, 2007-c).

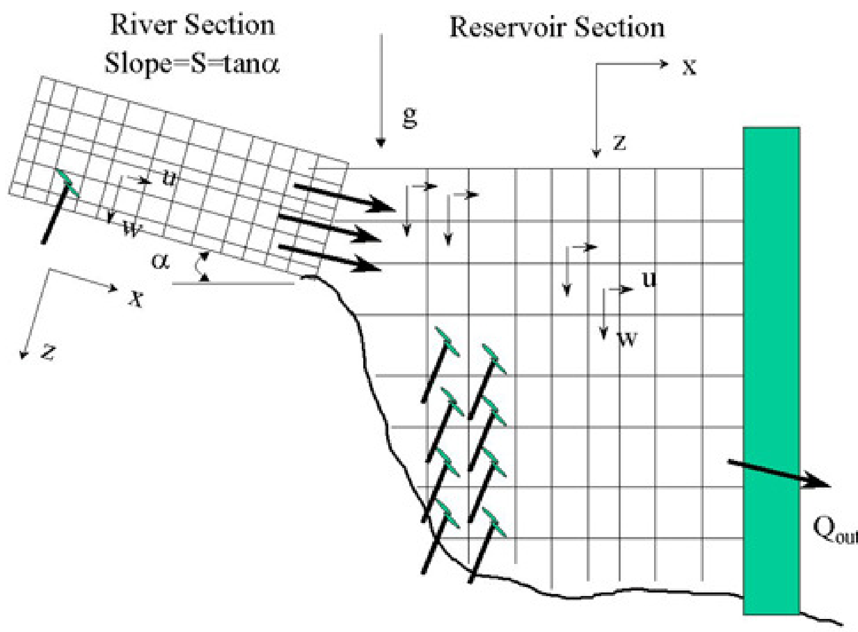


Figure 2-3: CE-QUAL-W2 Model Grid. Source: (Cole, et al., 2007).

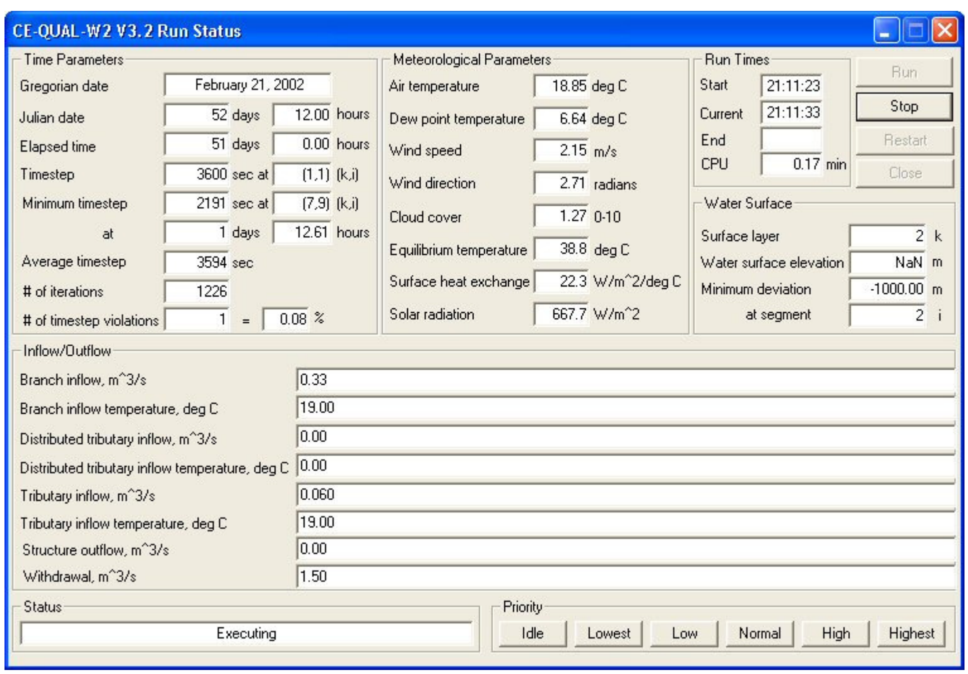


Figure 2-4: Stand-Alone Interface for CE-QUAL-W2.

EPD-RIV1 is a collection of programs to perform one-dimensional water quality and hydrodynamic models. It is based on the CE-QUAL-RIV1 model developed by the US-ACE. CE-QUAL-RIV1 does not have the ability to characterize the hydraulics or water quality of deeper reservoir systems or deep river pools that stratify (US-ACE, 1995).

2.4 Girded Surface Subsurface Hydrologic Analysis (GSSHA)

GSSHA is a two dimensional physically based, distributed-parameter, structured grid, hydrologic model that simulates the hydrologic response of a watershed subject to given hydro-meteorological inputs (Downer, et al., 2006).

Major components of the model include spatially and temporally varying precipitation, snowfall accumulation and melting, precipitation interception, infiltration, evapotranspiration, surface runoff routing, simple lake storage and routing, unsaturated zone soil moisture accounting, saturated groundwater flow, wetland peat layer hydraulics, overland sediment erosion, transport and deposition, in-stream sediment transport, and overland contaminant transport, and uptake.

In GSSHA, each process has its own time-step and an associated update time. During each time-step, the update time of each process is checked against the current model time. When they coincide, the process is updated, and updated information from that process is transferred to dependent processes. The time-step of dependent processes may be modified as part of the process update. This formulation permits the efficient simultaneous simulation of processes that have dissimilar response times, such as overland flow, evapotranspiration (ET), and lateral groundwater flow. This scheme also allows an integrated solution of processes coupled through boundary conditions and flux exchanges (Byrd, et al., 2005).

2.4.1 GSSHA vs. HSPF

The Hydrologic Simulation Program-Fortran (HSPF) is a semi-distributed continuous model that combines spatially distributed physical attributes into hydrologic response units (HRUs). Each HRU, in response to meteorological inputs and storage capacity factors, is assumed to behave in a uniform manner. In an essence, the geometric representation of a watershed in a GSSHA model may be coarser than its HSPF equivalent. HRUs are based on morphology and in general they are large polygons. GSSHA grid cell size may be unrealistically large and/or an HRU may be too small.

Surface runoff is simulated primarily as an infiltration-excess (Hortonian) process (Bicknell, et al., 2001). HSPF allows modelers to emphasize the hydrologic processes that are dominant in a watershed by specifying the major characteristics used to define HRUs, such as soil type or land use, and by adjusting parameter values during calibration. Although selection of parameter values that reflect watershed specific physical processes can improve model calibration, estimation of actual parameter values from physical measurements is difficult (Albek, et al., 2004). Therefore, optimum parameter values are generally obtained through the calibration process (Johnson, et al., 2003).

Previous research (Johnson, et al., 2003); (Salah, et al., 2005-a) indicated that HSPF requires cumbersome calibration process in the simulation of stream flow during summer period or in arid, semi-arid environments when saturation-excess (as opposed to infiltration excess) is a major factor in runoff generation.

2.4.2 Hydrodynamic and Water Quality Capabilities

Both GSSHA and HSPF can model the entire watershed including hydrodynamic, and water quality, capabilities to model rivers and reservoirs within the watershed. HSPF can be used to model a watershed in a holistic approach.

One of the major strengths resulting from linking CE-QUAL-W2 to GSSHA is the utilization of both extensive hydrodynamic and water quality capabilities of CE-QUAL-W2 linked with the detailed two dimensional distributed modeling tools in GSSHA. However, the statistical methodology proposed in this research can be implemented irrespective of the models used. The statistical methodology is universal and not model-specific. The implementation shown in the following chapters is specific, to these models, for the purpose of demonstration.

Modeling efforts could be carried out in two fashions; i.e. deterministic or stochastic. Each has advantages and disadvantages and may each be appropriate in unique situations.

2.5 Stochastic Processes

A quantitative description of a natural phenomenon is called a mathematical model. A model is usually judged using a single, quite pragmatic, factor, the model's usefulness. There is no such thing as the "best" model for a given phenomenon (Taylor, 1998). The pragmatic criterion of usefulness often allows the existence of two or more models for the same event, but serving distinct purposes.

The word "stochastic" derives from the Greek (στοχάζεσθαϊ: to aim, to guess) and means "random" or "chance". The antonym is "sure", "deterministic" or "certain". A deterministic model predicts a single outcome from a given set of circumstances. A stochastic model predicts a set of possible outcomes weighed by their likelihoods or probabilities (Taylor, 1998). However, phenomena are not in and of themselves inherently stochastic or deterministic. Rather, to model a phenomenon as stochastic or deterministic is the choice of the observer. The choice depends on the observer's purpose; the criterion for judging the choice is usefulness. To be useful, a stochastic model must reflect all those aspects of the phenomenon under study that are relevant to the question at hand. In addition, the model must be amenable to calculation and must allow the deduction of important predictions or implications about the phenomenon.

A stochastic process is a process that can be modeled stochastically. It involves at least one random, scalar or vector, variable that takes random values over time or space. In stochastic process, usually the word "random" is replaced by "stochastic". The set of all possible and observable values that a stochastic variable can take is called "state space" or sometimes called, for simplicity, population. (Lindsey, 2004)

2.5.1 Random and Stochastic Variables

Whenever measurements in any natural system are made, for example of an effluent biological oxygen demand (BOD_5) or a stream flow, some part of their variation cannot be explained and may only be attributed to chance or inherent randomness. Statistical methods offer means of evaluating this randomness in an objective way, rather than the all-too-common confusion of conflicting subjective opinions. They are capable of distinguishing between randomness (noise, random variability) and pattern (seasonality, trend) using repeatable procedures. Because of the role of variability in measurements, statistical methods allow hydrologic and water quality variables to be treated as random, or stochastic, variables. The basic distinction between random and stochastic variables is that a random variable does not imply some natural ordering of the results, while a stochastic variable does. For example, a time series of data at a particular site, or a set of samples down a river at the same time are considered stochastic variables. In either case, the value that a hydrologic or water quality variable may take has at least some element of randomness in it, and it needs to be recognized (McBride, 2005). It would be tidier perhaps to use just one term, but the literature uses both; random and stochastic, with stochastic being more commonly used in hydrologic science.

2.5.2 Probability Density Function (PDF)

A PDF is a statistical term for a frequency distribution constructed from an infinitely large set of values. It is a non-negative function (curve), with an integral of 1, related to a continuous random variable "A". The probability that "A" is less than a specified value "x" is the area under the curve up to the point "x". PDFs can be discrete or continuous. Discrete PDFs are usually called Probability Mass functions (PMF) (Mason, 2001).

To build a PDF for a specific parameter, it is necessary to follow three major steps to obtain a good representing distribution for the parameter under investigation (Bury, 1999); (Salah, et al., 2005-b).

1. First, the variable is examined to determine if it is discrete or continuous.
2. Then the physical process behind it should be inspected before a statistical package is used to suggest some distributions that have statistical relevance to the data.
3. Finally, a single distribution should be selected based on physical relevance to the variable examined (Salah, et al., 2005-b).

The two- and three-parameter gamma distributions have been widely used in hydrology, mainly for the purpose of modeling the frequencies of annual floods. The Log-Normal distribution is also used. Both are broadly similar in that both are skewed with a longer upper tail. Indeed, it is this characteristic which makes them suitable for representing annual floods and other hydrologic variables where skewness is invariably present (Clarke, 1994). Some other researchers indicate that a Beta distribution with specific shape parameter values may be suitable for representing other hydrologic and water quality parameters (Salah, et al., 2005-b).

The general formula (NIST, 2007); (Steeb, et al., 2000) for the probability density function of the Normal distribution is as follows, equation (2-1):

$$f(x) = \frac{e^{\frac{-(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} \qquad (2\text{-}1)$$

where:

$\mu =$ Location parameter,

$\sigma =$ Scale parameter.

The case where $\mu = 0$ and $\sigma = 1$ is called the standard Normal distribution. Thus, the equation for the standard Normal distribution is as follows, equation (2-2):

$$f(x) = \frac{e^{\frac{-(x)^2}{2}}}{\sqrt{2\pi}} \qquad (2\text{-}2)$$

For most earth science observations, a sub-set of the Beta distribution (Figure 2-5) may seem to yield a reasonable approximation to most parameters (Bury, 1999); (Heyman, et al., 1984); (Ricciardi, et al., 2005); (Valdes, et al., 1990). The Beta distribution is a set of continuous probability distributions bounded by the 0.0 and 1.0. The shape of each distribution is determined by the values of the shape factors. The general formula (Press, et al., 2005) ; (Palisade, 2007) for the probability density function of the Beta distribution is as follows, equation (2-3):

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{f(\alpha,\beta)} \tag{2-3}$$

where:

$\alpha, \beta =$ Shape Factors

$f(\alpha, \beta) =$ Normalization Function.

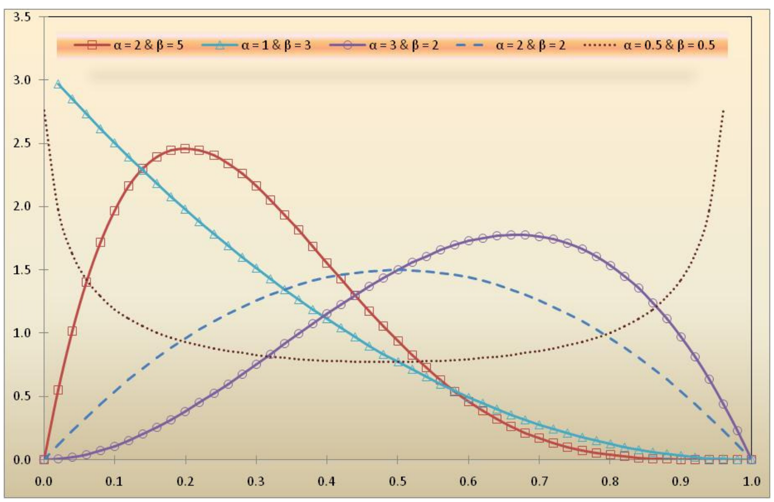


Figure 2-5: Sample Beta Probability Distributions.

The normalization function is dependent on the shape factor values and it can be considered as a normalization factor that sets the total probability of the Beta distribution, which is the area under the curve, to 1.0.

As seen in Figure 2-5 the shape factors affect the Beta PDF considerably. When both factors are set to 1, the PDF becomes almost identical to the standard uniform distribution. With other factors, Beta PDF can approximate a log-normal distribution.

Probabilities obtained from any PDF are used to estimate the corresponding z-score from an inverse cumulative distribution function (CDF) (Wichura, 1988). The formula for the CDF of the Normal distribution does not exist in simple closed formula and it is usually computed numerically (Wichura, 1988).

A CDF is the set of points, each of which equals the integral of a probability distribution starting at the minimum value and ending at an associated value of the random variable. A cumulative distribution is constructed by progressively adding the frequency across the range of frequency distribution (Palisade, 2007).

2.5.3 Population Statistics

For practical reasons, scientists usually use a sample to represent the population under investigation. Consequently, inferences made on this sample should not be generalized to populations other than the parent population investigated. Inferences, and estimates of the required parameter(s), are made based on statistics derived from the sample. A statistic is any quantity that can be calculated from observed data. An estimate is a statistic used to represent the value of the parameter. While the parameter remains unknown, its estimate can be calculated (Ramsey, et al., 2002).

A population mean, which is a statistical parameter, may be represented by the sample average which is in turn an estimate of the mean and therefore it is a statistic. Other sample statistics may include, but are not limited to, maximum, minimum, standard deviation, kurtosis, t-statistic and z-statistic.

Each statistic can be used in specific situations and where the assumptions underlying the equations are not violated. The Z-score (equation 2-4) is usually used for simplicity but it requires the standard deviation of the estimate which is usually unknown. However, if the population standard deviation is unknown, and the sample size is large, Z-tools could be used as opposed to the t-tools with little or no violation to the assumptions (Ramsey, et al., 2002).

$$\text{Z} - \text{Score} = \frac{\text{Estimate} - \text{Parameter}}{\text{Standard Deviation of Estimate}} \tag{2-4}$$

Basically, the need for the population standard deviation is to compute the standard error of the mean. Estimating it with the sample standard deviation would result in a biased estimate unless the sample size is large enough. In real world, it is almost always the case that enough field observations are not available and that assumptions must be made (Wu, et al., 2006). The assumption of independence is assumed in numerous hydrologic and hydraulic modeling situations (Zhang, 2001); (Melching, et al., 2001).

The standard error (equation 2-8) is a good measure of the population dispersion only if the population follows a Normal distribution. For such an assumption, it is expected, as an example, that 95% of the samples would have the sample mean within two standard errors of the mean of the current sample (Ramsey, et al., 2002).

$$\text{Standard Error} = \frac{\sum(x_i - \bar{x})^2}{\sqrt{n}} \tag{2-5}$$

where:

$\bar{x}$ = Sample mean.

x_i = Sample value.

n = Sample size.

2.5.4 Intervals

When we make a point estimate of a parameter, we obtain an indication of a possible value for it. But multiple estimates cannot be expected to return exactly the same value. Because of this variability in estimates, our interest may lie more in stating a region or an interval in which the true value of the parameter most likely lies (McBride, 2005). There are three types of intervals that may be used in such circumstances, depending on the questions addressed, namely, confidence, tolerance and credible intervals.

1. Confidence intervals are ranges in which the parameter may lie most of the time, in repetitive sampling.
2. Tolerance intervals are ranges covering a stated proportion of the population most of the time, in repetitive sampling.
3. Credible intervals are ranges in which the parameter probably lies.

The first two intervals are frequentist, in that they strictly only have meaning under repetitive sampling. The last is Bayesian, in that the probability statement made relates to the particular sample at hand (Taylor, 1998).

One may say of a 95% credible interval that there is a 95% probability that the parameter of interest lies between the interval limits (McBride, 2005). By "probability" we mean a Bayesian personal probability, in which some prior information has been incorporated by means of prior distribution.

A common error is to misinterpret the confidence interval. It is not an accurate statement to say that the probability that a parameter is included in a 95% interval is 95%. An accurate statement would be "if a large number of the 95% intervals is driven, it is expected that the true value of the parameter is included in the interval in 95% of the time". Further, the upper and lower bounds of the interval are random variables because they depend on the sample (Good, et al., 2006).

Another common error in statistics is the use of the notation in equation (2-6) to report the results of a set of observations (Good, et al., 2006).

$$\text{mean} \pm \text{Standard Error} \tag{2-6}$$

The main objective of using this notion is to report on:

- The "correct" result
- The precision of the estimate of the correct result
- The dispersion of the distribution from which the observation are drawn

The standard error is very sensitive to outliers and in this case, it may not be accurate to use equation (2-6) to report on any of the objectives listed. Other tools must be used instead.

Point estimates are seldom satisfactory. For continuous observations, the probability that a point estimate is correct is almost zero. In other cases, an estimate of the precision of the point estimate is required (Good, et al., 2006). A Common error in interval estimation is to use equation (2-7). This equation assumes Normal distribution of a random variable. Even in this case "k" should be determined from tables of the student's t-distribution (Good, et al., 2006) and not from the tables of Normal distribution.

$$\mathrm{CI} = \overline{\mathrm{X}} \pm \mathrm{k} \times \left(\frac{\sigma}{\sqrt{\mathrm{n}}}\right) \tag{2-7}$$

where:

$\overline{X}$ = Sample mean

n = Sample size

σ = Population standard deviation

k = Adjustment factor

Most of the time the population standard deviation (σ) is not known or available (Healey, 1999). In most cases, the population standard deviation can be estimated using the sample standard deviation (s). Equation (2-8), a modified version of equation (2-7), should be used when the population standard deviation is unknown.

$$\mathrm{CI} = \overline{\mathrm{X}} \pm \mathrm{k} \times \left(\frac{\mathrm{s}}{\sqrt{\mathrm{n}-1}}\right) \tag{2-8}$$

where:

$\overline{X}$ = Sample mean

n = Sample size

s = Sample standard deviation

k = Adjustment factor

2.5.5 Sample Size

Sampling distributions are the basis of making inferences about the population from a sample. The sampling distribution of the population mean is the probability distribution of the mean with repeated sampling. It is generally a function of the population distribution and the sample size. According to the Central Limit Theorem, if the population is not Normal, the sampling distribution of the sample mean will still be approximately Normal provided the sample size is sufficiently large (Lindsey, 2004). Unfortunately, the definition of how large the sample should be to satisfy the assumptions of the Central Limit Theorem is subjective. Some statisticians (Good, et al., 2006) consider a sample "large" if it is 25 or more in size. Obviously, larger sample size is better, if available, especially if the population and the sample exhibit a non-symmetric distribution.

Another loose rule of thumb (Healey, 1999) indicates that 50-100 sample size should be sufficient to conclude that Central Limit Theorem assumptions are not severely violated. In general, it can be assumed that the sampling distribution of the mean is Normal, even if the population distribution is not Normal, if the sample size is large enough. However, this also depends on the purpose of the model. A "reconnaissance" simulation is definitely different and requires less stringent guidelines than a well calibrated simulation used for policy analysis.

2.5.6 Bayesian Approach

There are generally two broad approaches for statistical inferences and decision; Frequentist and Bayesian. The former is sometimes referred to as the "Sampling-theory" or "Classical" approach. It is based on assumptions made on the population from the sample. In the Bayesian approach, information other than the sample is formally utilized. The motivation for the Bayesian approach is the desire to base inferences on "any and all" available information (Hays, et al., 1970). Contrary to Frequentists, Bayesians utilize subjective information in the analysis. The concept of "degrees of belief" is considered a formal part of the analysis. Yet, this does not prevent Bayesians from utilizing the "classic" sampling distribution.

In simple terms, Bayesians do utilize the sampling distributions most commonly encountered with frequentists, in addition to other information. For this reason, Bayesian approach may be considered an extension, when applicable, of the classical approach (Hays, et al., 1970).

In some cases, Bayesian and Frequentist approaches produce similar results. However, the interpretation may be a little different. Obviously, the amount of information and the effect of the information on the analysis pose an important factor in determining if the two approaches will be similar. If less additional information is used in the Bayesian approach (or its effect), it will produce similar results to those expected from the frequentist approach.

2.5.7 Statistical Simulation

A simulation is the imitation of the operation of the real world process or system over time. Simulation involves the generation of an artificial history, or future, of the system and through observations of that artificial history draw inferences concerning the operating characteristics of the real system that is presented (Law, et al., 2000).

Types of simulations (Law, et al., 2000) are:

1. Static Simulation models where time plays no role.
2. Dynamic simulation models where system evolves overtime.
3. Deterministic simulation models where no probabilistic components are involved.
4. Stochastic Simulation where at least one random input is involved.
5. Continuous simulation models
6. Discrete simulation models.

A pseudorandom process is a process that appears random but it is not. Pseudorandom sequences typically exhibit statistical randomness while being generated by an entirely deterministic computational process.

Monte Carlo methods are a class of computational algorithms for simulating the behavior of various physical and mathematical systems (Landau, et al., 2005). They are distinguished from other simulation methods by being stochastic, that is non-deterministic in some manner, usually by using random numbers or more often pseudo-random numbers, as opposed to deterministic algorithms. Interestingly, the Monte Carlo method does not require truly random numbers to be useful. Many of the most useful techniques use deterministic, pseudo-random sequences, making it easy to test and re-run simulations. Monte Carlo methods were originally practiced under generic names such as "statistical sampling" (Landau, et al., 2005).

Monte Carlo simulation may be considered as a computational algorithm for simulating the behavior of a number of potential realizations of the physical system of the watershed (Robert, et al., 2004). Monte Carlo simulation randomly generates values for uncertain variables sequentially to simulate a model. It was named for Monte Carlo, Monaco, where the primary attractions are casinos containing games of chance which exhibit random behavior.

2.5.8 Statistical Sampling

Statistical sampling is the selection of individual observations from a population to form a "sub-set". There are a number of sampling procedures, all of which depend on the objective of the project. In most cases, random sampling is required to minimize bias in the resulting inferences.

The statistical method, Latin Hypercube Sampling (LHS) (Wyss, et al., 1998) was developed to generate a distribution of a plausible collection of parameter values from a multidimensional distribution. In the context of statistical sampling, a square grid containing sample positions is a Latin square if, and only if, there is only one sample in each row and each column. A Latin hypercube is the generalization of this concept to an arbitrary number of dimensions, whereby each sample is the only one in each axis-aligned hyper plane containing it.

When sampling a function of N variables, the range of each variable is divided into M equally probable non-overlapping intervals. M sample points are then placed to satisfy the Latin Hypercube requirements. This forces the number of divisions, M, to be equal for each variable. Total number of instances of M for a total number of variables N would result in a total number of simulations equal to M^N. Also this sampling scheme does not require more samples for more dimensions (variables); this independence is one of the main advantages of LHS. Another advantage is that random samples can be taken one at a time, remembering which samples were taken so far. In general, LHS reduces the number of simulations required to describe the full distribution, including the tails (Wyss, et al., 1998).

In random sampling new sample points are generated without taking into account the previously generated sample points. Thus, one does not necessarily need to know beforehand how many sample points are needed. In Latin Hypercube sampling, one must first decide how many sample points to use and for each sample point remember which bracket it was taken from in the PDF.

2.5.9 Representing the Population

For samples of small sizes, summary statistics, like the mean and the standard deviation, are almost meaningless. In some cases, the median or the geometric mean is far more appropriate than the arithmetic mean to represent the population (Good, et al., 2006). Using the median is sometimes better than the arithmetic mean. The arithmetic mean gives a "weight" to every record in the data set, whereas the median does not, at least in the pure sense of the word, "weight" the sample values. Consider a company of 10 people; 9 employees make 35k and their supervisor earns 150k a year. Their arithmetic mean is 46k where as their median is 35K. So the median minimizes the effect that outliers have on the data set (Good, et al., 2006).

The Central Limit Theorem is the foundation for many statistical procedures. The Central Limit Theorem states that if the sum of the variables has a finite variance, then it will be approximately normally distributed. Since many real processes yield distributions with finite variance, this explains the ubiquity of the Normal probability distribution. The amazing and counter-intuitive thing about the Central Limit Theorem is that no matter what the shape of the original distribution, the sampling distribution of the mean approaches normality. It can be assumed that the sampling distribution is Normal with a mean equal to the population mean and a standard deviation equal to $\sigma/\sqrt{n}$ regardless of the shape of the population distribution. Deviations from the normality are generally ok in most cases (Ramsey, et al., 2002).

While testing for the mean, or a "population representative", and since population variance is, in most cases, not known, it can be approximated as indicated above, from the sample.

With all the capabilities of stochastic modeling and the inherit nature of uncertainty, it is possible to address, as needed, the level of uncertainty of most modeling endeavors.

The Central Limit Theorem also states that if the results from several random processes are combined, the resulting distribution will be normal, no matter what distribution types were originally used. This means that I expect the stochastic model results to conform to a normal distribution, even though my parameter distributions might not be normal; i.e. log-normal or even uniform.

2.6 Uncertainty for Water Resources Modeling

One of the greatest challenges facing those engaged in natural resources management is the quantification of variability and uncertainty of hydrology and water quality in natural systems. Stream networks exhibit wide temporal and spatial variability in flow rates due to regional climatic conditions, seasonal weather patterns and changing landscape conditions.

The natural variability in stream flows and chemical loads in streams makes it difficult to assign a single threshold for parameter concentration and load for regulatory purposes (Bonta, et al., 2003). Poor sampling designs, monitoring and laboratory sampling errors and improperly used analytical relations are among the factors that add to uncertainty. Field data used for watershed assessments are usually limited because of the high cost of sampling. Watershed models are often inaccurate because of spatial discretization, poor calibration and inadequate validation. For some water quality measures, such as toxic constituents, there are no models or data to support modeling them adequately (Bonta, et al., 2003).

There are two types of uncertainty; type I and type II (Takyi, 1995).

1. Type I is caused by lack of appropriate knowledge of the dynamics or relationships which may lead to use of inappropriate models.
2. Type II is characterized by the stochastic nature of the variables, incomplete datasets and errors made in measuring, processing of raw data.

2.6.1 Accounting for Uncertainty

Any prediction is uncertain, and in the mathematical modeling of hydrologic and water resources systems this uncertainty has to be accounted for (Hession, et al., 2000); (Bobba, et al., 1996). Of the two primary ways to model uncertainty, one method is based on considering all uncertain elements as random variables and on the use of probabilistic and statistical models. The other approach is based on the theory of fuzzy sets, where uncertainty is modeled by membership functions (Karamouz, et al., 2003).

Many researchers indicate that scientific uncertainties must be estimated and addressed (Hession, et al., 2000). Some researchers (Wu, et al., 2006) classify the complexity of the most common tools of addressing uncertainty ranging from simple to complex to the following:

- Sensitivity Analysis which is basically a “one variable at a time” approach (Hession, et al., 2000). This is usually the most traditional and quick approach. However, it does not include any likelihood of how different the current value is from the best-representing one.
- First Order Error Analysis (FOEA) (Wu, et al., 2006) which is derived from a Taylor’s series of a linear approximation around the mean. FOEA is generally simpler than simulation-based approaches and it can separate the model output uncertainty into its sources. The oversimplification, and occasional inaccuracies of FOEA is backed up by practical and fast application of the approach.
- Monte Carlo Simulation (MCS) is numerically capable of operating complex systems. Unlike FOEA, it is not restricted to linear and continuous systems. Modelers should pay close attention to the selection of the parameter PDF used and the number of simulations. Accuracy expected from MCS is largely affected by how well the PDF represents the input parameters (Landau, et al., 2005).

Modeling watersheds in a simulation-based framework can help considerably to account for uncertainties and hence the predictive capability of the model will be greatly enhanced (Salah, et al., 2005-b). In this sense, any model parameter that is considered as a random variable has to have a PDF or PMF instead of a single value. Other parameters that do not exhibit a large amount of uncertainty, based on experience or previous model sensitivity analysis or even lack of information, are entered into the model as a single value.

2.6.2 Examining Uncertainty

Uncertainty refers to random prediction error resulting from limitations in the data and models. The level of uncertainty determines the probability of achieving the desired standard at a specified frequency under some given conditions. Uncertainty can be reduced in many cases by collecting additional data and improving forecast models under adaptive management framework (Walker, 2003).

There are three types of uncertainties in most natural phenomenon. Any effort in reducing the uncertainty at any of these types will result in enhanced model output.

1. The natural uncertainty inherent in the phenomenon itself;
2. Model parameter uncertainty and
3. Model uncertainty. (Marco, et al., 1993).

2.6.3 GIS and Uncertainty

When GIS-based hydrologic and water quality models are used to evaluate the response of a watershed, every effort must be taken to minimize model uncertainties associated with input data. Results of some research (Cotter, et al., 2003) indicate that GIS data resolution has a significant impact on model output uncertainty. Models were found to be more sensitive to DEMs than land use and soil data in predictions. However, for a mid-size watershed, model predictions were most likely acceptable because of low relative error, up to pixel size of 150m (Cotter, et al., 2003).

In general, water quality management is complex because rivers and lakes are polluted from multiple sources. Therefore, it is necessary to use models that incorporate uncertainty and complex characteristics of the pollution problem. Stochastic modeling, in which stream flow and also various water quality parameters are assumed to be random variables, has been formulated by many researchers (Karamouz, et al., 2003).

2.6.4 Guidelines and Uncertainty

Federal guidelines (US-EPA, 1999-a) require consideration of variability and uncertainty in the development of TMDLs to meet water quality standards in impaired water bodies. Consideration of these factors is necessary to ensure that a TMDL implementation will meet objectives with a reasonably high probability of success and in a cost-effective manner. These requirements can be met using a variety of implicit or explicit approaches. Implicit approaches embed an MOS into one or more conservative assumptions in supporting analysis. If the MOS is not quantified, there is some risk that the resulting load control programs would be over-designed (resulting in unnecessary regulations and expenses) or under-designed (having a low probability of meeting objectives). If the MOS is explicitly quantified, control measures will generally be over-designed sufficiently to achieve the particular goal with a particular level (Walker, 2003).

2.6.5 Parametric Uncertainty

A number of concerns exhibit marked uncertainty for some aspects in integrated water quantity and quality management (Olufemi, et al., 2003).

First, in setting water quality standards, translating the narrative of designated uses, such as the requirement for fishable waters, into numeric target values like minimum dissolved oxygen concentration to support fish population is rather imprecise (US-EPA, 1999-a).

Second, uncertainty in a watershed's water body assessment results from the inadequate frequency and spatial distribution of observations of key water quality criteria, a situation that is derived from limited resources available to state and federal agencies responsible for ambient water quality monitoring.

Third, linkage analysis involves the use of simulation models. Such process-based models are characterized by structural and parametric uncertainties, the former resulting from an inadequate conceptual understanding of the internal behavior of the water body, and the latter from the inability to uniquely estimate the process parameters that quantify the component mechanisms in the model.

Fourth, the TMDL allocation formula includes the margin of safety that accounts for uncertainties in relating pollutant loads to receiving water quality.

Considering the current state of research, there is a question posed to water resources managers, and that is; will our decisions generally be better if we have some idea of the range of possible outcomes that might result? Many researchers believe the answer is yes (Reckhow, 2003). U.S. EPA also believes the answer is "yes", although their reasoning is unclear (US-EPA, 2005). EPA's perspective is implicit in their technical requirement for an uncertainty-based MOS in a TMDL application for instance; however, absent from EPA guidance is an explanation as to why decisions improve with an uncertainty analysis.

Decision makers are better off knowing the forecast uncertainty and for them to realize that they first need (1) motivation, that is, they must become aware of substantial magnitude of forecast error in many water resources assessment and (2) guidance, ideally, they need relatively simple heuristics that will allow them to use this knowledge to forecast error to improve decision making in the long run.

3 Methodology

This research is carried out in three interrelated parts; a conceptual framework, software development, and application demonstrations. The conceptual framework provides the theoretical basis and the software development describes the development of the tools used for its implementation. These tools are used to examine case studies that demonstrate the research concepts and provide a guide for implementing them in a watershed approach to water management.

3.1 Conceptual Framework

There are two interrelated driving forces for this research:

1. Develop an integrated water resources and quality modeling tool and,
2. Develop and implement a method for addressing uncertainty in integrated water resources modeling.

Any comprehensive integrated water resources framework should include routines that model both the land and water portions of the watershed, from both quantity and quality perspectives. As indicated in Chapter 2, currently, no model exists that combine modeling the land and water portions of the watershed with similar strength. This research makes use of a strong land model, GSSHA (Downer, et al., 2006), which uses spatial distribution to model localized land-based activities accurately and in a detailed manner, and a water body model, CE-QUAL-W2 (Cole, et al., 2007), which models the hydrodynamics and water quality of rivers, lakes and estuaries using a longitudinal grid to capture the changes over the water body profile.

A stochastic approach is used as a means of addressing uncertainties inherent in the hydrologic and water quality arena. In short, a deterministic approach still serves an important role in today's hydrologic analysis. However, using a stochastic approach allows decision makers to include uncertainty in their analysis.

All models operate on particular spatial and temporal domains. Integrating one model to another must take into consideration the spatial and temporal nature of the models involved. Linking GSSHA to CE-QUAL-W2 involves some spatial aspects that required innovations to implement. Integrating the two models requires a link between the two different domains (Figure 3-1) that successfully routes the water and chemical fluxes in a mass-preserving approach. The two models are linked in a framework (Figure 3-2) that models the land-water interface linked processes stochastically in order to address the uncertainty inherent in these types of models.

Conceptually, this approach will allow water resources managers to be able to model point and non-point source pollution throughout the basin, using a calibrated GSSHA model, route the overland flow to the adjacent, if any, water body and use these results as boundary conditions for a calibrated CE-QUAL-W2 model. The overland flow is routed to the adjacent water body through GSSHA grid cells or through a stream node. The stream node is the intersection of the GSSHA stream and the segment map.

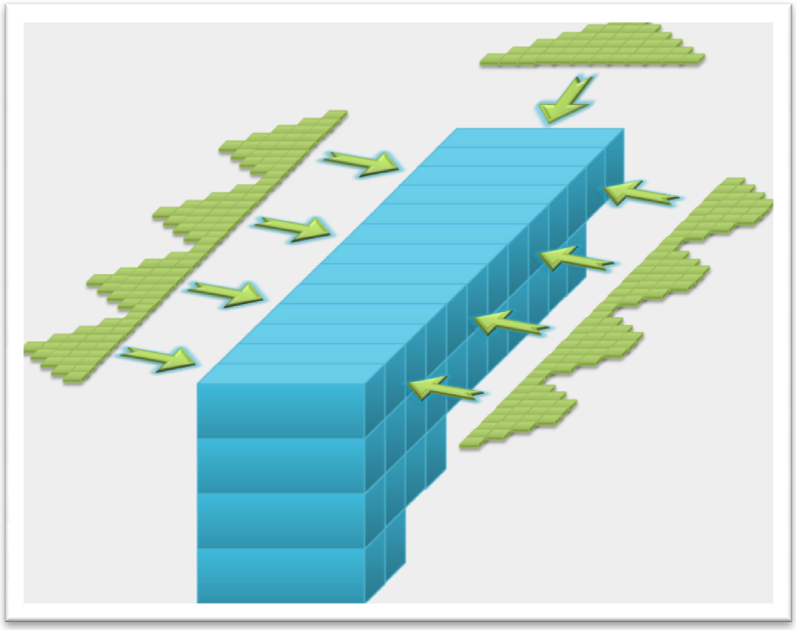


Figure 3-1: Linking Two Models with Different Domains.

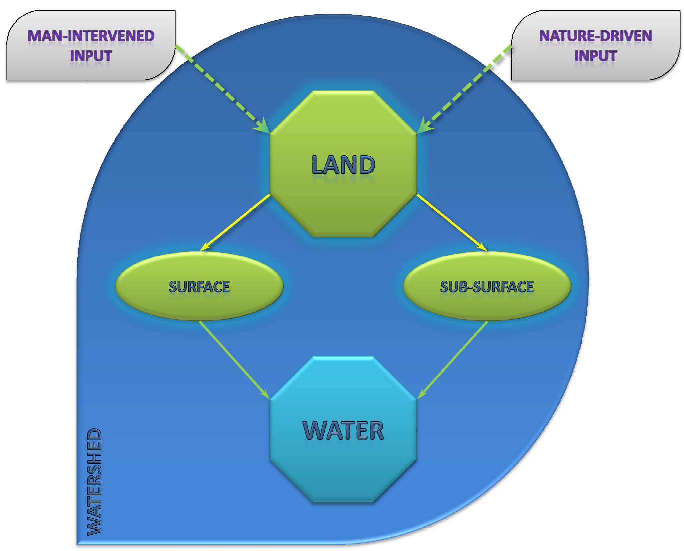


Figure 3-2: Integrated Water Resources Modeling; Conceptual Framework.

The CE-QUAL-W2 model will then be used to generate potential future scenarios, in a systematic and defensible manner that is required for any watershed-based management scheme. For example, a BMP, a TMDL study or a UAA can be simulated with an integrated modeling framework similar to the one proposed in this research.

Figure 3-1 shows the difference between the model domains. The horizontal 2-D grid represents GSSHA's grid where the land based activities are simulated. The vertical 2-D grid represents CE-QUAL-W2's laterally averaged longitudinal grid. As indicated, the output from the horizontal grid is used, in this research, as input to the vertical grid.

Figure 3-2 demonstrates the overall integrated process investigated in this research. Both nature-driven and man-made inputs to the land are modeled and the resultant output is routed through to the receiving water bodies prior to leaving the watershed through withdrawals.

3.1.1 Land Modeling

In the framework used in this research, GSSHA is used to model the overland flow and associated processes which occur in the watershed. As noted, GSSHA's strength is the ability to model hydrologic processes using a two dimensional spatial grid. Instead of using deterministic values for GSSHA input parameters, the framework developed in this research allows modelers to stochastically choose values from a pre-defined, or user-defined, parameter probability density functions (PDF) to help understand and quantify the uncertainty in various model parameters.

In some cases, values for some parameters are unknown or have a high degree of uncertainty due to measurement, instrumental or any other type of errors (Downer, et al., 2006). In this sense, the stochastic representation of model input parameters is very useful to manage, identify and include these uncertainties in subsequent management phases. Modelers will be able to simulate a potential range of system values using a large number of runs based on different values automatically picked from the pre-defined PDF (Salah, et al., 2005-b).

This research investigates if stochastic land modeling overcomes model and parameter-related uncertainties. The research focused on parameter-related uncertainties which are simply a result of, for example, measurement, instrumental and recording errors or true parameter variability. GSSHA-related uncertainties address outputs in the model that are a direct result of the uncertainties in the used processes in GSSHA and if they are best suited to the conditions of the watershed modeled. This research did not examine model-related uncertainties.

As indicated in Chapter 2, GIS data resolution can affect model output. In a similar fashion, and as demonstrated in this and the following chapters, the grid cell size of GSSHA affects model output and hence the overall results. However, since the main advantages of statistical simulation is that it overcomes many uncertainties and ambiguities in the choice of modeling factors such as the grid cell size. This research indicated that the effect of varying cell sizes have on output in stochastic modeling is not as apparent as for deterministic modeling.

Since GSSHA does not currently - 2009 - support irregular grids, selecting GSSHA grid cell size depends on many factors. Some considered in this research are:

1. Accuracy: cells need to be as small as possible to depict local changes in input parameters and produce the accuracy desired.
2. Data Availability: cell sizes are often determined by available data, modelers should not use smaller cell sizes than the available data can provide for, otherwise there will be a great deal of interpolation and estimation.

3. Sensitivity: depending on the case, watershed and the main parameter being modeled, some hydrologic models tend to be affected more by cell sizes as opposed to some other models that exhibit little or no sensitivity to cell size.
4. Possibility of irregularly sized grid. In most cases, relatively large grid cells are appropriate where either no great changes in model input parameters are expected or little accuracy is needed at a particular location with small cell sizes needed in areas where higher accuracy is desired and complimented by sufficient data.
5. Computing power: ideally, cell sizes are coordinated to not exhaust computing resources by creating overwhelming file sizes.

As part of this research, stochastics were used in GSSHA, and could be used in CE-QUAL-W2, interchangeably to change the parameters that the modeler thinks will possess a high degree of uncertainty and affect the model results. At the end of this phase of modeling, multiple GSSHA runs, as opposed to a single deterministic run, are simulated based on a number of stochastic inputs. These runs need to be processed before they can be used in the following water modeling phase of the linkage.

3.1.2 Water Modeling

In this research, CE-QUAL-W2 is used to model the hydrodynamics and the water quality aspects associated with a reservoir serving as the receiving body within a watershed. CE-QUAL-W2 creates a rectangular grid that approximates the orientation, length and width of the actual physical boundary of the water bodies (Cole, et al., 2007). These approximations are determined by the modeler and, of course, accurate approximations are required for reliable models. In this regard, it is important to represent this interface correctly.

In CE-QUAL-W2, a water body is divided into one or more branches. Branches may be established according to natural sub-boundaries in the water body or according to special modeling requirements. Each branch is divided into one or more segments. Similar to branches, segments may be established according to natural geometry of the water body or to account for other modeling requirements. Each segment may, then, be vertically divided into one or more layers (Figure 2-3). Based on that, the water body is modeled with a vertical two dimensional grid where each segment is laterally averaged and hence any bank-to-bank changes are compiled into one cell. Like any model, the selection of cell, branch, segment and layer sizes, depend on the accuracy desired.

3.1.3 Spatial Linkage

This research defines the spatial linkage or the land water interface as the geometric boundary between the land and water portions of the watershed. It is the shoreline of the water bodies modeled by CE-QUAL-W2.

The main concept of this definition is to get all inputs from land to the water (Figure 3-3). As seen in Figure 3-3, conceptually, the linkage between GSSHA and CE-QUAL-W2 is not difficult. However, when the two model domains are discretized in dissimilar fashions, the interface between the domains becomes much more complex.

A major part of the challenge of linking these two specific models is the geometric dissimilarity of the model domains. GSSHA models a land segment using a horizontal two dimensional grid whereas CE-QUAL-W2 is a two dimensional laterally average model (vertical two dimensional). The aggregated input from GSSHA is generated at the top layer of the CE-QUAL-W2 model as input.

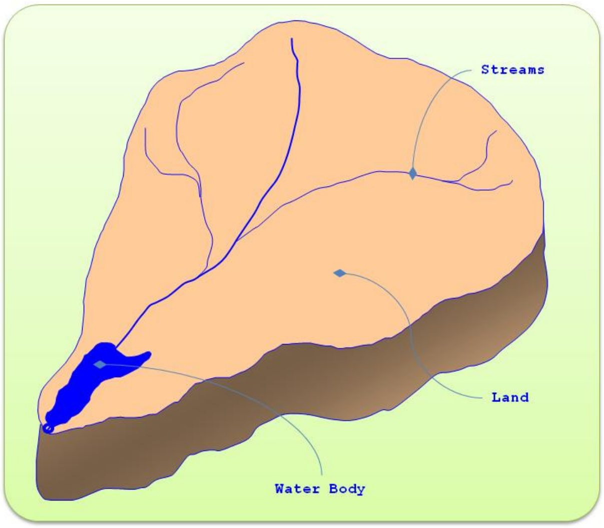


Figure 3-3: Modeling Domains.

A main assumption of this research is that the overland flow, from GSSHA, is combined into the top layer for each CE-QUAL-W2 segment. Thus, all the water fluxes from GSSHA are aggregated into the top layer of each CE-QUAL-W2 segment. GSSHA does compute groundwater flows, but there is nothing in the code that would allow "extracting" the information easily enough to include in a stochastic modeling sequence (Downer, et al., 2006).

On the other hand, CE-QUAL-W2 allows tributary inflows to be distributed among model layers. Options for distributing the inflow among layers are:

- Even distribution among model layers [DISTR].
- Density-specific distribution [DENSITY].
- User specified range of layers [SPECIFY].

In this research, the third option is assumed. It is assumed so, as currently, GSSHA underground flows are not contributing to lower, than the top layer, layers in CE-QUAL-W2. Thus, for areas where groundwater exfiltration to a water body is significant this approach would not result in a reliable model unless the users manually implement either one of the first two options.

On the spatial scale, three options were identified in this research to model the linkage between the land and water in this approach:

1. A GSSHA Lake: A lake in GSSHA is defined by a set of cells representing the initial, minimum and maximum water surface (Byrd, et al., 2005). Lakes polygons in the WMS interface for GSSHA are used as reference features and are not actually used by GSSHA. GSSHA lakes are not allowed to shrink below the minimum water surface elevation or extend beyond the maximum. This research indicated that if a GSSHA Lake is used in the spatial linkage, the stream network must extend to the minimum water surface elevation cells, not the maximum or initial (Figure 3-4). Figure 3-4, A indicates stream network that is not extended to the minimum water surface elevation cells; i.e. the inner most and dark cells. Whereas, Figure 3-4, B shows the stream network goes beyond the outer most, light grey, lake grid cells, to be connected to the minimum water surface elevations. The linkage proposed in this research results in direct impact to the water surface elevations in the CE-QUAL-Model as a result from the input from GSSHA, the linkage (if a GSSHA Lake method was used) keeps track of the lake water surface

elevations and update them as appropriate. The spatial linkage in this research is not assumed to be bi-directional and thus this option is not considered.

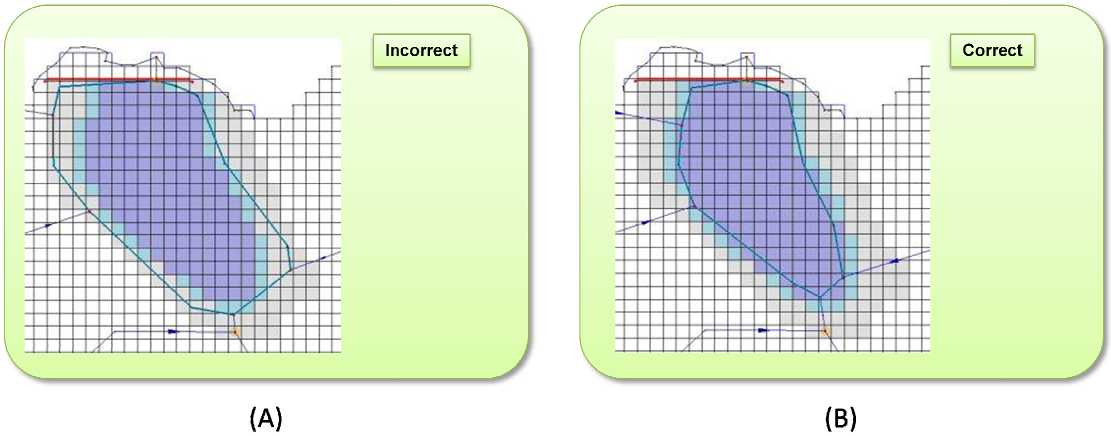


Figure 3-4: GSSHA Lake/Stream Network (Source (Byrd, et al., 2005)).

2. The Land-Water Interface: as defined earlier, the land water interface is the geometrical border that separates water from land in a watershed. In this research the following are taken into consideration in accurately defining the geometrical aspects of the Land-Water interface.

➢ Boundary cells: This research effort has identified water body, and specifically segment polygon, boundary cells as the cells overlapped by arcs comprising the segment polygon provided that it satisfies the cell centroid condition (Figure 3-5), (Figure 3-6). The dark outlined cells in Figure 3-5 show seven GSSHA cells that would be selected by the algorithm to be boundary cells for the most upstream segment of the water body; i.e. Segment 2. It must be noted that, for the purposes of this linkage, the inner cells that lie between segments and do not represent any water body shore line are not considered as boundary cells.

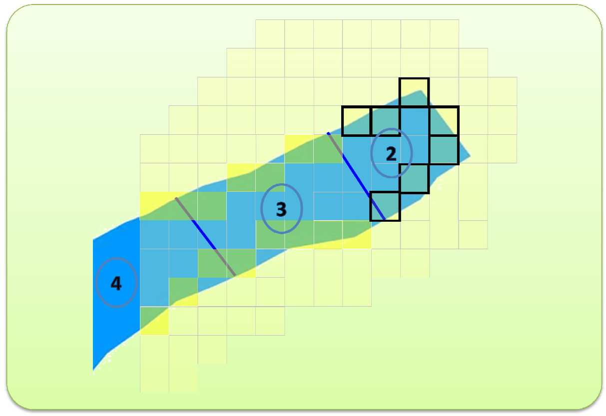


Figure 3-5: Boundary Cells.

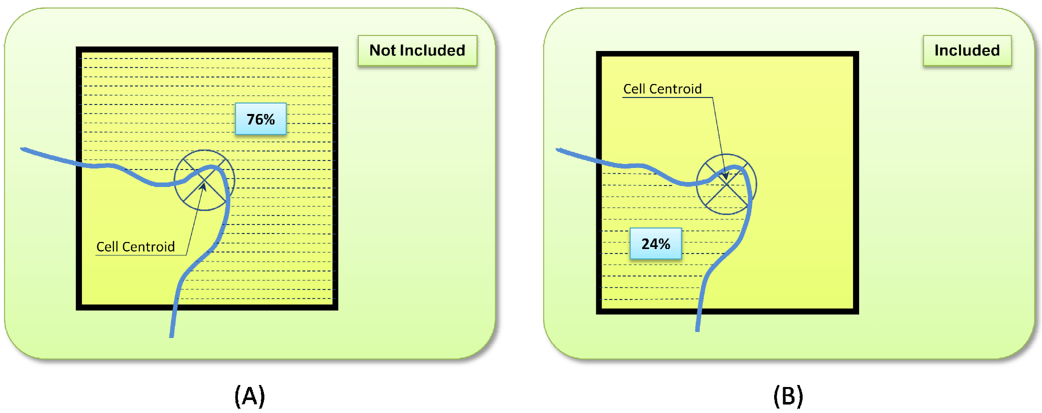


Figure 3-6: Cell Centroid.

- Cell centroid: In this research, a cell is considered a boundary cell if the cell centroid is inside the segment polygon. The overlapping percentage between the cell and the segment polygon is not considered (Figure 3-6). For example, Figure 3-6, A shows a cell that is 76% covered by the segment but the cell centroid is outside the segment area, thus it is not included as a boundary cell. On the very contrary, Figure 3-6, B shows a cell that only overlaps a segment polygon by about a quarter of its area but the cell is included as a boundary cell because the cell centroid lies within the segment polygon.

➢ Double Cell Selection: I paid close attention towards selecting a cell more than once. If the boundary cell selection is solely based on arcs compounding the segment polygon, a cell may be selected twice (Figure 3-7, A). For a spaghetti topology, a start node of a polygon arc is the end node of another. To avoid duplicate cell selection, my algorithm keeps track of selected cells and checks to see if a new cell that need to be added is already selected or not.

➢ Meandering Shoreline: when the general direction of the shoreline is orthogonal; i.e. parallel to either side of the GSSHA grid, (Figure 3-7, B) and depending on the relative meandering in the shoreline to the grid cell size, boundary cell selection should avoid double cell selection or any L-Adjacent cells. In some cases meanders in the shoreline would lead to selection of two cells from the same row or column.

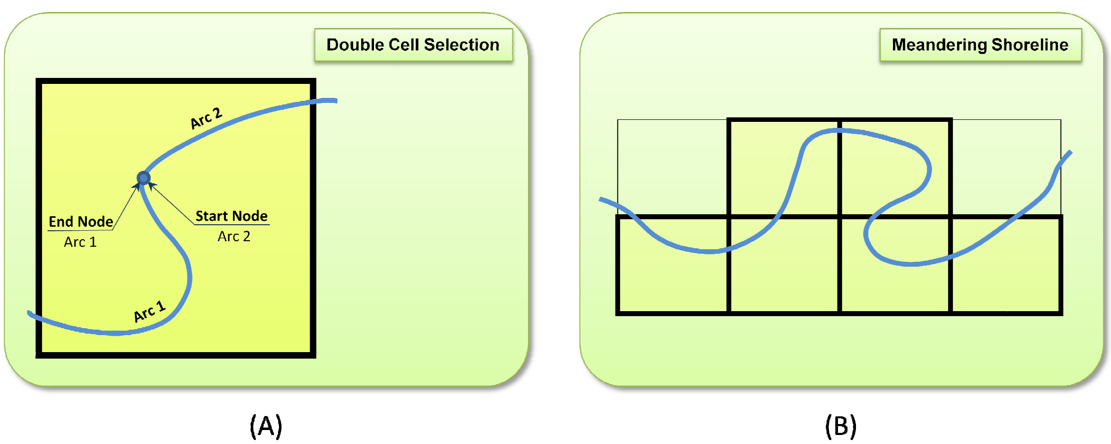
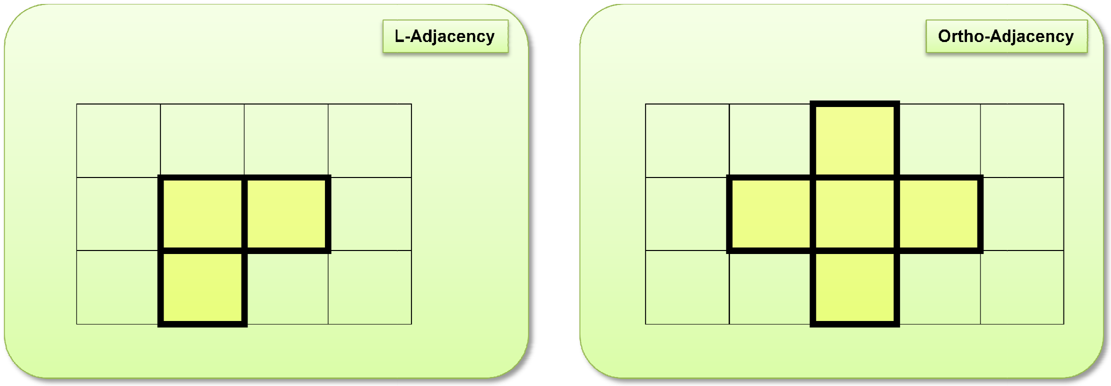


Figure 3-7: Double Cell Selection & Meandering Shorelines.

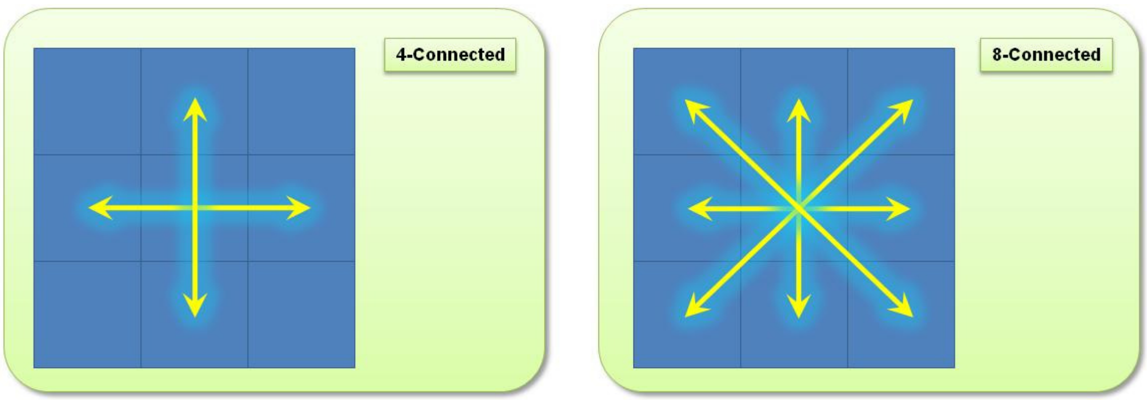
➢ L-Adjacency: with the various boundary cell selection take place, L-adjacent cells may result (Figure 3-8, A). L-Adjacent cells may result in double or inaccurate inflows to the designated segment. Thus L-Adjacent cells must be avoided. In this case, the first cell that has its centroid outside the segment polygon is eliminated.

- Ortho-Adjacency: Similar to L-Adjacency, boundary cell selection may result in ortho-adjacent cells (Figure 3-8, B). Ortho-Adjacent cells may result in double or inaccurate inflows to the designated segments. Thus ortho-Adjacent cells were avoided. This research uses proposes spatial filtering to avoid Ortho-Adjacency.
- Spatial Filtering: as used in image analysis, filtering is a selective process meant to enhance features in a grid. In this algorithm, a 3x3 filter is used to scan through the bounding box of the water body, as opposed to the entire GSSHA grid, from the top left corner to the lower right. The filter evaluates every 9 cells, a group at a time, and eliminates both the Ortho and L-Adjacency issues outlined above. The filter is designed to get rid of one of the non-collinear cells in an L-Adjacent cells situation (Figure 3-8, A) and the two non-collinear cells for ortho-adjacent cells (Figure 3-8, B).
- 4 or 8-connected: The direction of flow from a GSSHA cell to a CE-QUAL-W2 segment, orthogonal (Figure 3-9, A) or ortho-diagonal (Figure 3-9, B) is not crucial in this approach since it does not affect inputs to CE-QUAL-W2.
- One-Cell Boundary: In some cases where the segment polygon is too narrow in relation to the GSSHA model grid cell size, the algorithm developed in this research is set to select one cell wide for the segment even though the cell would not cover the centroid or have a significant overlap with the segment polygon (Figure 3-10).



(A) (B)

Figure 3-8: L & Ortho-Adjacency.



(A) (B)

Figure 3-9: 4 and 8–Connected Cells.

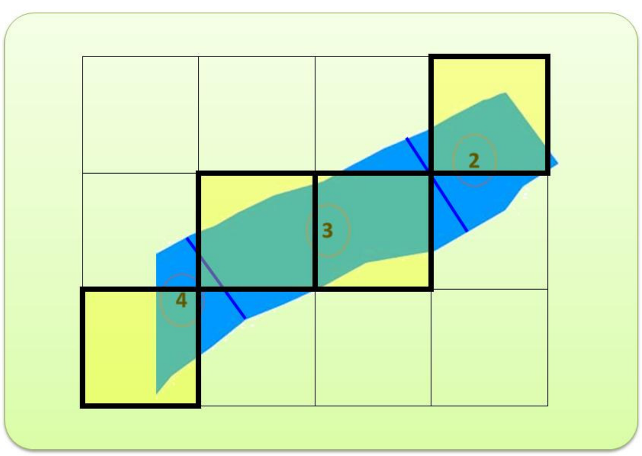


Figure 3-10: One-Cell Boundary.

- Duplicated Segment Designation: in some cases, a cell may overlap more than one segment. For obvious reasons, the algorithm does not allow multiple segment designation for the same cell. In such a situation, the algorithm designates a segment to the cell based on the shortest distance from the cell centroid to the segment polygon centroid.

3. A Segment ID Index Map: the water body in the CE-QUAL-W2 model is divided into branches and each branch is divided into segments. An index map is generated for the whole catchment with the same grid cell size as the GSSHA model. An integer value is assigned to each cell in this index map as follows:

- Zero: for cells that are not overlapped by the water body.
- Non-zero: for the cells that are overlapped by the water body. Each value represents the segment number that is mapped to the cell.

Out of the above available three options, the first one is ruled out because of potential programming issues in both WMS and GSSHA. The "index map" (option3) was favored for consistency with the current GSSHA interface in WMS. Based on personal communication with the GSSHA development team, it was recommended to proceed with a combination of the last two options as appropriate.

In summary, this is how the spatial linkage algorithm was implemented:

1. Verify that both branch and segment coverages exist for the water body.
2. Verify that a GSSHA Grid exists and covers at least the bounding box of all segments in the CE-QUAL-W2 model.

3. Get the segment ID range; i.e. minimum and maximum segment ID. Noting that the minimum segment ID should always be 2 for a successful CE-QUAL-W2 model.
4. Scan through all the GSSHA grid cells and determine the cell centroid.
5. Get the first segment polygon that encloses the cell centroid.
6. Assign the segment ID to the grid cell.
7. Create the index map file for further processing.

A major source of temporal input is the stream flow. In this approach, stream input to the water body is obtained from the GSSHA output hydrograph at the stream nodes. Stream nodes are nodes where streams intersect CE-QUAL-W2 segments. The hydrograph is added as a boundary condition to the particular segment (Figure 3-3).

3.1.4 Temporal Linkage

For the temporal linkage, the duration, time step, start time, and end time of the two models should be related. For this research, there are basically two ways of temporally linking the GSSHA and CE-QUAL-W2 models:

1. Identical time-stepping: using the same time step, start and end date for the two models, and
2. Interpolated time step: using a different time step for the two models and interpolating the results for one to the other.

Practically, the CE-QUAL-W2 time step should be set equal to the hydrograph output frequency in GSSHA (Hyd_Freq), and identical time-stepping should be used for ease of modeling (Downer, et al., 2006) ; (Cole, et al., 2007). This would allow aggregated fluxes from GSSHA to be considered as direct input in the CE-QUAL-W2 segment inflow file. Interpolated time step data requires an added effort that can be avoided. Another disadvantage is the speculation that interpolated inputs may not represent the actual input at the interpolated time step.

GSSHA model run duration does not need to be the same as CE-QUAL-W2 run duration. Generally there are three cases for the run durations of the models:

1. GSSHA run duration is longer than CE-QUAL-W2 run duration. In this case, the linkage is designed to "trim" the flux files generated by GSSHA to match the total run duration of CE-QUAL-W2, specified in the control file.
2. GSSHA run duration is shorter than CE-QUAL-W2 run duration. In this case, the linkage is designed to "extend" the flux files generated by GSSHA using a dummy record that has a flow of zero and Julian date matching the end of the CE-QUAL-W2 model specified in the control file.
3. GSSHA run duration is the same as the CE-QUAL-W2 run duration. In this case, the linkage is designed to make sure and slightly adjust, if necessary, the last time step in the GSSHA-generated flux files.

3.1.5 Stochastic Approach

As noted in Chapter 2, some parameters are best handled stochastically, which means that their values are chosen from a pre-defined PDF, as opposed to using a single value. The framework developed in this research is designed to allow modelers to define a PDF of their choice or use pre-defined standard PDFs, for a given set of parameters.

I used static simulation model; i.e. Monte Carlo Simulation, to select parameter values. The selection process is assumed to be either random or follow an LHS scheme.

The list of currently available distributions in the interface developed for this research includes:

- Beta Distribution (with default shape factors of $\alpha = 2$ and $\beta = 7$ by default)
- Normal Distribution,
- Log-Normal Distribution, and
- Uniform Distribution.

In accordance with other research efforts (Ricciardi, et al., 2005); (Heyman, et al., 1984); (Valdes, et al., 1990), the Beta distribution was selected to be the default distribution in this interface. The main reasons the Beta distribution (Figure 3-11) was incorporated in the interface and set as the default distribution are as follows:

- Positive: most hydrologic modeling parameters are positive.
- Bound between 2 values (minimum and maximum): it is evident that most hydrologic parameters have extreme values that they do not go beyond.
- Not truncated: as opposed to Normal distribution, the minimum and maximum values are not represented by a truncated probability distribution function. The minimum and maximum are actually features of the distribution.
- Can be used to represent outliers and extreme values: Most hydrologic parameters are generally distributed around the median with very few occurrences of "extreme events".
- Can approximate normal, log-normal and other types of distributions.

As an example, the PDF with $\alpha = \beta = 7$ (Figure 3-11) may be a good representation of some variables where the majority of the parameter values lie in the middle zone with minimal probability of occurrence on the two edges. Also, the PDF with $\alpha = 2$ and $\beta = 7$ (Figure 3-11) may be a good representation of some variables where the majority of the parameter values lie in the lower zone with minimal probability of occurrence on the higher ranges.

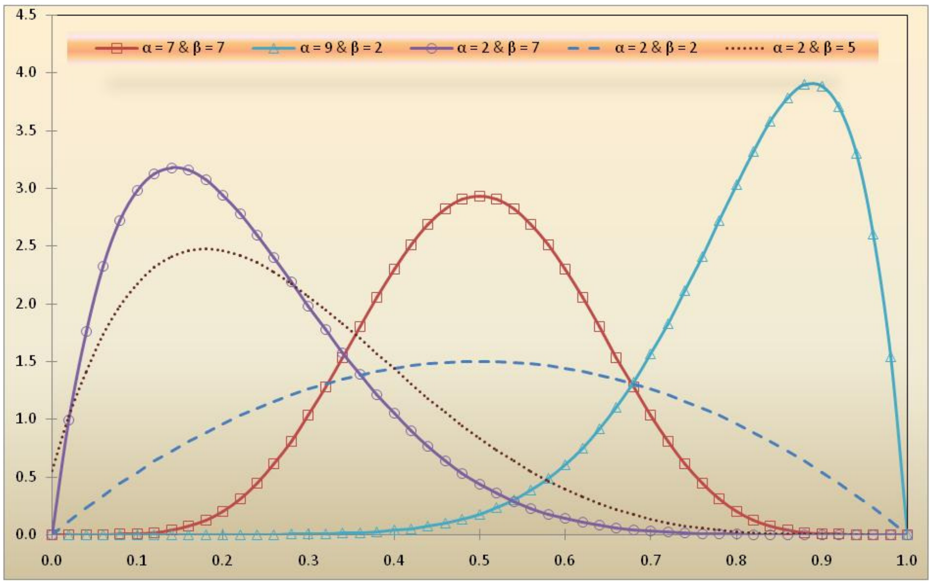


Figure 3-11: Sample Beta Distribution Set for Earth Science Observations.

3.1.6 Credible Intervals

As indicated in Chapter 2, equation (2-8) was used to estimate the credible interval for the time series plots. The interface I developed in WMS is designed to indicate the following set of probabilities. These probabilities are utilized to satisfy equation (3-1) and Table 3–1:

- 50%
- 80%
- 90%
- 95%

- 98%
- 99%
- User defined

$$\text{Credible Interval Probability} = 1 - \alpha \quad (3\text{-}1)$$

where α = Confidence level.

Table 3–1: Z Scores for Various Confidence Levels.

Confidence Level	α	$\alpha/2$	Z-Score
50.0%	0.500	0.2500	±0.675
80.0%	0.200	0.1000	±1.282
90.0%	0.100	0.0500	±1.645
95.0%	0.050	0.0250	±1.960
98.0%	0.020	0.0100	±2.326
99.0%	0.010	0.0050	±2.576
99.9%	0.001	0.0005	±3.291

Building an initial model helps in refining the results. If the interval width is wider than anticipated for the desired credibility level (i.e., 95% credible interval), a desired interval width can be determined using the initial model results.

Assuming the standard deviation of the sampling distribution remains the same, the number of simulations required to get a specific interval width for the corresponding credibility level can be estimated using equation (3-2).

Referring to equation (2-8) and since in most cases, the population standard deviation is not known, equation (3-2) was used in estimating the credible interval width.

$$\text{Width}_{CI} = k \times \left(\frac{s}{\sqrt{n-1}}\right) \tag{3-2}$$

where:

$Width_{CI}$ = Credible interval width

n = Sample size

s = Sample standard deviation

k = Adjustment factor (e.g. Z-Score)

In this research, the number of simulations is viewed as the "sample size". Following equation (3-2), and using the values for the sample standard deviation and desired credible interval width and the credibility level, modelers will be able to determine the required number or model runs (sample size) using the developed interface.

Each point in a time series output, or an output grid, of CE-QUAL-W2, as an example, is basically a representation of the population which can be viewed as all the possible values that could have happened for this particular location and time step. In this research, this point is considered as a sample mean which may vary from one sample to another due to the variability of the input parameters used to obtain it.

As indicated in Chapter 2, it is assumed that sample means (point values) tend to be normally distributed even though the population, which considers all potential values for the parameter under investigation at the respective point in time and space, may not follow a Normal distribution.

Currently, WMS users can generate a deterministic time series from CE-QUAL-W2 output representing parameter values at a specific segment and layer over time. These values are generally considered what would be expected at the specific location and time. In the stochastic approach presented in this research effort, these values are only considered representatives of the "population" of that parameter at the respective location and time. The "population" of the parameter is all the possible values that could occur for the parameter at the specific location and time. The population of an output parameter in CE-QUAL-W2, like most other models, is a direct result of model and parameter uncertainties.

To illustrate the methodology followed in this research to represent stochastic time series, Figure 3-12 shows a portion (only four time steps) of a deterministic time series plot of a pollutant. As indicated in the figure, the last time step was investigated for the potential population. The solid line, with marking symbols, in Figure 3-12 shows a portion of time series of a deterministic output. This portion of the time step shows 4 time steps, 3 of which represented by small squares where the most right one, "the investigated time step", is represented by a star. The bell shaped curve indicates a hypothetical distribution of the same parameter at the last time step (the furthest to the right). As shown, the relative position of the deterministic line and the hypothetical value distribution divides Figure 3-12 into four possible cases:

A. The deterministic value seems to correspond to the median of the distribution of the parameter at the investigated time step. This is evident from the fact that the value of the time series at the investigated time step seems to center vertically with the population PDF. This would mean that this value seems to be a good representation of the population. In this case, the deterministic modeling may be a good representative of the actual value at the investigated time step. However, this may not be the same across all other time steps (Figure 3-12, A).

B. The deterministic value seems to be at the lower (left) tail of the distribution of the parameter at the investigated time step. In this case, the deterministic modeling may have under-estimated the actual value at the investigated time step. However, and as with the previous case, this may not be the same across all other time steps (Figure 3-12, B).

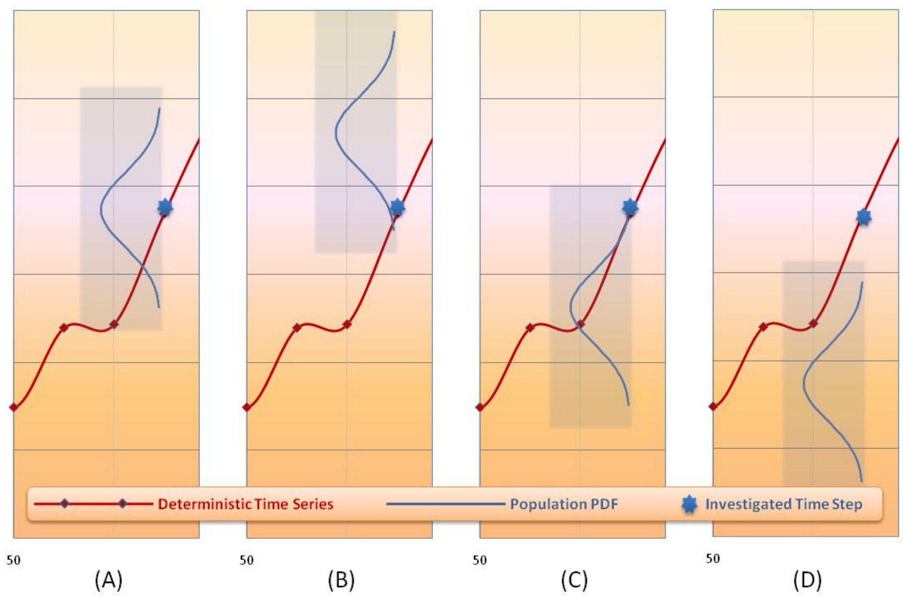


Figure 3-12: Distribution of a Parameter and a Deterministic Time Series.

C. The deterministic value seems to be at the upper (right) tail of the distribution of the parameter at the investigated time step. In this case, the deterministic modeling over-estimated the actual value. However, and as with the previous cases, this may not be the same across all other time steps (Figure 3-12, C).

D. The deterministic value seems to be as an outlier in relation to the distribution of the parameter at the investigated time step. In this case, the deterministic modeling may have considerably over-estimated, or under-estimated if the outlier is on the other side, the actual value at the investigated time step. However, and as with the previous cases, this may not be the same across all other time steps (Figure 3-12, D).

It is evident from Figure 3-12 that one cannot generalize a trend for all time steps. A deterministic model may result in over- or under-estimation of the parameter under investigation or anything in between. Further, it can change across time steps. A deterministic model may result in an over-estimation in some time steps and under-estimation in other time steps for the same model. A better way to represent this variation is with a credible interval that contains the mean as well as the upper and lower bounds of the interval across the time series plot (Figure 3-13).

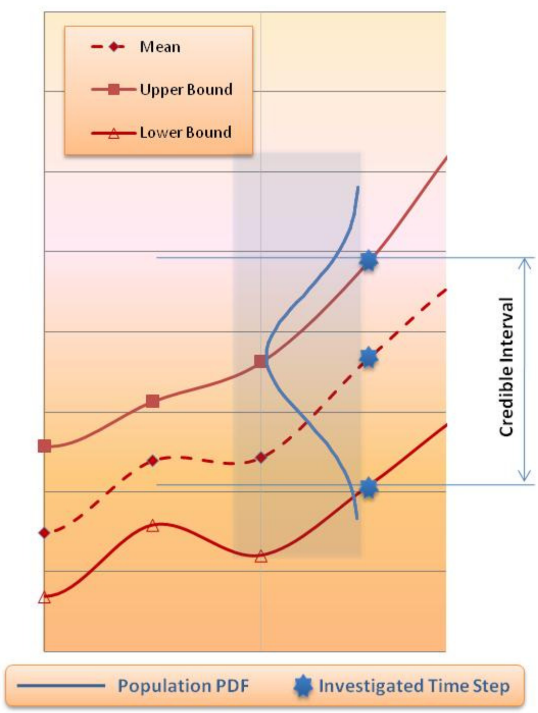


Figure 3-13: Credible Interval Time Series Output.

Figure 3-13 shows the same portion of the time series with the four time steps shown and only one of them being investigated. As seen, the credible interval is constructed using the PDF at the investigated time step (the furthest to the right) to indicate a lower and an upper bound. The PDF is theoretically underlying all the time steps, but it is shown only by this particular time step for illustration. The credible interval is defined by the upper and lower bounds. Depending on the level of credibility desired the difference between the lower and upper bounds can vary.

The parameter importance (Williams, et al., 1995) is a relative mean to compare how wide the variations in each model input parameter are. These variations will eventually affect output parameters. An "important" parameter usually means the combination of two things:

1. The model is very sensitive to variations in parameter values.
2. There is more uncertainty in the input parameter.

Either of the two or both would affect the population of the output parameter. A more "important" input parameter is likely to contribute to the uncertainty in output because of the inherited variations due to model sensitivities or uncertainties embedded in the parameter.

One of the goals of this research is to estimate an unbiased representative of the population that addresses uncertainty. Hence, random sampling of important model input parameters is believed to be a good ground for an unbiased experiment (in this case, it is the linked models) that would result in a sample (output parameter values).

The linkage proposed in this research is viewed as merging GSSHA and CE-QUAL-W2 into one "experiment" (Figure 3-14). In statistical terms, an experiment usually describes the process of drawing a sample (Miller, 2006). In a deterministic approach one value for each input parameter (out of the whole parameter population) is used as an input to the experiment. The "experiment" results in a "sample" which is a direct representation of the output parameter. Figure 3-14 is an illustration to the concept followed in developing the methodology for this research. It shows three major cases of sampling from a population:

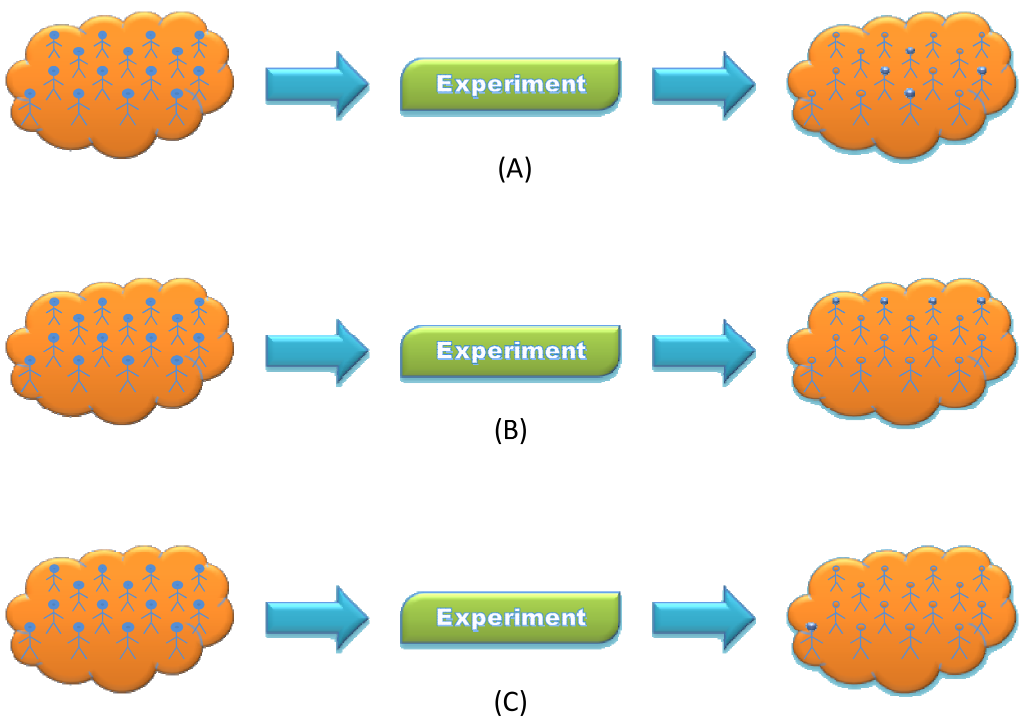


Figure 3-14: Sampling from a Population.

➢ The experiment (sampling method) results in an un-biased stochastic sample because more than one row and column are represented. That can be considered a good representative of the population. As indicated in the previous chapter, additional sampling may result in a different sample and hence a different mean (Figure 3-14, A). This highlights the process followed in this research to obtain an un-biased stochastic sampling.

- The experiment results in a biased stochastic sample where the selected objects are all in the top row. That may be considered a bad representative of the population. It is important to examine why it is biased to avoid biased samples in future similar experiments (Figure 3-14, B). This research effort is designed to avoid a biased stochastic sampling.
- The experiment results in a deterministic sample by selecting only one object. This sample can be biased and accordingly give misleading results (Figure 3-14, C). Unfortunately, this may or may not be a good representative of the population as this object may be an outlier (Figure 3-12). This highlights the deterministic approach currently used in WMS.

In the "experiment" of this research, sampling (sample size = n) was done from the general input parameter population through stochastic GSSHA. The sample size "n", was determined at the first step when we specified the number of runs for GSSHA. CE-QUAL-W2 ran for the same number of runs and consequently generated an "n" number of output data sets. The "n" output data sets were used to build the output population distributions needed for credible intervals.

As indicated in Chapter 2, a Bayesian approach was used in this research. The following is a list summarizing the subjective information used in this research:

- Selecting "some" GSSHA variables.
- Selecting a RANGE of values for these variables/parameters.
- Linkage in itself is subjective.

- The fact that the GSSHA input parameters modeled stochastically will affect the "population" of the output parameter.
- We apply the concept of "degrees of belief" in the sense that we accept the assumption that some distributions will be close to Normal or at least the normality, known variance underlying a particular sampling and frequency distribution, indicate that this distribution can be given a subjective interpretation.

3.1.7 Parameter Selection

Conceptually, the framework developed in this research is set so that all parameter sets in GSSHA can be simulated stochastically. The modeler, based on watershed information, parameter uncertainty and importance and the available computer resources, decides which parameters to consider stochastically. Obviously, more parameters modeled stochastically require more computing resources and/or more time needed for GSSHA to run the simulations.

The framework is run-homogenous; i.e. each parameter selected for stochastic simulation would have the same number of runs with different values, based on the parameter. The number of runs is not parameter-specific but based on the total number of parameters considered stochastically especially if LHS is selected.

It is important to determine which parameters to stochastically vary in the simulations. Parameter importance analysis, previous experience, knowledge of the area, previous models, or sensitivity analysis should generally be used to determine which parameters should have higher priority for stochastic treatment and which parameters can be used deterministically (Salah, et al., 2005-b). In the case studies tested in this research, previous experience and knowledge of the area were used to determine the parameters used stochastically.

Land use changes, ground water interaction and soil types could be considered, for example. By evaluating scenarios and defining the required accuracy level, the approach can be tailored and focused towards meeting the objectives of the modeler and enhance the process and reduce the overall simulation time.

This probabilistic approach helps in representing the actual "population" of each stochastic parameter properly by selecting values from the pseudo-population of that particular parameter. This will help eliminate and quantify the uncertainty issues modelers usually have in representing a population of, for instance, initial soil moisture which varies greatly and has a major impact on runoff predictions (Downer, et al., 2006). Choosing a PDF that best represents each parameter (best pseudo-population) and selecting values from the designated PDF does not actually reduce uncertainty, yet, it accounts for it and helps address the resulting uncertainty in output (Salah, et al., 2005-b).

3.1.8 Parameter Importance

Normalizing the sensitivity of the model to a parameter by its uncertainty, results in a parameter importance value. Some parameters exhibit a large influence on model results, but at the same time, possess very low uncertainty and hence would have very low importance and therefore would not need to be modeled stochastically. As an example, most hydrologic models are very sensitive to small changes in water viscosity, however, there is almost always a very low uncertainty associated with viscosity values of water. Hence while models might be very sensitive to viscosity, viscosity has a relatively low importance compared to other model input parameters (Williams, et al., 1995).

One of the major enhancements in the parameter selection proposed in this research is addressing the importance of each parameter. Parameters that are found to exhibit large influence on the model are not necessarily important in the simulation. When a model is said to be sensitive to changes in some parameters, then these parameters usually exhibit large influence on the model. For example, the model may be sensitive to a change in value, but they are well known so there is little uncertainty in this value, thus the parameter is considered “less important” (Williams, et al., 1995).

In this case, the PDF used to define the stochastic values of this model would not be as wide as other parameters that might exhibit a higher importance.In summary, the linkage (Figure 3-15) is organized in 4 phases:

1. Phase I: Stochastic GSSHA
    - Step 2: involves the stochastic input to GSSHA
    - Step 3: running GSSHA stochastically
    - Step 4: flux files are generated.
2. Phase II: GSSHA/CE-QUAL-W2 link
    - Step 1: spatial linkage where a segment ID index map is generated.
    - Step 5: temporal linkage where the flux files are broken down.
3. Phase III: Stochastic-Driven CE-QUAL-W2
    - Step 6: when CE-QUAL-W2 runs multiple times, each using a newly modified run-specific control file, and a specific set of input files for

the specific run and a static bathymetry file (and if applicable other input files).

4. Phase IV: Stochastic Output

   - Step 7: multiple CE-QUAL-W2 outputs are read into data sets into WMS.
   - Step 8: a gage is generated in WMS at a desired location in the water body profile and a time series plot, with credible interval, is generated.

3.2 Software Development

This section presents the software development to implement the conceptual framework outlined above. As mentioned earlier, the primary software program that is used in this research is WMS (Nelson, 2008). WMS has existing model interfaces for GSSHA and CE-QUAL-W2 and it is used in two ways:

- As a platform to implement the required linkage of the two models and enhancements to the existing interfaces to be able to utilize the tools, and
- A tool for analyzing and testing the developed tools for the case studies.

By linking GSSHA and CE-QUAL-W2 we can create multiple scenarios that can run to develop a range of probable results. The WMS drainage coverage is used to delineate watersheds and sub-basins with associated land use and soil coverages to define the different activities within the watershed. A GSSHA model is built for the watershed to best model the land portion of the watershed. Feature polygons; i.e. branch and segment coverages, can be used to represent water bodies in a CE-QUAL-W2 model. By integrating these processes directly through the conceptual model, the linkage between

GSSHA and CE-QUAL-W2 can be defined spatially and temporally (two dimensional grids of GSSHA surrounding lakes/river polygons automatically linked to CE-QUAL-W2 water bodies) (Figure 3-15) and (Figure 3-16) which is the basis of this research.

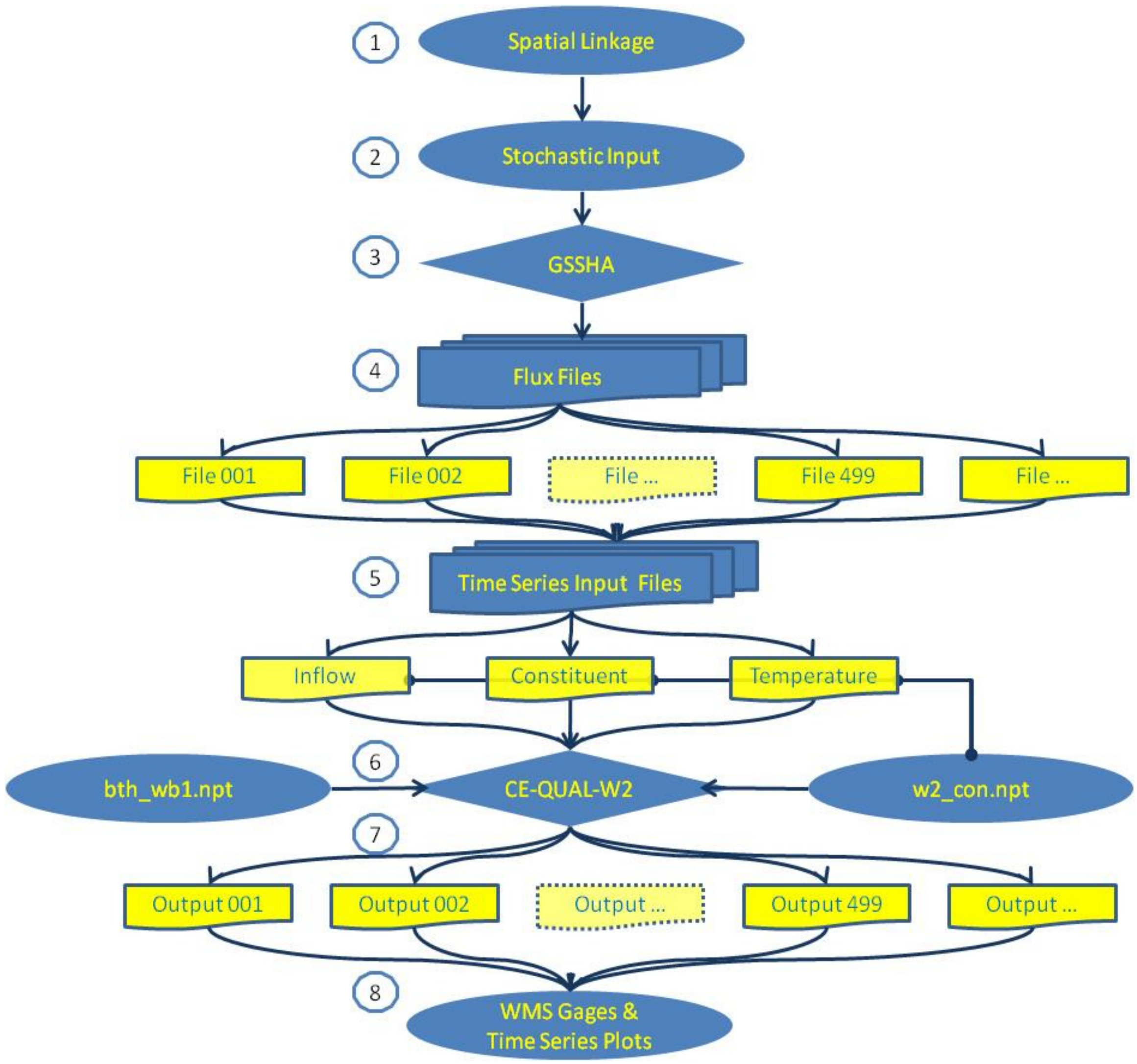


Figure 3-15: Overall Conceptual Framework.

The more accurate the supporting data, such as soil and land use, the more reliable is the output. With respect to software development, and consistent with the conceptual framework outlined above, this research is divided into four phases:

1. Updating the GSSHA interface in WMS to incorporate stochastic modeling,
2. Linking GSSHA to CE-QUAL-W2 for a single case (i.e., single GSSHA output aggregated to be the input for a CE-QUAL-W2 model),
3. Upgrading the link to include all the stochastic runs from GSSHA into CE-QUAL-W2, in what is called the "stochastic-driven" phase.
4. The aggregation of results and post processing.

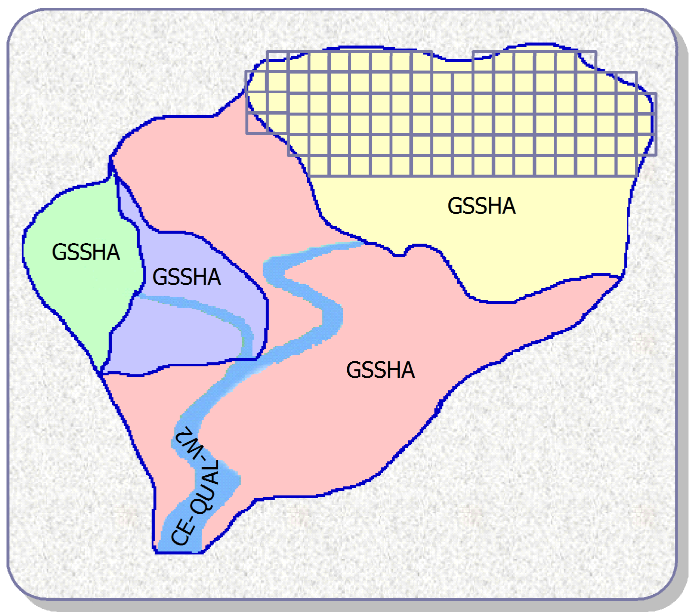


Figure 3-16: Conceptual Linking of GSSHA to CE-QUAL-W2.

3.2.1 Phase I: Stochastic GSSHA

In phase 1, called the "stochastic GSSHA", GSSHA parameters are considered stochastically. This phase involves two main stages:

1. GSSHA input parameter selection, and
2. Statistical simulation: this involves the simulation technique and the selection of the most appropriate PDF to represent each parameter (i.e., creating the pseudo-population). Most of the time, the modeler selects the PDF during the

modeling process. Other distributions, like Normal, Log-Normal, Beta and user defined distributions have been identified and incorporated in the WMS interface to give modelers some flexibility. Most of the time, however, the hydrologic parameters are found to follow Uniform, Normal, Log-Normal or Beta distributions (Clarke, 1994).

A dialog (Figure 3-17) is developed to enable modelers to choose a set of parameters and the associated PDFs. It also enables modelers to choose the method of sampling the PDF and the number of GSSHA runs desired. In this dialog, modelers do not necessarily have to choose PDFs for all parameters, but only the ones they think are most important.

Modelers can add or delete parameters to be stochastically represented. Default values (Table 3–2) for the mean, minimum, maximum and standard deviation are populated automatically. As indicated earlier, the simulation method used is the Monte Carlo Simulation whereas the sampling, from the PDF, is random, LHS or user defined. Modelers can re-populate runs with new sampling values or edit the individual values for each of the fields in the dialog.

As shown in this dialog, modelers determine the number of simulations, which are also considered the "sample size". This value depends on the selection method used to sample values from the PDF. For example, if the LHS is selected, the modelers have no direct control on the number of simulations. Instead the number of simulations is determined by the number of instances and the number of the stochastic parameters selected in this dialog. The number of instances in this case represents the number of equal areas sub-sections of the respective distribution is divided into. The total number of simulations in this case is the product of the number of sub-sections for each parameter. Therefore, if capillary head has "A" sub-section and hydraulic conductivity has "B" sub-section, the total number of simulations is A× B.

The dialog shows each stochastic parameter with a key value (Figure 3-18, A) which must always be a negative integer. This key is used in the standard input file within the GSSHA interface in WMS (Figure 3-18, B). That way, WMS and GSSHA would model it stochastically by using one of the listed values from within the dialog. These non-unique keys can be used by multiple parameters.

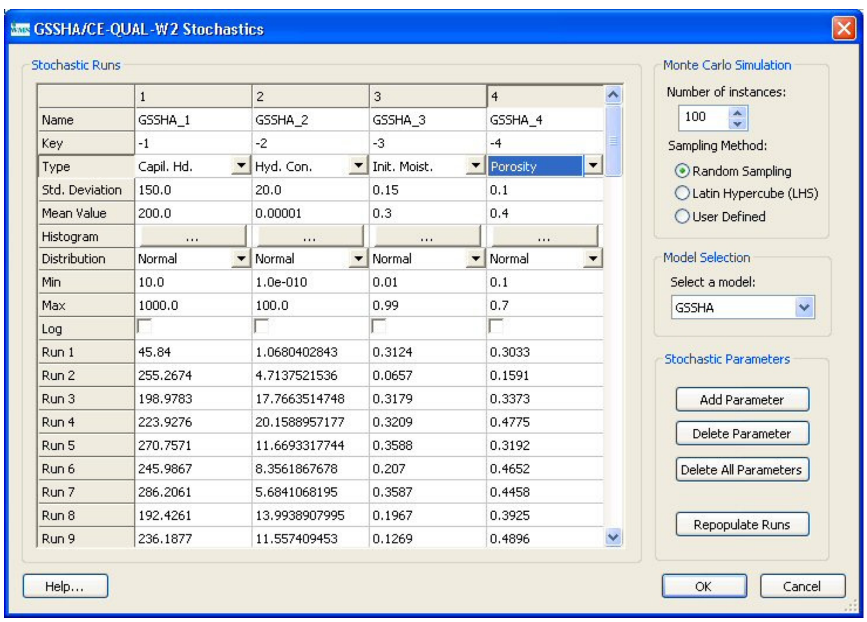


Figure 3-17: Stochastic GSSHA Dialog.

Default values (Table 3–2) for the parameters have been estimated based on the common range of values. The parameters names are designed to be unique; i.e. WMS does not allow duplicate keys.

Concurrent research efforts and GSSHA model data requirements (Byrd, et al., 2005) identified key input parameters that were found to be most important to the model in a lot of applications. These parameters may include, but are not limited to, the following:

- Capillary head,
- Hydraulic conductivity,
- Initial moisture conditions,
- Manning roughness,
- Porosity,
- Surface Albedo,
- Interception Coefficient,
- Rainfall.

(A) (B)

Figure 3-18: Stochastic Parameter Handling in GSSHA.

Table 3–2: Default Values, Minimum and Maximum, for Stochastic Parameters.

Parameter	Units	Min	Max	Standard Deviation
Capillary Head	cm	10	1000	150
Hydraulic Conductivity	cm/hr	1e-10	100	20
Initial Moisture	--	0.01	0.99	0.15
Manning's "n"	--	0.005	0.500	0.10
Porosity	--	0.10	0.70	0.10
Precipitation	mm	10	200	25

In this research, parameter selection is assumed to be based, solely, on the choice of modelers, as opposed to a pre-defined set of parameters. This is assumed because an important parameter for model A is not necessarily an important parameter for model B. Uniqueness of models and conditions forces modelers not to use preset or universal conditions. The framework developed in this research demonstrates the ability to model watersheds in an integrated scheme. However, it is important to realize that this framework, like other modeling efforts, has limitations and cannot be universally applied.

Once the modeler has identified which parameters will be modeled stochastically and has identified the associated model parameters, WMS will help with automating the GSSHA runs. In order to facilitate the researched methods, a tool to have WMS generate two additional files which are the parameter and value files, in addition to the regular project files. The parameter file lists all the parameters that are stochastically modeled whereas the value file lists all the associated values that are actually available and editable in the Stochastic GSSHA dialog (Figure 3-17).

The stochastic dialog provides input for these two files: (Figure 3-19):

1. Parameter File,
2. Value File.

These files are used by GSSHA to run in stochastic mode. The two files will be created and saved with the GSSHA project file once the modeler saves out the GSSHA project. The "Save GSSHA Project File" dialog has multiple tabs that handle each individual file category. In addition, a third file is indicated and saved. This "siminput.txt" enables modelers to reload the parameters and their values when the project is re-opened. Figure 3-20 show an example of a GSSHA project file. The project file is divided into four main sections:

1. Introduction: includes the title and any additional remarks
2. Batch Mode: includes where the parameter and value files are saved (Figure 3-20, A). This section is optional and is not required by GSSHA for a deterministic run.
3. Regular Input: includes the default input values for the run.
4. GSSHA to CE-QUAL-W2: includes the flux file, Segment ID index map and the stream nodes file (Figure 3-20, B). This section is optional and not required by GSSHA for a stand-alone GSSHA run; i.e. not linked to CE-QUAL-W2.

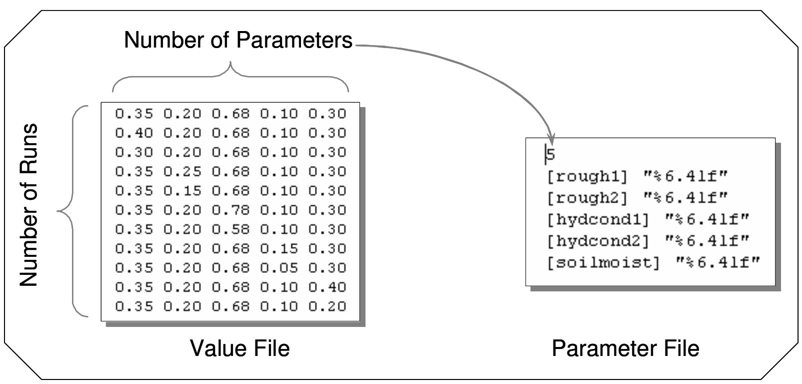


Figure 3-19: Value and Parameter Files Example.

The flux file is the GSSHA output that compiles the output to each individual segment for each time step. The flux file is the file that will be broken down to individual CE-QUAL-W2 input files for each segment. The stream nodes file is another GSSHA output that compiles the streams input to the water body on all the nodes selected by the modelers. This file may not be generated if the modeler elects to not model streams in

GSSHA. Contrary to these two files, the index map is the only file in this section that is not a GSSHA output, rather it is a GSSHA input built on the CE-QUAL-W2 model.

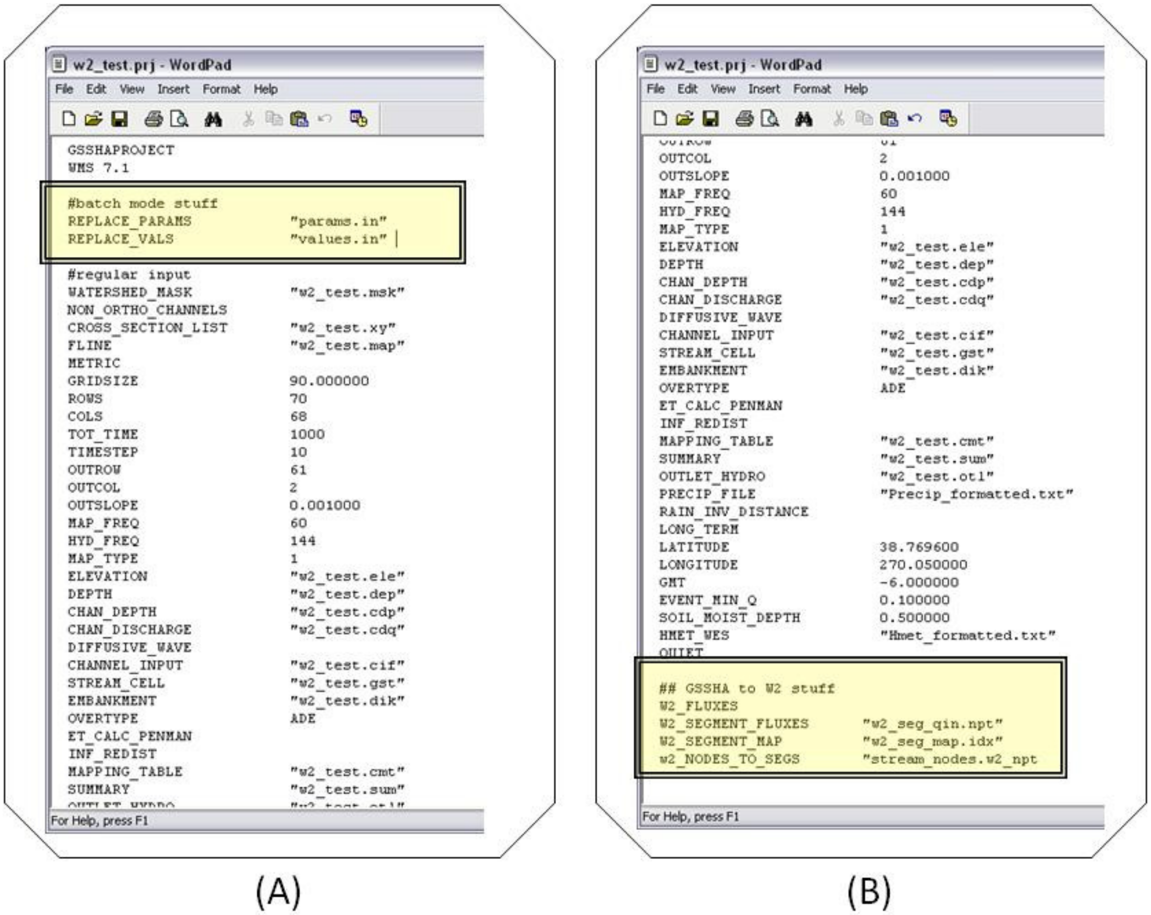


Figure 3-20: Sample GSSHA Project File.

All of the five files indicated in the above two cards must be present in their respective locations for a successful stochastic GSSHA run that is linked to CE-QUAL-W2 and obviously a stochastic version of GSSHA is needed to run in batch mode.

3.2.2 Phase II: GSSHA / CE-QUAL-W2 Link

In this section, the term "GSSHA output" will be used to refer to both water quantity and quality. However, the current version of GSSHA (v4.3C) does not have an extensive water quality module (Downer, et al., 2006). For this reason, this research focuses on GSSHA water quantity output only. However, the methodology developed is parameter-independent and can be applied to water quality constituents as well, once a robust nutrient module is developed in GSSHA.

GSSHA output is used to incorporate the inflows to CE-QUAL-W2 as a flow tributary inflow file (*_qtr.npt). A tributary inflow file is where modelers can specify what additional input is contributing to the receiving water body either through a stream inflow or direct from GSSHA cells. GSSHA is designed to generate one tributary inflow file for all segments and the linkage is designed to break this file into individual segment files. Once the GSSHA nutrient module is in place, a constituent tributary inflow concentration file (*_ctr.npt) and temperature (*_ttr.npt) (Cole, et al., 2007) can be generated and incorporated in the analysis as necessary.

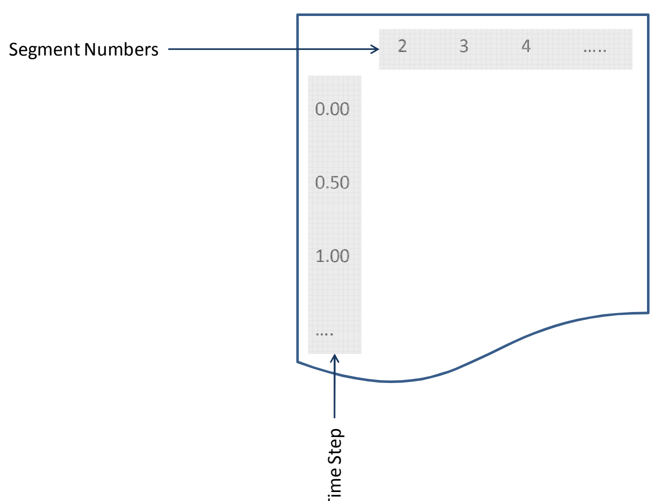
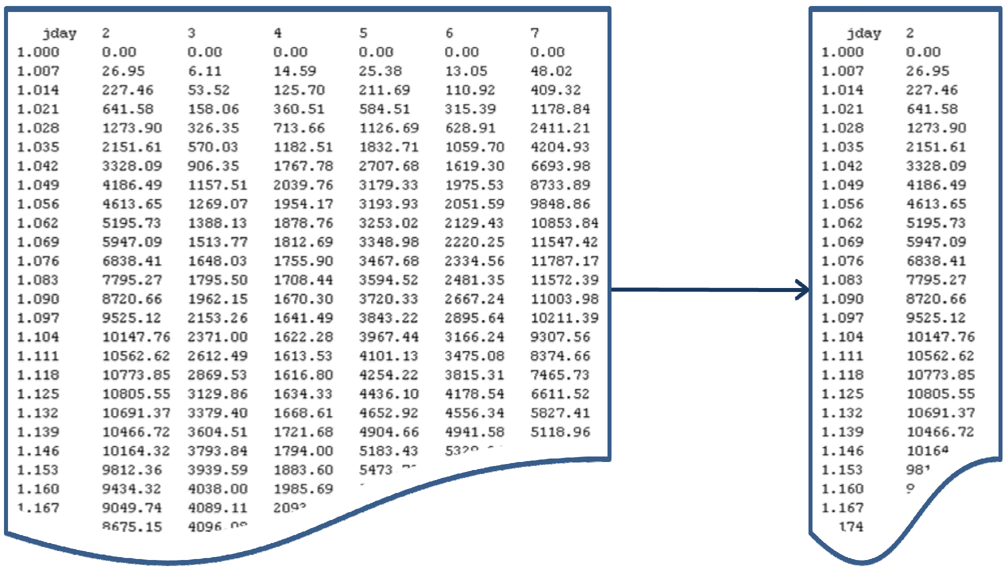


Figure 3-21: Flux Files Generated by GSSHA.

There are two features required for the link between CE-QUAL-W2 and GSSHA. The first feature is a direct feature of the spatial linkage. This is to create an index map of all the segments in the CE-QUAL-W2 model. The index map should be created using the same grid cell size of the GSSHA model. This cell size restriction is essential to establish the linkage properly. This feature is programmed as an item in the CE-QUAL-W2 menu in WMS. As indicated in Figure 3-20, the segment ID index map must be generated before GSSHA runs. It is necessary for GSSHA to map the necessary files for the link.



jday	2	3	4	5	6	7
1.000	0.00	0.00	0.00	0.00	0.00	0.00
1.007	26.95	6.11	14.59	25.38	13.05	48.02
1.014	227.46	53.52	125.70	211.69	110.92	409.32
1.021	641.58	158.06	360.51	584.51	315.39	1178.84
1.028	1273.90	326.35	713.66	1126.69	628.91	2411.21
1.035	2151.61	570.03	1182.51	1832.71	1059.70	4204.93
1.042	3328.09	906.35	1767.78	2707.68	1619.30	6693.98
1.049	4186.49	1157.51	2039.76	3179.33	1975.53	8733.89
1.056	4613.65	1269.07	1954.17	3193.93	2051.59	9848.86
1.062	5195.73	1388.13	1878.76	3253.02	2129.43	10853.84
1.069	5947.09	1513.77	1812.69	3348.98	2220.25	11547.42
1.076	6838.41	1648.03	1755.90	3467.68	2334.56	11787.17
1.083	7795.27	1795.50	1708.44	3594.52	2481.35	11572.39
1.090	8720.66	1962.15	1670.30	3720.33	2667.24	11003.98
1.097	9525.12	2153.26	1641.49	3843.22	2895.64	10211.39
1.104	10147.76	2371.00	1622.28	3967.44	3166.24	9307.56
1.111	10562.62	2612.49	1613.53	4101.13	3475.08	8374.66
1.118	10773.85	2869.53	1616.80	4254.22	3815.31	7465.73
1.125	10805.55	3129.86	1634.33	4436.10	4178.54	6611.52
1.132	10691.37	3379.40	1668.61	4652.92	4556.34	5827.41
1.139	10466.72	3604.51	1721.68	4904.66	4941.58	5118.96
1.146	10164.32	3793.84	1794.00	5183.43	532	
1.153	9812.36	3939.59	1883.60	5473		
1.160	9434.32	4038.00	1985.69			
1.167	9049.74	4089.11	209			
	8675.15	4096				

jday	2
1.000	0.00
1.007	26.95
1.014	227.46
1.021	641.58
1.028	1273.90
1.035	2151.61
1.042	3328.09
1.049	4186.49
1.056	4613.65
1.062	5195.73
1.069	5947.09
1.076	6838.41
1.083	7795.27
1.090	8720.66
1.097	9525.12
1.104	10147.76
1.111	10562.62
1.118	10773.85
1.125	10805.55
1.132	10691.37
1.139	10466.72
1.146	10164
1.153	981
1.160	9
1.167	
174	

Figure 3-22: Flux File Break Down.

The second feature serves both the spatial and temporal linkages and it should come at a later stage of the modeling process. This feature de-aggregated the generated flux files from GSSHA and breaks it down to multiple CE-QUAL-W2 input files. These files are broken out by segments (Figure 3-22).

A sample flux file is shown in Figure 3-21, the first row in the file shows the segment numbers (based on the index map), and the first column shows the time steps (based on GSSHA time steps). Figure 3-22 illustrates how a flux file is broken into multiple CE-QUAL-W2 input files; i.e. one for each segment.

As mentioned earlier, the linkage is divided into two main levels:

1. Spatial linkage: The CE-QUAL-W2 segment coverage (Figure 3-23) in WMS is used to build a GSSHA segment IDs index map (Figure 3-24). Any GSSHA grid cell that is not overlaid with a segment polygon is given an index of zero, otherwise it is given the segment ID it is overlaid by. This way, a water body boundary is determined by the fact that any flow coming from a zero-valued cell to any non-zero-valued cell is a flow from GSSHA to CE-QUAL-W2 water body. The CE-QUAL-W2 segment index map tool, developed in this research, creates an index map for each CE-QUAL-W2 segments in the segment coverage. Before a CE-QUAL-W2 segment index map is created in WMS, all segments are checked to determine if they are mapped to CE-QUAL-W2 branches. On the other hand, branches are checked to see if each branch has at least one segment assigned.

The first step in generating a Segment ID index map is done in the WMS interface for CE-QUAL-W2. The WMS interface checks to see if there is a GSSHA grid that covers the segment coverage in CE-QUAL-W2 and then generates the GSSHA index map. It also checks to see if both branch and segment coverages are already defined and if a CE-QUAL-W2 model is already initialized.

As noted in Figure 3-24, we can see a GSSHA grid overlaying a polygon feature that represents a lake. The spatial linkage is manifested by the following features:

- Zero-valued cells indicating cells that are not within the water body.
- Non-zero based cells which are within the water body boundary. Values of these cells represent the Segment ID that the cell lies within.

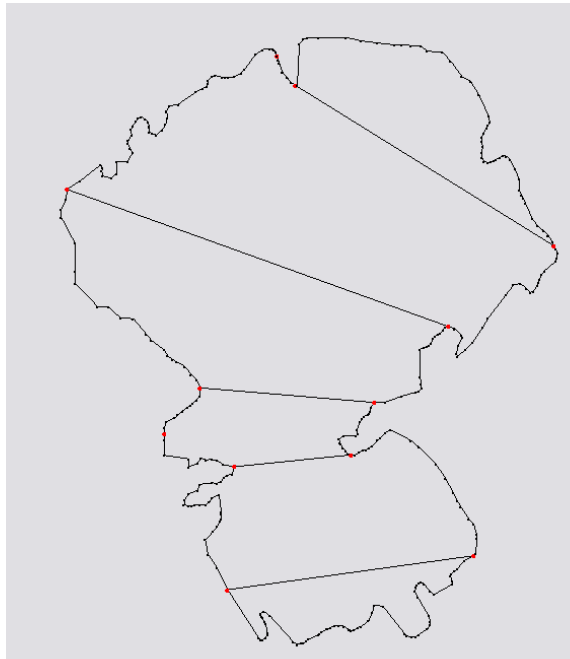


Figure 3-23: Segments in a CE-QUAL-W2 Model.

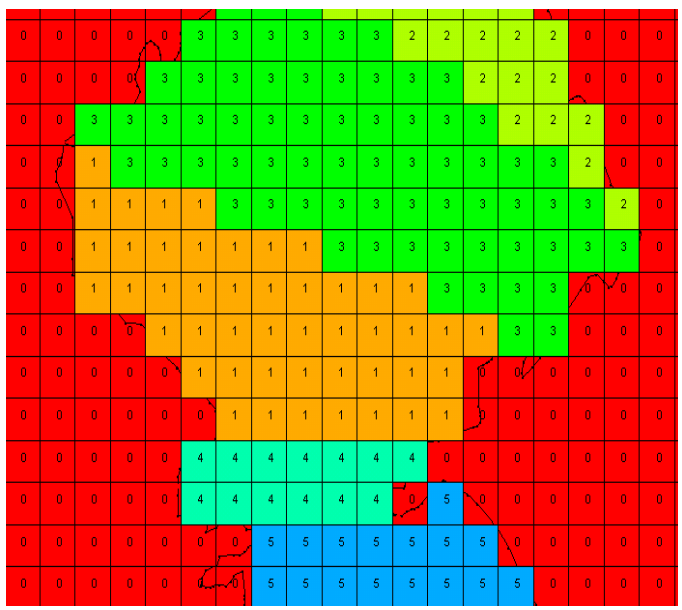


Figure 3-24: Spatial Linkage, Segment ID Index Map.

- Water body arc which is the water body boundary as modeled by the CE-QUAL-W2 branch and segment coverages.

➢ Major streams out falling in the water body are modeled through the stream nodes file. The stream node file (Figure 3-25) defines what “Link” discharges into which segment in the CE-QUAL-W2 model.

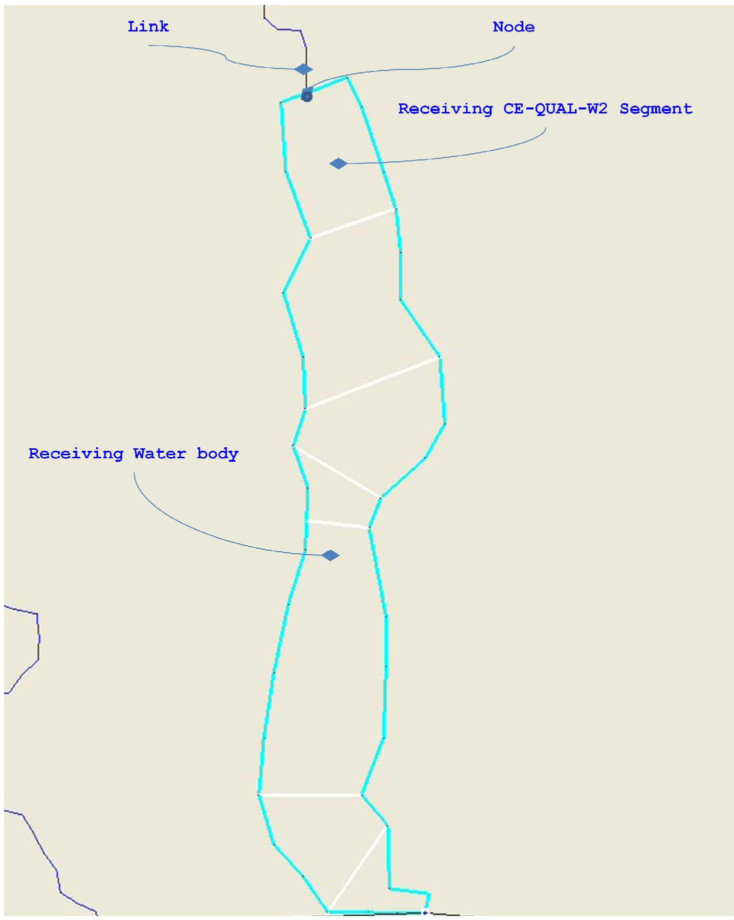


Figure 3-25: Link-Node-Segment Map.

2. Temporal linkage: the temporal linkage is done through various time steps. The GSSHA-generated flux files (Figure 3-21) at the specified cells at the time steps defined, is then computed and aggregated for CE-QUAL-W2 segments (Figure 3-22). In this research, a tool that converts this output to the format the respective CE-QUAL-W2 input file requires (i.e. inflow, concentration and temperature).

In Summary, the linkage, so far, is done at two interdependent fronts:

1. WMS: Index maps are generated based on the segment IDs and passed to GSSHA. At the same time stochastic values for a set of modeler-chosen parameters are passed to GSSHA in the parameter and values files.
2. GSSHA: GSSHA generates fluxes aggregated by segment. There are some additions in the GSSHA code to help the data input and output interchange smoothly.

If both phases are implemented correctly, a number of flux files, equal to the number of simulations (sample size), will be generated. Similarly, the stream inflow files are used as the major contributing flows from stream network (if modeled in GSSHA). These files are the basic building block for the following phase.

3.2.3 Phase III: Stochastic-Driven CE-QUAL-W2

This phase is referred to as the stochastic link where multiple GSSHA outputs generate multiple input files for CE-QUAL-W2. Once the spatio-temporal and data linkage is developed in phase II, the interface is ready for multiple CE-QUAL-W2 runs for the same model, based on the stochastic inputs to GSSHA. The multiple GSSHA runs generated by the stochastic values for GSSHA parameters are used to generate some of the input files required by CE-QUAL-W2. At the end of a stochastic GSSHA run, GSSHA generates a set, equal to the number of runs selected by the modeler, of segment flux files which contain the water quantity input required by CE-QUAL-W2. Each one of these files is broken down to multiple files that will be used by one CE-QUAL-W2 run (Figure 3-26).

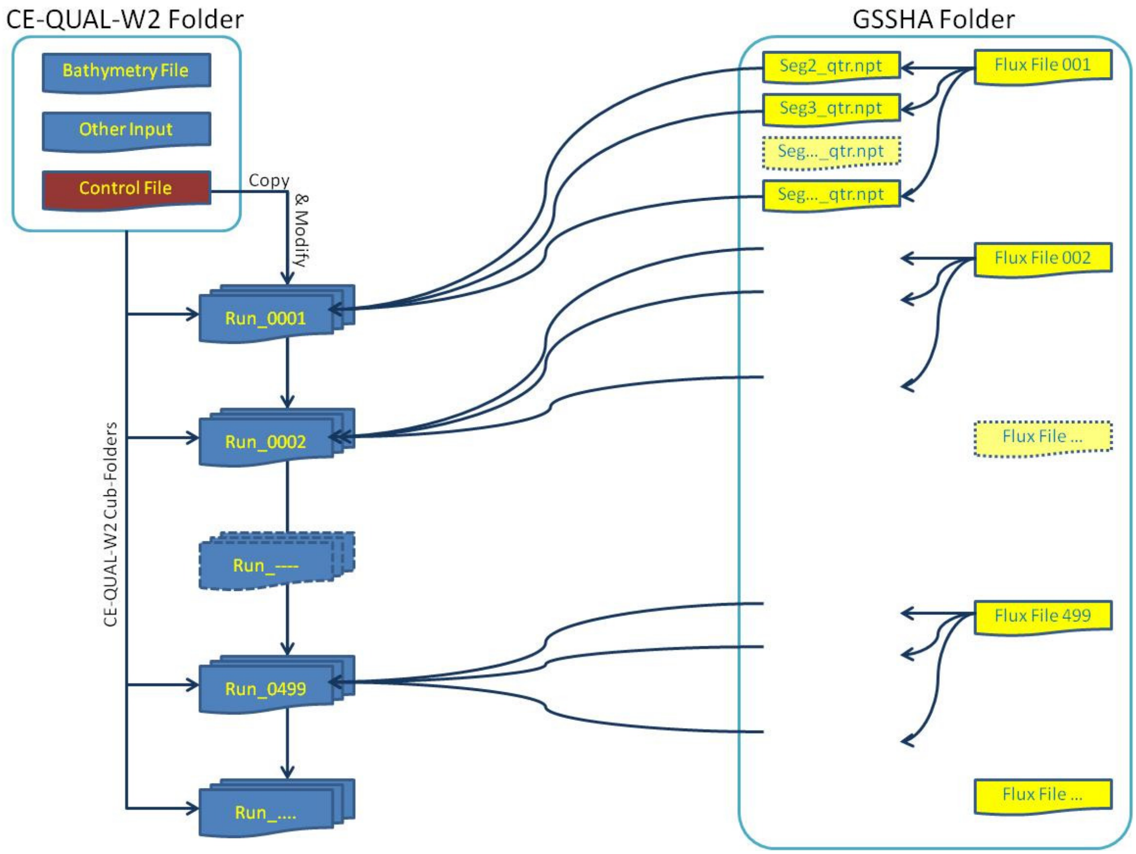


Figure 3-26: Stochastic-Driven CE-QUAL-W2.

As indicated in section 3.1, the bathymetry file does not change for different CE-QUAL-W2 runs. This is because the stochastic runs do not alter any of the input parameters of the bathymetry file. In the same process the original CE-QUAL-W2 control file is copied multiple times, equal to the number of runs, in separate sub-folders. Once copied, each control file is modified to reflect the changes pertinent to the stochastic input and the newly generate input files. Each of these sub-folders becomes a separate CE-QUAL-W2 run that will eventually contain the CE-QUAL-W2 output.

The control file may have small changes that correspond to the changes occurring in the model based on the stochastic inputs. The main difference in the control file is the different path of the input files. So, it is basically multiple control files and single bathymetry file, in addition to the multiple tributary inflow files. This defines multiple runs of CE-QUAL-W2.

WMS modifies the control file, specifically the absolute paths of the other input files. The broken down flux files are written out in their respective run sub-folder; e.g.:

- C:\...\Parent Folder\Run_0002\w2_seg2.npt,
- C:\....\ Parent Folder\Run_0002\w2_seg3.npt,
- …etc. All other input files are saved in the parent folder (e.g., C:\...\Parent Folder\bth.npt).

This phase is the most time-consuming phase of the linking process in terms of implementation and model run time. Depending on the size of the watershed, grid cell size, number of segments/branches, and the accuracy needed, a complete stochastic run can take a relatively long time (Section 3.3). As indicated in the following chapter, modelers in some cases have to increase the grid cell size and/or do some other model simplifications, at least at the initial stages, to produce a reliable model.

The distributed version of CE-QUAL-W2 comes with a simple status dialog that gives interactive summary information about the run (Cole, et al., 2007). For the purpose of this research, using the dialog would not be efficient, as modelers would not want to manually initiate numerous instances of the dialog for each run. Additionally, there is no reason to duplicate the CE-QUAL-W2 executable multiple times in each sub-folder of the runs.

Alternatively, a generic CE-QUAL-W2 executable was compiled that could be called directly from WMS to execute the model. The executable is saved with the other WMS model executables. As part of this research, the source code was programmatically modified, compiled and linked using FORTRAN 90 to take an argument which is the control file full path. That way, the executable could be called multiple times, and passed the full path of the respective control file, for each CE-QUAL-W2 run without any redundancy. Based on that, modelers can run CE-QUAL-W2 from within WMS either for one run or for stochastic runs.

3.2.4 Phase IV: Stochastic Output

After a successful completion of the runs, WMS reads the CE-QUAL-W2 solution files equivalent to the number of runs/simulations that are saved in the sub-folders as indicated in Figure 3-26. For a stochastic solution, the interface requires the path of the parent folder, containing all of the sub-folders for all the runs, as opposed to the full path of the CE-QUAL-W2 output file (as is the case for a deterministic solution). For each solution to be read, a longitudinal profile needs to be created first. Each branch of the CE-QUAL-W2 will have a profile showing the segments and layers which will be later used to display contours of the constituent analyzed.

Each solution is used to generate a dataset for the selected constituent(s) provided that these constituents are simulated in the CE-QUAL-W2 run. All the read datasets are used to estimate the credible interval (i.e. lower and upper bounds). The level of credibility is determined by the user (Figure 3-27). As indicated in the previous chapter, the higher the credible level, the more confident modelers are in the results and the wider the interval is.

All the generated datasets, from the multiple runs, and the selected credible interval are used to generate three additional data sets:

1. Credible interval lower bound.
2. Mean.
3. Credible interval upper bound.

These generated datasets will be utilized to build the stochastic time series. The width of the credible interval is a direct function of the desired credibility. Higher credibility generally results in a wider range and vice versa.

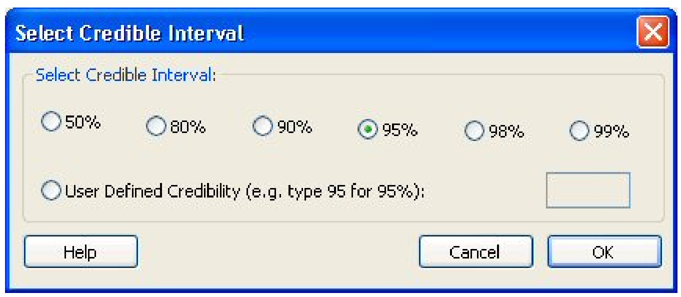


Figure 3-27: Credible Interval Dialog.

In summary and in light of the overall modeling process in WMS (Figure 2-2), batch runs (Figure 3-15) are planned to comprise the four following ordered steps:

1. Running GSSHA stochastically to generate multiple outputs.
2. Aggregate GSSHA output which produces multiple input files equivalent or less than the original number of runs.
3. Run CE-QUAL-W2 stochastically.
4. Display the outputs of CE-QUAL-W2 on longitudinal profiles and/or time series plots of the water body.

4 Results

The developed approach is an original effort to integrate a land surface model and a water model into a stochastic integrated modeling environment in order to address the uncertainty from both water quantity and quality perspectives. It contributes to the integrated water resources modeling by providing a tangible tool, and at the same time addresses uncertainty, which is not given the due attention in current practice. This approach has some advantages to similar integrated water resources modeling environments in a GIS context. As seen later in this chapter, the developed approach has shown some noticeable advantages in the overall modeling process.

4.1 Integrated Water Resources Modeling

The main advantage of this approach, as far as the integrated water resources modeling is concerned, can be summarized as follows:

1. Most integrated models are not stochastic and stochastic models are not integrated.
2. The strength of statistical simulation overcomes the lack of data. In most cases, modelers do not have values for all the parameters required by the model. What is even worse is that modelers do not have 100% certainty, or even close, for some of the values of these parameters. Even if modelers are

certain of a specific value, it might not be the best value to represent the population of that parameter in a particular location at a particular point of time. The statistical simulation allows for the selection of the parameter values from the given PDF and simulating the model multiple times would produce outputs that incorporate uncertainty and credible intervals as opposed to a rigid single answer that is suspect because of uncertainty.

3. This approach is scalable in time, space, parameters and tools.
    - The time domain of the model can expand depending on the available computing resources and the accuracy needed. This means modelers can use a larger duration of the model to simulate and/or reduce the time step to capture changes in model output that would not be captured by a "coarser" temporal resolution.
    - Similarly, the spatial context of the model could be enhanced. Finer resolution could be used to address more detailed and localized issues if needed and data are available.
    - Grids can be rectilinear, or what could be referenced as "smart grids", to reduce data requirements as possible. More parameters can be used to demonstrate the stochastic nature in modeling.
    - Also, more statistical simulation techniques such as dynamic simulations can be used as opposed to the one used (even though there are two sub groups of parameter selection). Also the available PDFs can be expanded to encompass a wider selection of distributions that might provide a better fit to certain parameters.

4. Automating the separate steps (i.e. pre-processing of data, aggregating information) saves time especially for repeated model trials.
5. The proposed system is ideally suited for TMDL studies, UAAs, and to model point and non-point source pollution simultaneously.

4.1.1 Land-Water Interface

The two models, GSSHA and CE-QUAL-W2, are used interactively to generate a complete integrated run for the interface between the land and water portions of the watershed. Some GSSHA parameter values are picked from a PDF or PMF for continuous and discrete parameters respectively. This produces a PDF for the output constituent under investigation (or outflow) as opposed to single value at each time step. Moreover, this generated PDF is time and location-dependant.

If a time series is needed, the third dimension in Figure 4-1 can be represented as time (as opposed to different scenarios) and the PDF would rather be a PDS.

Because each PDF is time, parameter and location dependant, the PDS (Figure 4-1) varies along the third dimension time or scenarios. The PDS can be considered discrete or continuous. A PDS is considered continuous if values between the component PDFs are feasible and they can actually be used to represent parameters at the specific time, parameter and location.

4.1.2 Scenario Development

The developed approach can easily be used in a scenario development where some of the parameter values can be manually entered as opposed to selection from a PDF, based on a specific scenario. This is referred to as a low level scenario development which is primarily meant to incorporate some land use/management changes to see how that would affect the overall objective of the study at hand. High level scenario development is when statistical simulations take place in developing the scenarios themselves. Typically, the scenarios depend largely on the overall objective of every unique study.

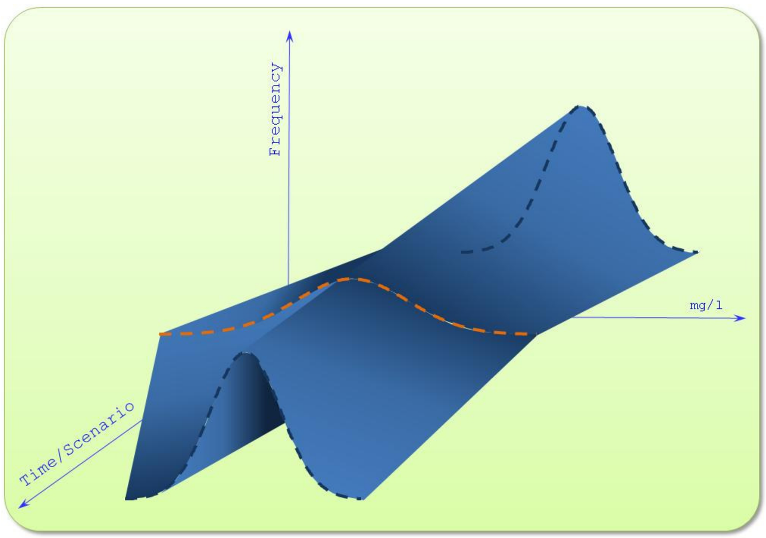


Figure 4-1: A Probability Density Surface.

4.1.3 Cascaded Reservoirs

In some cases, a watershed might have two (or more) reservoirs inter-connected by a river. This system is sometimes referred to as cascaded reservoirs (Figure 4-2). This is an added sophistication to the approach outlined above. However, a sophisticated system like this can still be modeled, using the developed approach, with greater computational power to reduce the time needed for a successful and complete linked run.

Cascaded reservoirs are a common feature of the watershed hydrology where reservoirs A or B (Figure 4-2) can either be man-made or natural. An example of this is the Eau Galle watershed.

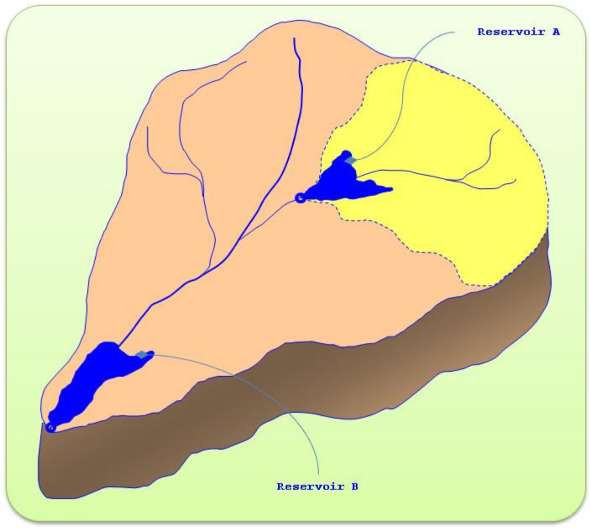


Figure 4-2: Cascaded Reservoirs Modeling, A and B.

4.1.4 Statistical Distribution

The default statistical distribution used throughout this research is Beta distribution. To evaluate if using Beta distribution would be different than using Normal or Log-Normal distributions, I compared Beta distribution to both the Normal and the Log-Normal distributions. This comparison was done using raw values generated from WMS before it was used to run GSSHA stochastically Table 4–1. In addition, the comparison was done to compare GSSHA output as a result of both distributions. Both comparisons are made to track the difference from the raw data to model output.

Values were generated for four GSSHA parameters; i.e. hydraulic conductivity, capillary head, initial moisture and manning's n, using three statistical distributions; i.e. Normal, Log-Normal, and Beta (with shape factors of 2, 7) in WMS. Each set of values was pair-wise compared to another set using simple linear regression analysis. These comparisons indicated how different the distributions were at the input level (Table 4–1).

Table 4–1: Statistical Distribution Comparison – GSSHA Input.

Parameter	**p-Value**		
	Beta* vs Normal	**Beta* vs Log-Normal**	**Normal vs Log-Normal**
Hydraulic Conductivity	0.597	0.436	0.067
Capillary Head	0.199	0.218	0.248
Initial Moisture	0.576	0.266	0.383
Manning's n	0.403	0.764	0.104

* Shape factors of 2, 7.

This research aimed at looking at the stochastic input, as well as the effect of various stochastic distributions on the output. Thus it was important to verify that the used parameter values, generated by different distributions, are different. Table 4–1 indicates that there is strong evidence that hydraulic conductivity stochastic values generated from a Beta distribution are different from stochastic values generated from a Normal distribution (p-Value = 0.597). Similarly, there is strong evidence that stochastic values generated from a Beta distribution are different from those generated from Log-Normal distribution (p-Value = 0.436). Also, there is evidence that stochastic values generated from a Normal distribution are different from those generated from Log-Normal distribution (p-Value = 0.067). Even though the differences are imperative (since different distributions are used) we could see from Table 4–1 that using different distributions has a different effect on different parameters. For example, the p-Value comparing Normal and Log-Normal for Initial Moisture (0.383) is significantly higher than for the Hydraulic Conductivity (0.067).

Similarly, stochastic values generated by Beta distribution are different from those generated by Normal and Log-Normal distributions for capillary head, initial moisture and manning's n (Table 4–1).

In addition to comparing distribution at the raw data level, Beta and Normal distribution values were both used in separate GSSHA runs and the flux files were compared to see if stochastic runs using these two distributions would result in similar flux files. For this comparison, GSSHA fluxes resulting from Beta distribution raw values for each run for all segments are stacked to be statistically compared, using t-tools, to the equivalent list resulting from normally distributed raw values (Table 4–2).

Table 4–2: Statistical Distribution Comparison – GSSHA Output.

	CE-QUAL-W2 Segment						
	2	**3**	**4**	**5**	**6**	**7**	**8**
p-Value	0.321	0.149	0.965	0.951	0.275	0.684	0.893

High p-values in Table 4–2 indicate that there is no reason to believe that both distributions produced different flux files for all the seven segments of Eau Galle Reservoir.

Based on what we see on this analysis, it looks like the choice of the statistical distribution (i.e. Log/Normal, Beta) to represent the hydrologic input parameter values for GSSHA is not likely to affect the stochastic output substantially, especially the output used in generating CE-QUAL-W2 input. This may be explained by Normal distribution tails and outliers. In a Beta distribution tails are not unlimited. It must be noted, however, that this conclusion is valid for the above shape parameters of Beta distribution. Other research findings (Ashkar, et al., 1998) indicate that the choice of statistical distribution becomes critical when working with extreme values. Also, this analysis indicated that Beta distribution, through the use of shape factors, can be used to approximate normal and log-normal distributions.

4.1.5 Spatio-Temporal Uncertainty

Another major accomplishment in this framework is the ability to address uncertainty. The flexibility to report the output, either a time series or a constituent profile, on a credible interval or the median (current approach) adds to it. Modelers can still view a "one line" representation of a time series the same as with current techniques or better yet, display the range of values (Figure 4-3) and (Figure 4-13). Similarly, the final longitudinal profiles of CE-QUAL-W2 would have ranges of regions that, exceed water quality standards using the specified credible interval.

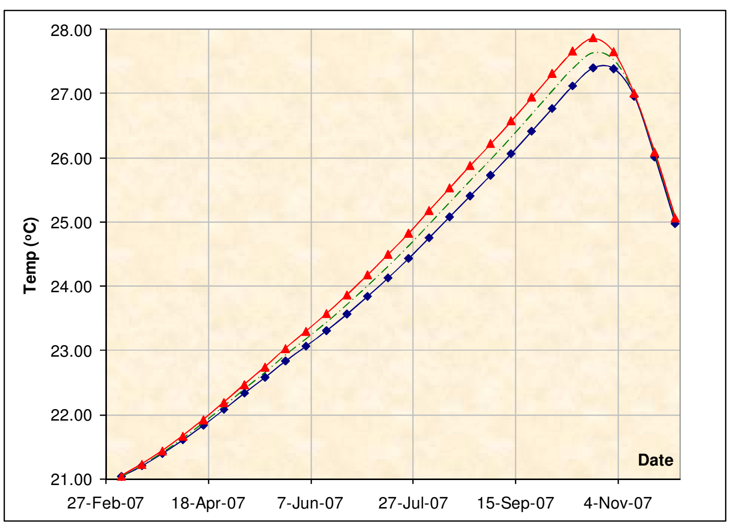


Figure 4-3: Temperature from Sample Run for Lake Zapotlan (95% CI).

For time series, instead of having a single value of the constituent under investigation, whether a pollutant concentration or flow of water, each time step has a range of values of minimum and maximum and a median, based on a certain credible interval. In that sense, modelers would be able to say that at a specific time step, there is a 95%, for instance, chance of having the value lie between the indicated lower and upper bounds (Figure 4-3).

The result of the linkage process is multiple data sets of the parameter(s) under investigation. These data sets are typically used to generate a time series plot of a parameter at a point of interest, for instance water withdrawal location from a reservoir. The population distribution that can be generated in the process at each time step can actually be used to draw a different inference. As an example, the generated time series may be used to infer that there is, for example, a 10% probability that the value of the parameter at certain time step/location will exceed a given concentration. This concentration can be the water quality standard or any other arbitrary threshold the modeler or decision maker has previously set.

In a selective withdrawal scheme, multiple time series plots can be generated along with the associated probabilities. This will result in a relative comparison of the different withdrawal locations out of the reservoir. In turn, this leads to a more informative decision of which withdrawal location in the withdrawal tower to use and which time step is the most suitable.

4.2 GSSHA / CE-QUAL-W2 Link

The two model domains with their strong underlying numerical representations of both the land and water have been linked in a GIS context. A two dimensional plenary grid (GSSHA) is efficiently linked to another two dimensional, but longitudinal, grid (CE-QUAL-W2).

4.2.1 Interface

Mapping/disctretizing one domain to the other is done from a horizontal two dimensional model to a vertical two dimensional model (Figure 3-1). This process can be generalized to any other set of models. Thorough understanding of the two model domains is essential in customizing this approach for other sets of models. The continuity of "manual" data dissemination between various models is a tedious and difficult task, especially if the process is bidirectional.

This research effort shows how WMS, as a common platform, utilizes pre-existing tools and functionalities to automate this interface and facilitate the linkage process.

As seen in Figure 3-18, a simulation input file is generated along with the parameter and value files. This file allows users to reload the stochastic parameters and their values used in the previous simulation. Re-populating the runs or selecting other distribution/settings will change these values. However, these changes will not be saved until the GSSHA project is saved.

To run GSSHA in the batch (stochastic) mode, a "-b", appended by the total number simulations, must be added to the command line used to execute the stochastic version of GSSHA. This is implemented in GSSHA's interface in WMS.

The FORTRAN code of the generic version of CE-QUAL-W2 is used and modified programmatically to generate a model executable that could take arguments, specifically the full path of the control file name. The newly generated executable is used in WMS to run CE-QUAL-W2 from within WMS. This is meant to facilitate running the model especially in the batch run mode where modelers had to manually run the stand-alone version of CE-QUAL-W2 multiple times. Two CE-QUAL-W2 executables are created for the two versions of 3.2 and 3.5.

According to the specific case of the GSSHA, CE-QUAL-W2 and the methodology (3.1.4), WMS breaks down flux files so that each segment flux file has a start and end Julian dates that matches those specified in the main model control file.

4.2.2 Batch Runs

The batch runs facilitate the modeling process and reduce the time taken to run all simulations. In addition, it helps in automating the process which is, in most cases, considered a major advantage in a repeated model run environment. To automate the linkage, the following are achieved:

- CE-QUAL-W2 can be run directly within WMS.
- CE-QUAL-W2 executable and graph.npt are both saved in the same folder.
- W2_con.npt may be located in a separate path.

4.2.3 Modeling Guidelines

In the event that back water exists (i.e., water that is flowing outside the water body) or in other words water is flowing from CE-QUAL-W2 water bodies to GSSHA cells, modelers are warned to re-set their boundary conditions and/or refine their GSSHA model. The linkage must be redefined because, currently, this negative flow is ignored and not factored into the model linkage. This research handles negative fluxes this way because; a negative flux could mean one of two things:

1. An issue in the real world that needs to be looked at closely or
2. A problem with the mathematical representation of the real world.

Fixing and/or refining the model and its boundary conditions should, in most cases, eliminate the negative fluxes. If it does not and a negative flux is actually predicted repeatedly, then some other flooding-specific model should be used for this particular area.

4.2.4 Modeling Limitations

Like any modeling process, there are some assumptions and limitations that are associated with this approach. These assumptions can be listed as follows:

- There is only one static simulation technique assumed; i.e. Monte Carlo Simulation.
- The approach has limited number of stochastic variables in GSSHA. These variables are set to follow certain distributions (with a default of Beta distribution).
- The approach works well with equal sized grid cells. However, it can be expanded to work with irregular (smart) grids if needed in the future.
- Whenever a test of containment is needed, the cell centroid, as opposed to the entire cell, is used to test if the cell lies within or outside a given polygon (in this case, CE-QUAL-W2 segment polygon).
- The aggregation of cells is done with the help of extending the segment boundary to intersect the cell.
- The importance of a variable is used as opposed to sensitivity of the model to it.
- It is assumed that all the inflow from GSSHA cells is contributing to the top layer of each segment in the CE-QUAL-W2 model. This is assumed

indicating that the major process in the transition of water from land to water portions of the watershed is done through overland flow (surface water) as opposed to sub-surface/ground interaction. None of the overland flow is assumed to be distributed through the vertical profile of a segment.

- The linkage is designed to be a quasi 2-way interaction between the two models. A input; i.e. segment ID index map, is required by GSSHA when at a later stage, an output from GSSHA is used as input to CE-QUAL-W2.
- The stochastic input is attenuated from GSSHA to CE-QUAL-W2. Water surface elevations might show it a little better.

4.2.5 Compatibility and Expandability

Most models undergo continuous development and updates. It is expected that GSSHA will have a comprehensive nutrient module. Likewise, CE-QUAL-W2 is expected to have further modifications in the future. Some of the planned enhancements (Cole, et al., 2007) for CE-QUAL-W2 include sediment diagenesis and a three dimensional version among other enhancements. For the linkage to perform the way it is currently intended for, there are some steps to account for:

1. CE-QUAL-W2 Executable: As updates occur and new versions are available for use, new CE-QUAL-W2 model executables and the corresponding graph input file must be generated and distributed to WMS users. The executable is generated by modifying the code distributed with the generic version of CE-QUAL-W2. The modification should enable command line argument; i.e. the full path of the control file name. Currently, the generic version of CE-QUAL-

W2 is programmed using FORTRAN-90. This way, WMS can call this executable and run it multiple times.

2. CE-QUAL-W2 Control File: WMS creates a control file that is compatible with CE-QUAL-W2 v.3.2. Differences between v. 3.2 and v.3.5 (Cole, et al., 2007) are minimal and include few lines of text. It must be noted that v. 3.5-compatible control files are still readable by WMS. Currently, WMS 8.1 is distributed with the two CE-QUAL-W2 versions; i.e. 3.2 and 3.5, along with the appropriate graph input file. For example, if a newer version of CE-QUAL-W2 is released, modelers would have the choice of modeling their watershed with v. 3.5, or have WMS generate the 3.5-compatible control file and edit it manually. Alternatively, a new CE-QUAL-W2 executable should be generated and distributed to users. Furthermore, modifications in WMS to generate a control file that is compatible with the new version usually can be obtained from CE-QUAL-W2 manual; i.e. differences between versions (Cole, et al., 2007).
3. GSSHA: A GSSHA executable capable of running in the batch mode, as indicated earlier, must be available. This version must be able to utilize a segment ID index map, stream inflow file and both value and parameter files to generate the flux files for the number of runs corresponding to the value file.

Any parameter that is modeled in CE-QUAL-W2 can be viewed using the credible datasets once the nutrient module in GSSHA is fully functional. Modelers must select the parameter that is under investigation to be modeled by CE-QUAL-W2 and enable constituent output time series. When reading the CE-QUAL-W2 solution in WMS, modelers must select the desired parameter, from the list of parameters, to be read. If a parameter is not modeled by CE-QUAL-W2, or if the modeler does not select the right output file, WMS would generate a warning message that the desired parameter is not read. However, by default, WMS will generate a number of temperature datasets equivalent to the number of valid output files in valid run folders; Run_001, Run_002, …etc.

Parameters in CE-QUAL-W2 could also be handled stochastically using a similar methodology done with GSSHA stochastics. The stochastic dialog would have a list of the key parameters and modelers would input a negative value in this dialog that is used where the parameter value is needed in CE-QUAL-W2. In the current setup, modelers must use the same number of runs in CE-QUAL-W2 as for GSSHA.

4.2.6 Similar Model Linkage

The linkage process outlined above can be utilized between other models. To do so, the general guidelines used in this research can be summarized as follows:

A. Dimensionality: it is not necessary that the linked models have the same dimension. For instance, a 1-dimensional model could be linked to a 2-dimensional model. However, the linkage must be designed in such a way that input/output exchange is made to preserve information and does not result in major loss of accuracy of model representation/input data. Assumptions made must not degrade the available input data quality.

B. Spatial domain: in a similar fashion, it is not necessary for the linked models to have the same dimension on a spatial domain. As an example, this research links

GSSHA, a horizontal 2-dimensinoal model to CE-QUAL-W2, a vertical 2-dimensional model (1-dimensional on a horizontal scale).

C. Temporal domain: Two factors must be addressed as far as the temporal domain is concerned.

- Time step: the linked models do not need to have the same time step so long as interpolated results are acceptable. However, in interactive bi-directional linkage, using the same time step may be the best way for the linkage.
- Total duration: in a bi-directional model linkage, the total duration for linked models, generally, need to be the same. Otherwise, and this research as an example, run durations in the linked models do not need to be the same. Moreover, the "feeding" model run duration can either be longer or shorter that the run duration of the "fed" model. However, it is recommended that the run duration of the feeding model, if applicable, be shorter, or at most equal to, the run duration of the "fed" model. That way, the "fed" model captures all the input from its counterpart.

D. Statistical domain: this is an optional feature in the linkage process. Stochastic approach provide a means for addressing uncertainty in the modeling process and in general, helps overcome the inherent variability in the values of model input parameters and in the modeling process itself.

E. Bi-directional linkage: in this research, the two models do not require dynamic linkage where input/output are exchanged back and forth between them. However, in other cases, this might be the best way to link models.

With these guidelines in mind, and as applicable, other models may be linked in an automated fashion to build integrated temporal and spatial modeling frameworks. Considerable research effort is yet to be performed to make this linkage process not model-specific. However, some parts of the linkage process may need minimal efforts to customize it. As an example, the random generation of model input parameters following a pre-set statistical distribution only need the determination of which parameter for the linked models that may be modeled stochastically. In this case, the current research and the developed tools may be utilized for other models.

4.3 Case Studies Results

This section will highlight the application of the developed framework on two case studies; Lake Zapotlan, Mexico and Eau Galle Reservoir, Wisconsin. Each of the case studies represents a unique set of conditions and was used to test the tools and functionalities developed in this research. It is beyond the scope of this research to calibrate these models. This research focused on testing the reliability of both models from a stochastic linkage perspective. This process is not intended to replace calibration but rather, provide powerful prediction to one that already is.

As indicated in the previous chapter, the framework developed in this research requires three hypothetical layers:

1. Drainage layer which involves WMS drainage coverage with a delineated watershed along with elevation and other supporting data.
2. GSSHA layer which has a GSSHA grid for the delineated watershed along with a running GSSHA model for the whole watershed.

3. CE-QUAL-W2 layer which has WMS branch and segment coverages and a running CE-QUAL-W2 model for the water body(ies), of interest, in the watershed.

4.3.1 Lake Zapotlan

Lake Zapotlan closed basin is located in the southern part of Jalisco State, Mexico bounded by 19° 34' & 19° 53' North and 103° 24' & 103° 38' West. It is a relatively shallow endorheic lake with average depth of about 4-6 meters (Jimenez, et al., 2006) (Figure 4-4). A GSSHA model was developed for the lake and its watershed (Gautirrez, 2007). Due to lack of data, the model only examined two processes, i.e., overland flow and infiltration (using Green and Ampt). The grid cell size used was 267 meter (Figure 4-5). The total number of rows was 127 and number of columns was 90. The duration of the simulation time was 500 minutes with a computational time step of 10 seconds.

Figure 4-4: Lake Zapotlan, Mexico.

Because the GSSHA model was planned to run multiple times, it was important to determine what grid cell size was optimum for the run time (Figure 4-6). The GSSHA model was optimized by varying the grid cell size, computational time step, type of processes involved, and some of the output cards, some of which are "printing" cards that take running time to print results to grids (Downer, et al., 2006).

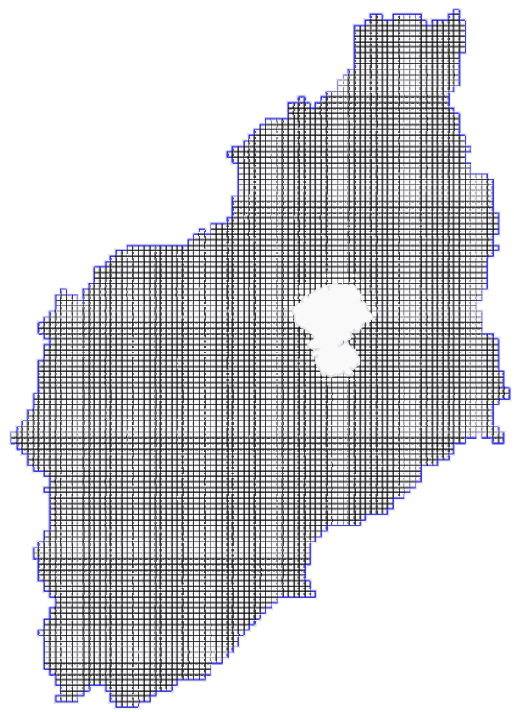


Figure 4-5: Lake Zapotlan Watershed GSSHA Model.

For optimization purposes, the GSSHA model was modified and run for seven additional grid cell sizes; i.e. 350, 400, 450, 500, 1000, 2000 and 4000 meter. Each of these was run multiple times to investigate the effect of the time step on the total successful run time GSSHA takes. Results of these runs are illustrated in Figure 4-6. Since it is beyond the objectives of this research to calibrate/validate a GSSHA model, this section focused on finding a "good-enough" grid cell size to test the linkage between

GSSHA and CE-QUAL-W2. The GSSHA model was finally set to have a grid cell size of 500 meter. This model was taken as the main land model and all lake cells (i.e. cells overlaid by the lake) are actually ignored in the GSSHA model.

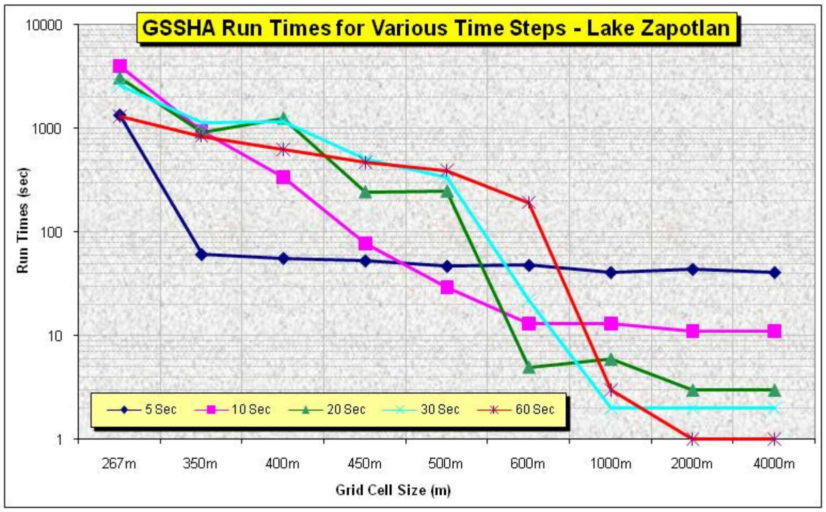


Figure 4-6: Run Times for the Lake Zapotlan GSSHA Model.

It was found that the running time of the 500m model is acceptably appropriate for the linking purposes. It must be noted, however, that a uniform cell size of 500 m is not optimum at least for this watershed and at least as far as the GSSHA model output is concerned. Modelers should always try to maintain a balance between the accuracy desired and available computing power. With this size of a watershed and with the amounts of stochastic runs that are required, modelers may need to compromise the cells size to get a functional stochastic model.

As we can see in Figure 4-6, the general trend is a decrease in GSSHA model run time with increase in the model grid cell size. Also, in general the higher the computation time step, the lower the run time for the same total run period.

A water quality, hydrodynamic model, with one branch and six segments, was set up for Lake Zapotlan using CE-QUAL-W2 (Paz, 2007) (Figure 4-7).

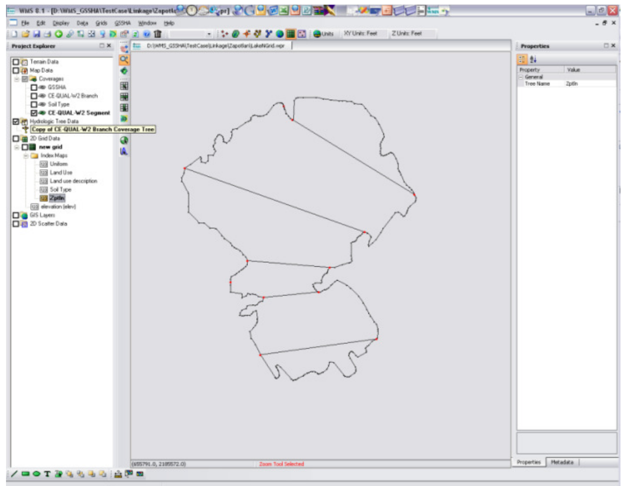


Figure 4-7: Lake Zapotlan CE-QUAL-W2 Segments.

The CE-QUAL-W2 model (Figure 4-7) was used to generate the segment ID index map (Figure 4-8). All the cells that overlay the lake are assigned a number greater than zero and all non-lake cells are assigned a zero. A stochastic version of GSSHA was run 50 times in the batch mode and accordingly 50 flux files (Figure 4-9) were generated.

The total run durations and the time steps were different in the GSSHA model than in the CE-QUAL-W2 model for lake Zapotlan, with the total duration in the CE-QUAL-W2 model longer than the total duration in the GSSHA model. This is generally acceptable, in terms of the linkage, as opposed to the opposite, total duration of GSSHA longer than it is for CE-QUAL-W2, because if the duration of the GSSHA model is longer than the duration of the CE-QUAL-W2 model, this would result in a loss of available modeling data. It is important to note, however, that the flux files generated by GSSHA need to be extended to match the total run duration by CE-QUAL-W2 with zero input. If the flux files generated by GSSHA are left alone, CE-QUAL-W2 model will not run because of "incomplete" input files.

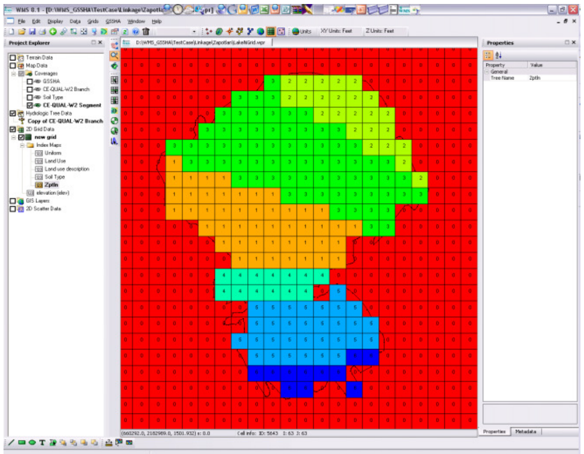


Figure 4-8: Lake Zapotlan CE-QUAL-W2 Index Map.

As confirmed by the sample flux file Figure 4-9, the CE-QUAL-W2 model of Lake Zapotlan has six segments (2 through 7) thus six tributary inflow (QTR) files (Figure 4-10) were generated out of each flux file, one for each segment. WMS created 50 sub-folders, named "Run_0001", "Run_0002" and so forth. Each of these sub-folders contained the copied and modified control file along with the six segment inflow files. Other CE-QUAL-W2 input files that are not affected by the stochastic modeling or the linkage remain in the parent folder. CE-QUAL-W2 was run stochastically 50 runs for each of the sub-folders. The resulting output was read into temperature, as an example, datasets (Figure 4-11).

A WMS gage (Figure 4-12) was placed in one of the longitudinal cells to be able to plot a time series displaying the temperature variations among the various time steps of the model (Figure 4-13). This gage can generally represent a monitoring site in a reservoir or even representing a withdrawal point in a selective withdrawal tower.

0006_w2_seg_qin - WordPad

File Edit View Insert Format Help

jday	2	3	4	5	6	7	8
1.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.004	0.00	0.02	0.08	0.02	0.07	0.01	0.00
1.008	0.54	0.18	0.84	0.19	0.65	0.19	0.04
1.012	0.34	0.44	1.73	0.25	1.62	0.50	0.16
1.017	0.89	0.81	2.36	0.34	2.63	0.97	0.45
1.021	1.65	1.27	2.96	0.68	3.83	1.33	0.84
1.025	2.48	1.70	3.70	0.82	4.26	1.75	1.22
1.029	3.39	2.37	4.64	1.15	4.93	2.28	1.78
1.033	4.62	2.95	5.63	1.32	5.71	2.76	2.55
1.037	5.24	3.58	6.69	1.54	6.56	3.33	2.67
1.042	6.13	4.18	7.70	1.60	7.11	3.98	3.31
1.046	5.35	3.27	4.05	1.09	7.29	2.86	2.20
1.050	3.41	2.12	2.91	0.48	6.38	1.74	1.27
1.054	2.25	1.32	2.12	0.12	4.34	1.19	0.54
1.058	1.56	0.84	1.49	0.08	2.84	0.62	0.14
1.063	1.15	0.40	0.85	0.06	1.89	0.24	0.03
1.067	0.85	2.58	0.40	0.01	1.28	0.11	0.00
1.071	0.55	1.35	0.17	0.00	0.78	0.06	0.00
1.075	0.49	0.75	0.02	0.00	0.48	0.04	0.00
1.079	0.34	0.42	0.01	0.00	0.26	0.04	0.00
1.083	0.27	0.21	0.00	0.00	0.15	0.04	0.00
1.088	0.23	0.06	0.00	0.00	0.07	0.00	0.00
1.092	0.14	0.00	0.00	0.00	0.04	0.00	0.00
1.096	0.11	0.00	0.00	0.00	0.00	0.00	0.00
1.100	0.06	0.00	0.00	0.00	0.00	0.00	0.00
1.104	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.108	0.04	0.00	0.00	0.00	0.00	0.00	0.00
1.113	0.01	0.00	0.00	0.00	0.00	0.00	0.00
1.117	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.121	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.125	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.129	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.133	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.137	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.142	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.146	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.150	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.154	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.158	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.162	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.167	0.00	0.00	0.00	0.00	0.00	0.00	0.00

For Help, press F1

Figure 4-9: Sample Lake Zapotlan Flux File.

w2_seg2 - WordPad

File Edit View Insert Format Help

jday	2
1.000	0.00
1.004	0.00
1.008	0.54
1.012	0.34
1.017	0.89
1.021	1.65
1.025	2.48
1.029	3.39
1.033	4.62
1.037	5.24
1.042	6.13
1.046	5.35
1.050	3.41
1.054	2.25
1.058	1.56
1.063	1.15
1.067	0.85
1.071	0.55
1.075	0.49
1.079	0.34
1.083	0.27
1.088	0.23
1.092	0.14
1.096	0.11
1.100	0.06
1.104	0.00
1.108	0.04
1.113	0.01
1.117	0.00
1.121	0.00
1.125	0.00
1.129	0.00
1.133	0.00
1.137	0.00
1.142	0.00
1.146	0.00
1.150	0.00
1.154	0.00
1.158	0.00
1.162	0.00
1.167	0.00

For Help, press F1

Figure 4-10: Sample Lake Zapotlan Flux File Broken Down to Segment 2.

Figure 4-13 shows 3 lines representing the mean, lower and upper bounds of a 95% credible interval of the temperature time series at the specified location. If desired, modelers can still utilize the “one line” time series by ignoring the upper and lower bounds of the credible interval. However, the credible interval might aid water resources managers in the decision making process.

As indicated earlier, the higher the credibility, the wider the range between the lower and upper bounds. This can be viewed as follows:

- Credibility level of 50% (Figure 4-14).
- User-defined credibility level of 77% (Figure 4-15)
- Credibility level of 99% (Figure 4-16).

This case study shows the implementation of the developed tools on a closed basin. The second test case, Eau Galle reservoir is different as seen in the following section. Following are some of the differences between the two test cases; Lake Zapotlan and Eau Galle Reservoir:

- Lake Zapotlan watershed is a closed basin while the Eau Galle watershed is a sub-watershed of the Chippewa River; i.e. a tributary of the Mississippi river.
- In Lake Zapotlan model, there is no stream input where as in Eau Galle Reservoir; input from Eau Galle River is accounted for.
- In Lake Zapotlan model, there is no infiltration process taking place, whereas Eau Galle model accounted for the infiltration process.
- Eau Galle watershed has cascaded reservoirs.

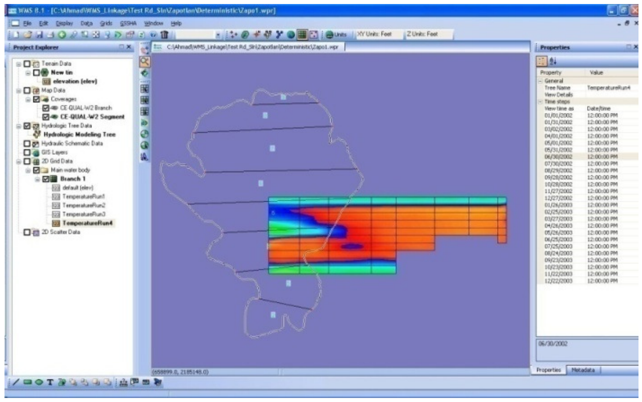


Figure 4-11: Temperature Grids for Lake Zapotlan.

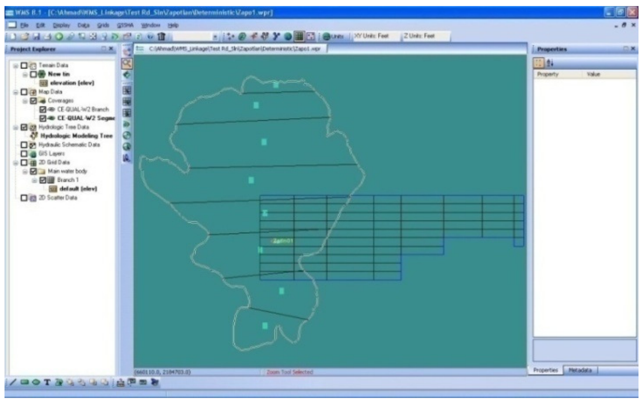


Figure 4-12: Location of Monitoring Point (WMS Gage).

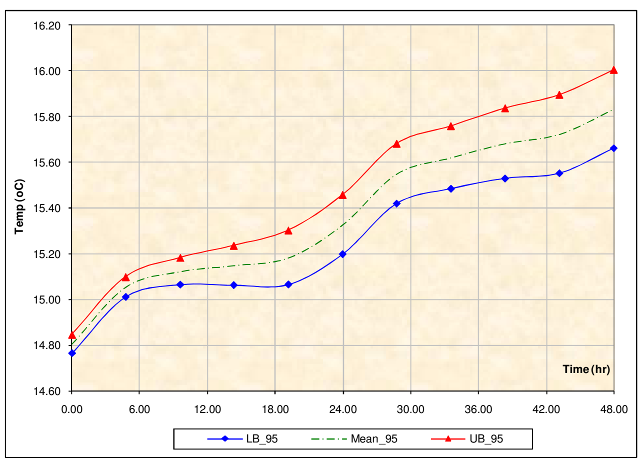


Figure 4-13: Stochastic Time Series Plot at the Monitoring Point – 95% CI.

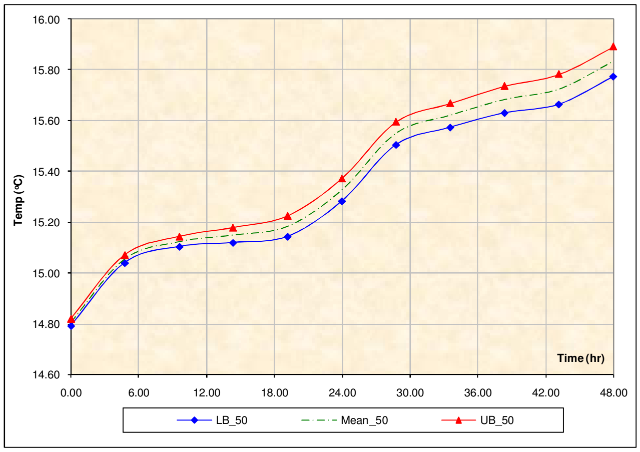


Figure 4-14: Stochastic Time Series Plot at the Monitoring Point – 50% CI.

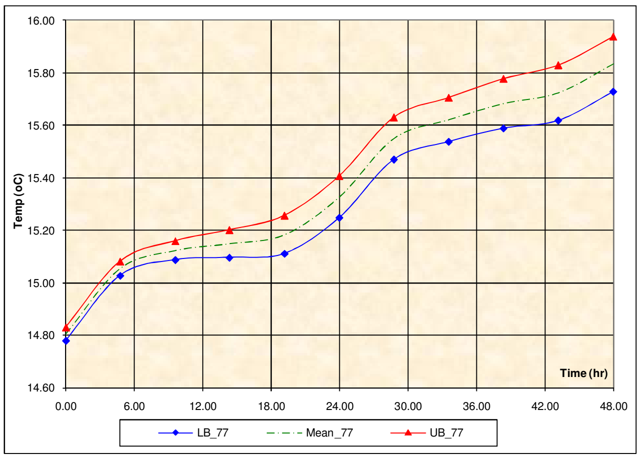


Figure 4-15: Stochastic Time Series Plot at the Monitoring Point – 77% CI.

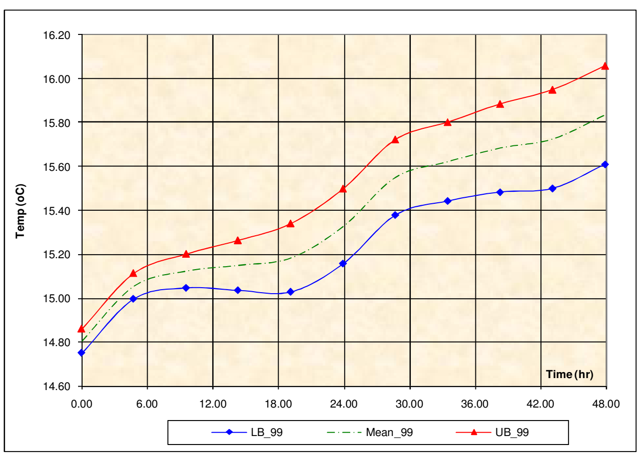


Figure 4-16: Stochastic Time Series Plot at the Monitoring Point – 99% CI.

4.3.2 Eau Galle Reservoir

Eau Galle Reservoir (Figure 4-17) is a 150-acre impoundment located just north of Spring Valley, Wisconsin and 50 miles east of the Twin Cities and 40 miles west of Eau Claire, Wisconsin (US-ACE, 2007). The Eau Galle River is a tributary of the Chippewa River in western Wisconsin in the United States. It is about 35 miles long. Via the Chippewa River, it is part of the Mississippi River watershed. The reservoir is formed by a US-ACE dam located near Spring Valley, Wisconsin. Lake Eau Galle is hyper-eutrophic, shallow lake with very poor water quality and poor water clarity. As a lake, Eau Galle is high in nutrients and can support a large biomass (Donkel, 2002).

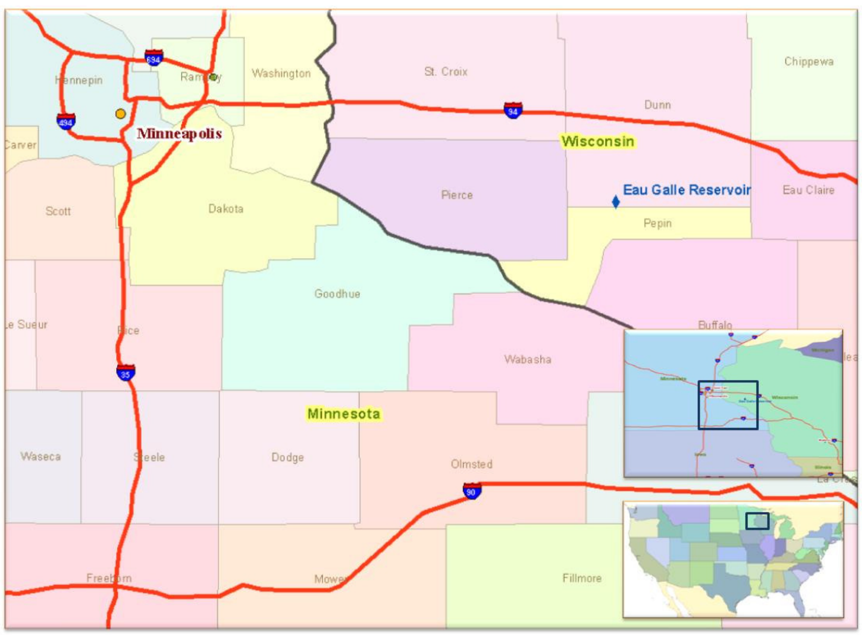


Figure 4-17: Eau Galle Reservoir Location.

A GSSHA model was setup for Eau Galle Reservoir Watershed (Figure 4-18). A grid cell size of 100 m with a total number of rows of 468 and number of columns of 305 was used in this model. Total duration was 1000 minutes with a computational time step of 5 seconds. The Green and Ampt method was used for infiltration. Also, diffusive wave was used as channel routing for Eau Galle River. No evapotranspiration method was simulated in this model.

A CE-QUAL-W2 model was setup for Eau Galle Reservoir (Figure 4-19). One branch and seven segments were used to simulate the reservoir. Input to the reservoir in this model comes from the grid cells in addition to the inflow coming through the river. Similar to Lake Zapotlan, the CE-QUAL-W2 model for Eau Galle Reservoir was used to generate the segment ID index map. GSSHA was run stochastically for 50 runs.

To investigate the various aspects of the stochastically linked GSSHA to CE-QUAL-W2, the following runs were performed:

- Deterministic run: This represented a single GSSHA run linked to a single CE-QUAL-W2 run. The GSSHA model start date was identical to CE-QUAL-W2 model start date. However, the duration of the GSSHA model was considerably less the duration of the CE-QUAL-W2 model (Table 4–3). This is considered the base deterministic model (Figure 4-22).
- Base stochastic run: This represents the 50 stochastic GSSHA runs that were de-aggregated to re-build 50 CE-QUAL-W2 runs. The output from these runs shows variations (Figure 4-23). Similar to the base deterministic model, the GSSHA model start Julian date coincides with the CE-QUAL-W2 model start Julian date on all 50 runs. This is considered the base stochastic model.

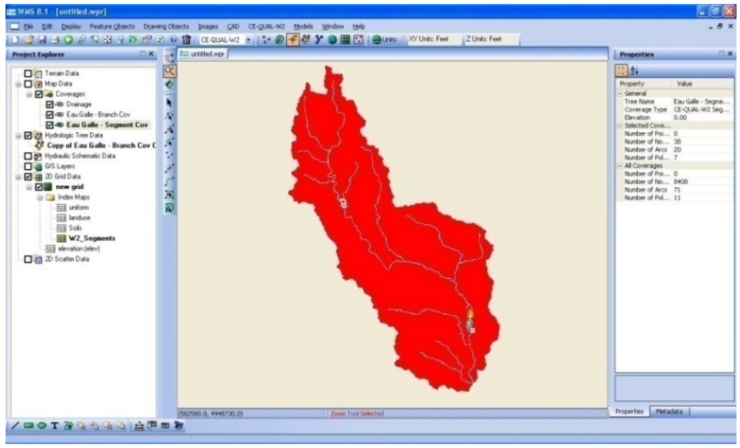


Figure 4-18: Eau Galle Reservoir Watershed.

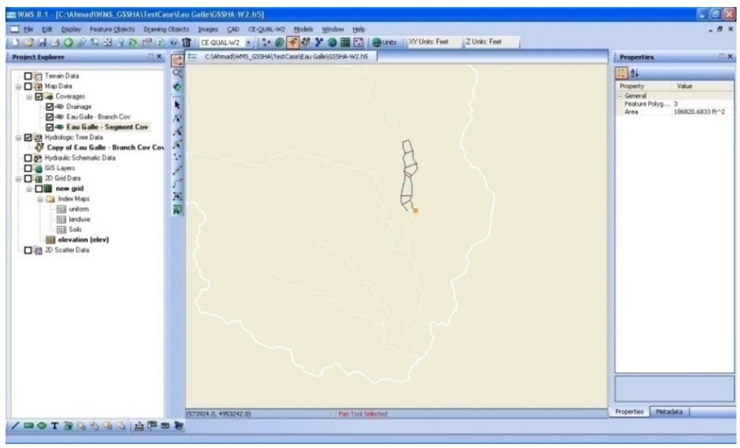


Figure 4-19: Eau Galle Reservoir CE-QUAL-W2 Model.

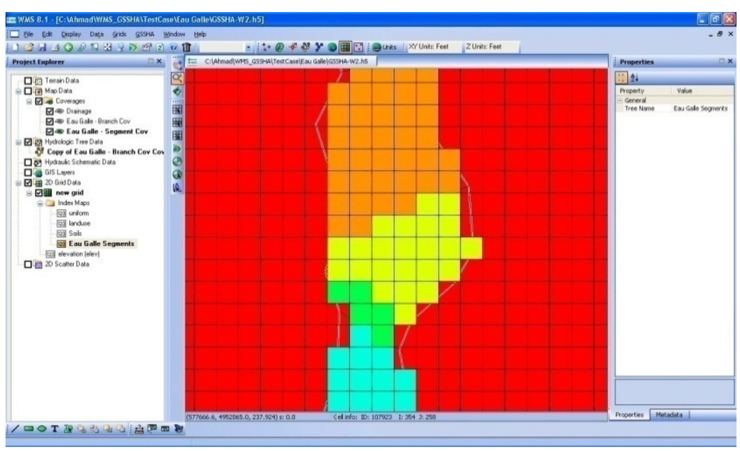


Figure 4-20: Segment ID Index Map for Eau Galle Reservoir.

Table 4–3: Models Temporal Linkage.

Model	Start Julian day		Model Duration (days)	
	GSSHA	CE-QUAL-W2	GSSHA	CE-QUAL-W2
Base Deterministic	1	1	1	400
Base Stochastic	1	1	1	731
Delayed Input	381	1	1	731
Stochastic Temperature	1	1	1	731

- Delayed stochastic run: This is identical to the base stochastic model with GSSHA input delayed to Julian date 381 (Table 4–3). In this model, the first 380 days have essentially the same input and therefore, the credible interval width is zero up until the stochastic GSSHA input comes into effect (Figure 4-24).
- Stochastic temperature input run: This is identical to the base stochastic model with the tributary temperature being modeled stochastically. The tributary input temperatures are set to vary following a Normal distribution PDF (Figure 4-25).

For all of these runs, a time series plot of the temperatures was developed for eight locations along the profile of Eau Galle Reservoir (Figure 4-21). Figure 4-22, Figure 4-23, Figure 4-24, Figure 4-25 show time series plots for three gages out of the eight within the profile. The top time series is for the gage in segment 3 and layer 2. The middle time series is for the gage in segment 7 and layer 7. The lower time series is for segment 8 and layer 9.

The location of gages is selected to capture a comprehensive and comparative picture of output across the surface and depth of the profile. As we can see from Figure 4-21, gage S7L7 is located in segment 7 and layer 7. It can be seen that segment 8; i.e.

the most downstream segment in the reservoir, has 3 gages. Two gages are located at the top surface in segments 3 and 8. Three gages are located at the lower layer of the reservoir in segments 4, 6 and 8 while three other gages are located in the middle of the profile in segments 4, 7 and 8.

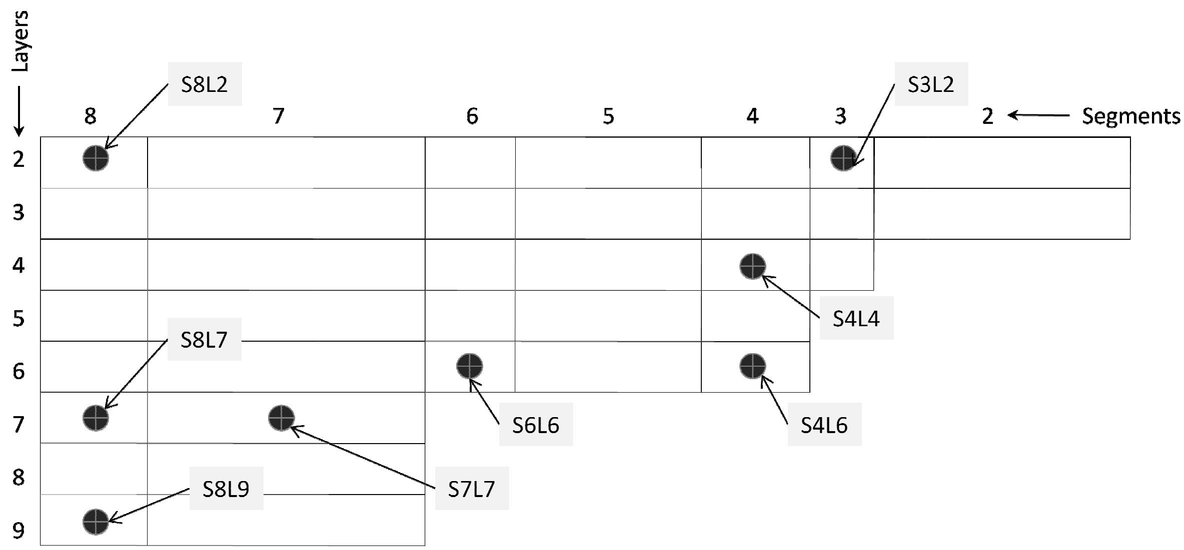


Figure 4-21: Eau Galle Reservoir Grid and Gage Locations

The temperature time series were analyzed to determine the width of the credible intervals for the last three runs outlined above; i.e. base stochastic, delayed and stochastic temperature. Figure 4-26 shows the credible intervals width for the gage located in segment 6 and layer 6 (S6L6) for the base stochastic run. Similarly, Figure 4-27 and Figure 4-28 show credible interval width for the same gage for the delayed runs and the stochastic temperature runs, respectively.

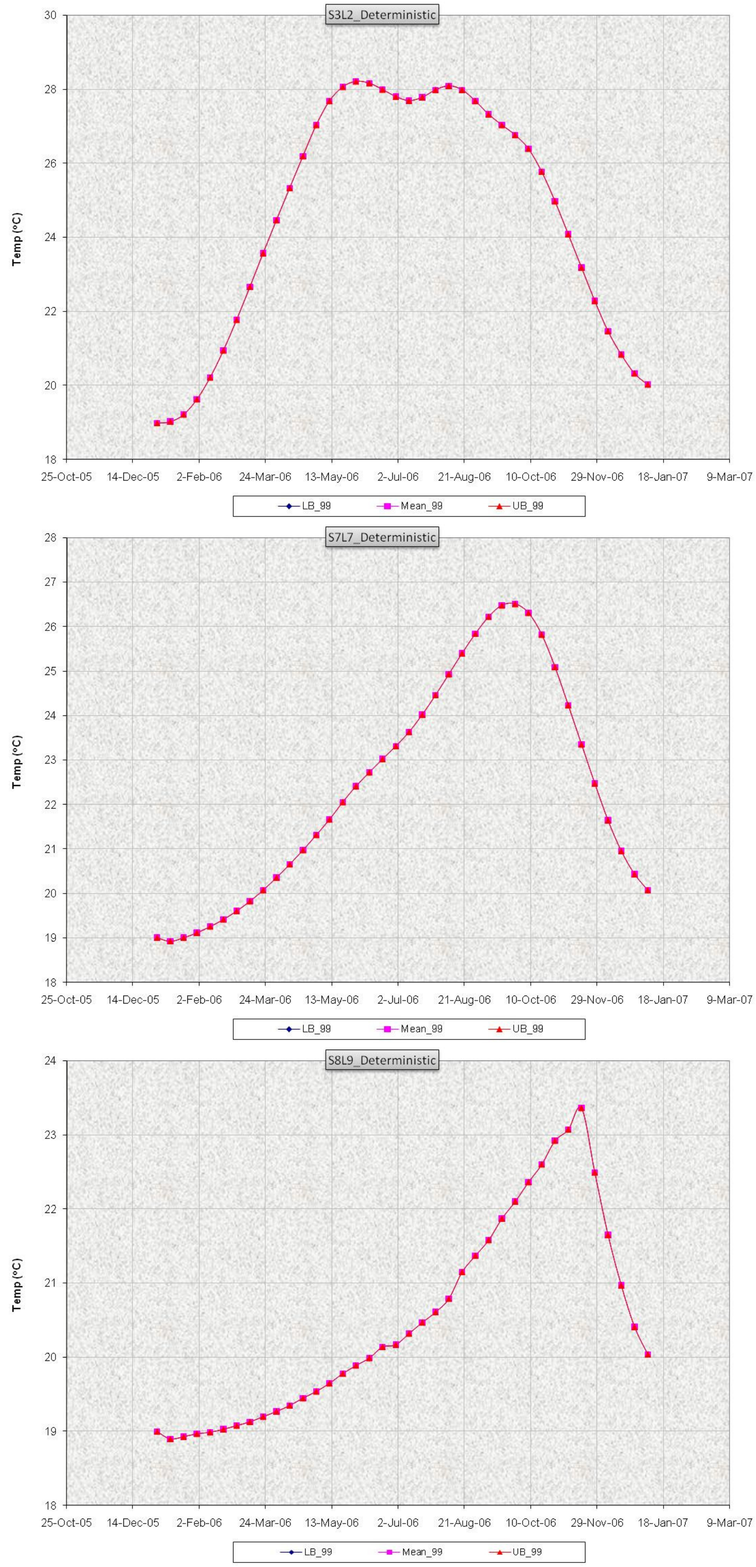


Figure 4-22: Eau Galle Reservoir Temperature Time Series – Deterministic.

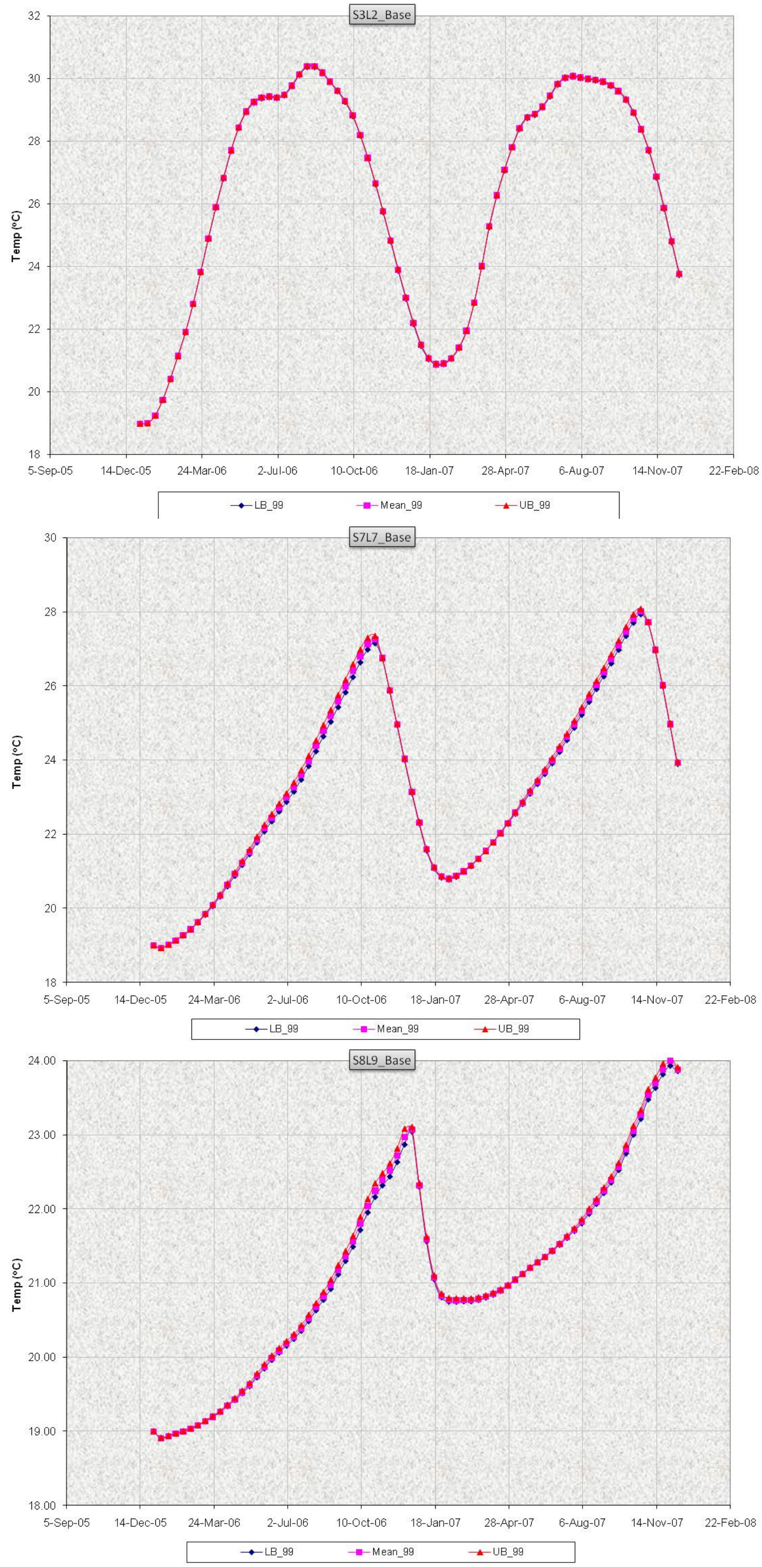


Figure 4-23: Eau Galle Reservoir Temperature Time Series – Base Stochastic.

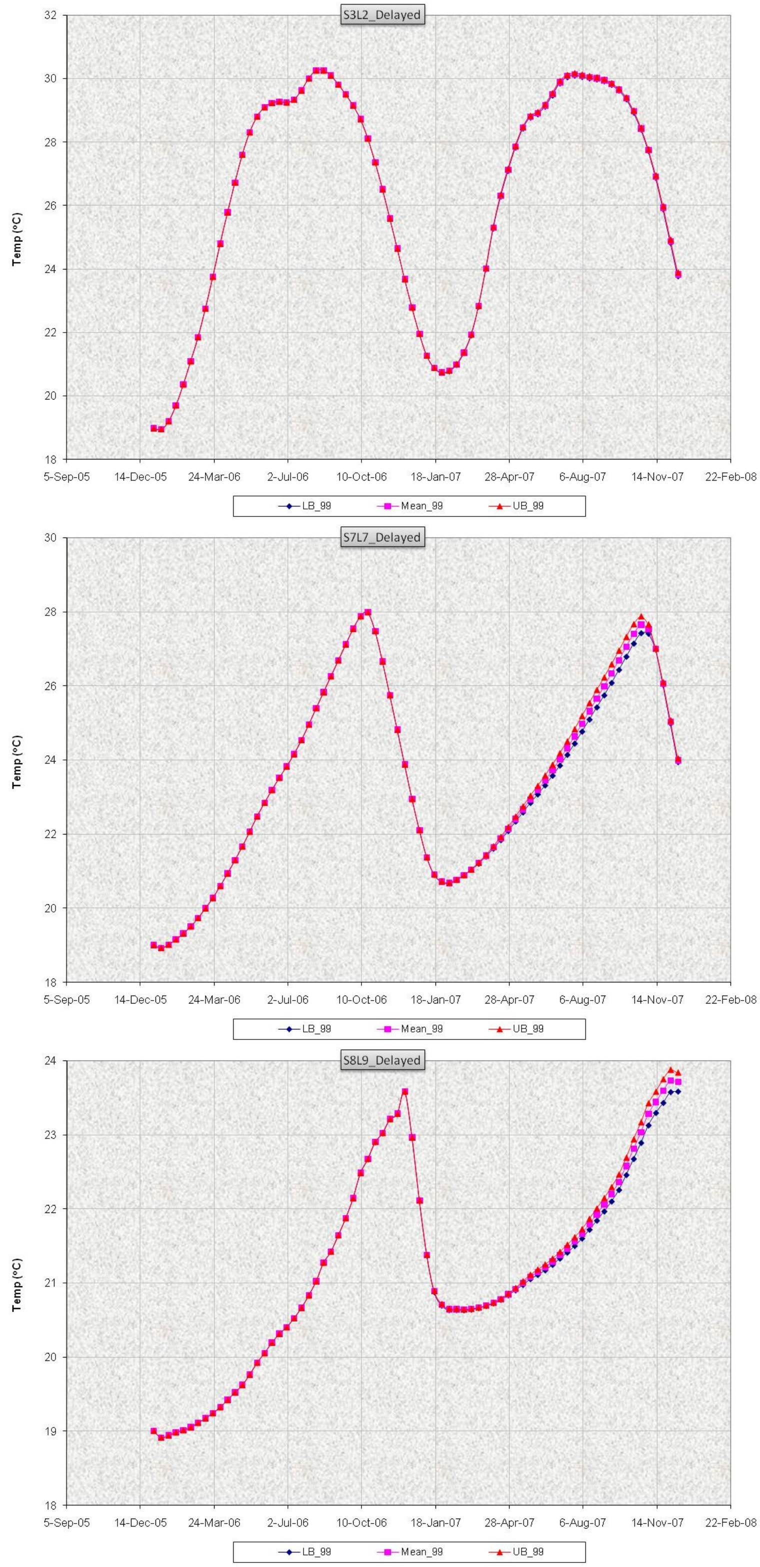


Figure 4-24: Eau Galle Reservoir Temperature Time Series – Delayed.

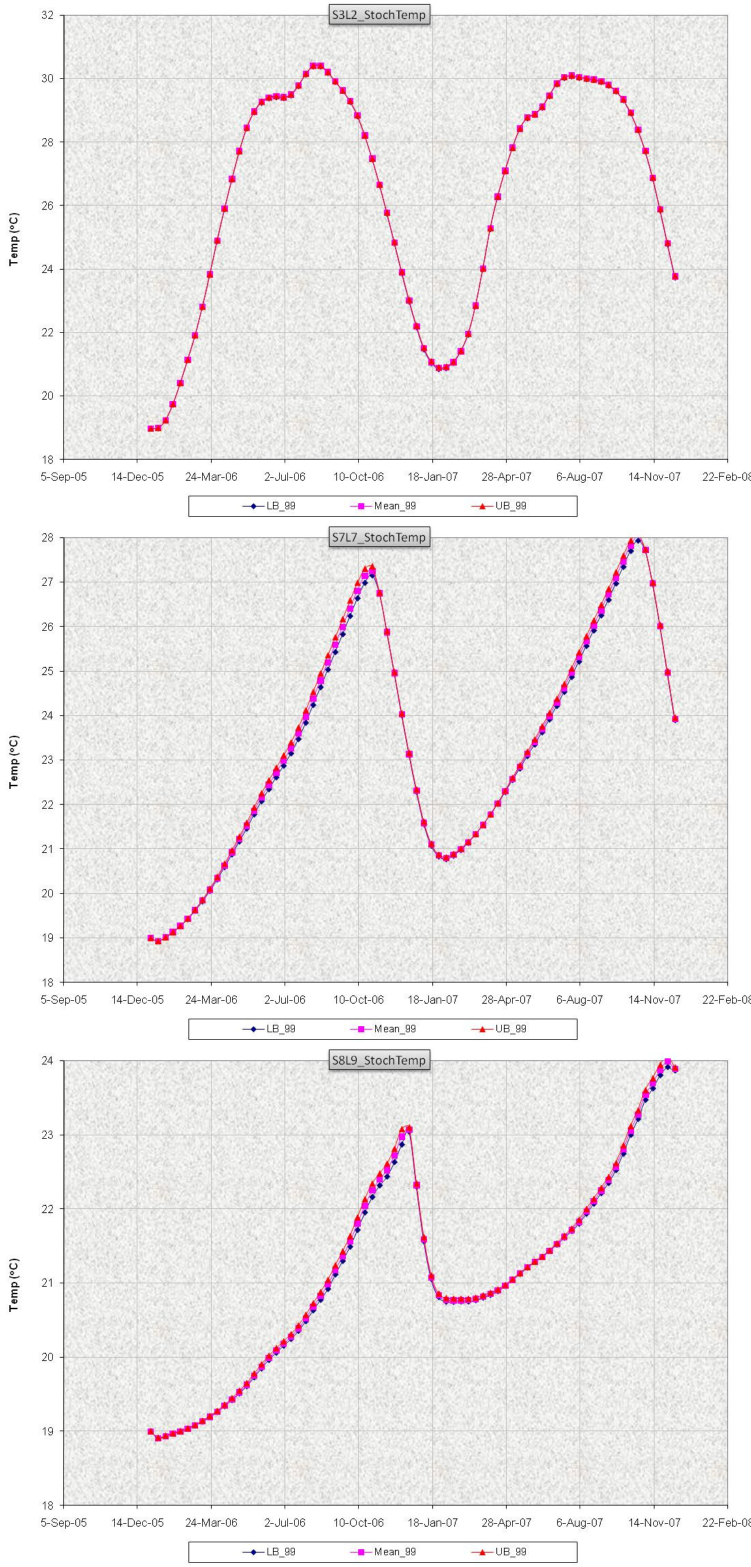


Figure 4-25: Eau Galle Reservoir Temperature Time Series – Stochastic Temp.

Figure 4-22, Figure 4-23, Figure 4-24 and Figure 4-25 indicate that surface gages experienced minimum temperature fluctuations and that gages closer to the reservoir bottom showed small fluctuations whereas the mid-depth gages showed maximum fluctuations. Surface layers in the water body generally experience direct atmospheric effects that play a large role in determining water surface temperature. Also, bathymetric layers have direct contact with the reservoir soil beds making them more susceptible to temperature changes attenuation.

Temperature time series plots for the S8L9 gage seem to be the "flatter" plot of all eight gages. This suggests that the deeper the reservoir gets, the less variation there is in water temperature across seasons. It can be seen from the deterministic run (Figure 4-22) that the model duration is half the model duration of the three other stochastic runs. The deterministic run is done as a reference.

As seen in Figure 4-24, the delayed run values for various credible intervals are identical up until GSSHA input time step where we start to see some variations, and accordingly, interval width. Relatively speaking, these widths are seen 380 days earlier in the base stochastic run (Figure 4-23). In comparing the base stochastic run (Figure 4-23) to the stochastic temperature run (Figure 4-25), there is no major difference noticed.

Figure 4-26 shows that the effect of the stochastic input on Julian Day 1 (1/1/2006), is propagated in the entire duration of the model. It is noticed that the credible interval width is relatively high in summer months. It is also noticed that the credible interval widths approaches zero around March and November. As indicated earlier, a 99% credible interval is expected to be wider than a 95% interval and a 95% interval is expected to be wider than a 70% interval. This is evident in Figure 4-26.

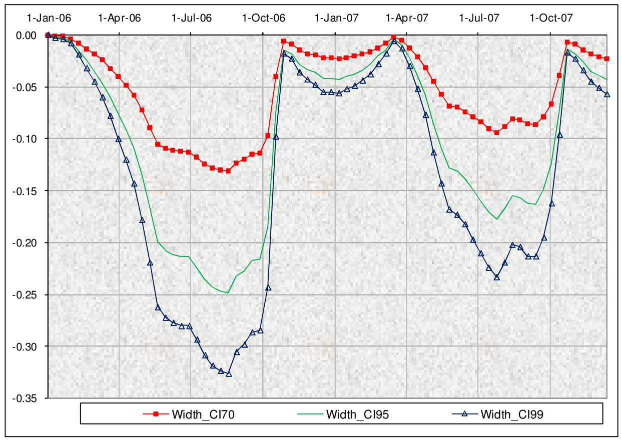


Figure 4-26: Credible Interval Widths for S6L6 – Base Stochastic.

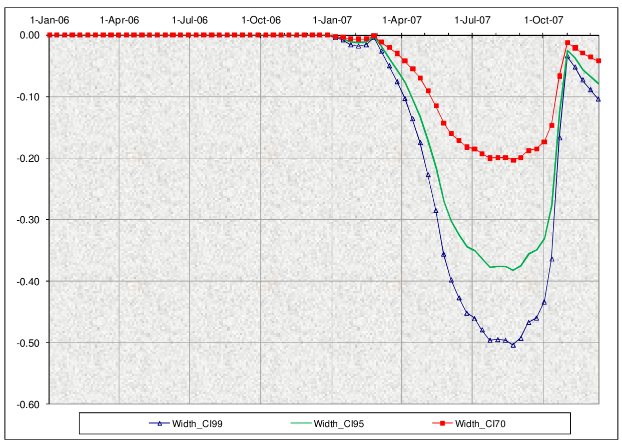


Figure 4-27: Credible Interval Widths for S6L6 – Delayed Stochastic.

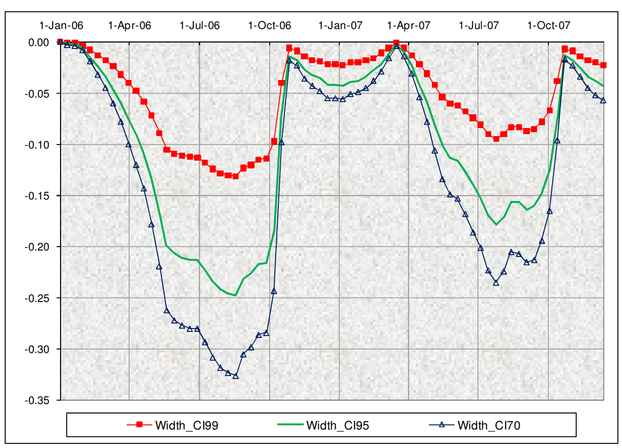


Figure 4-28: Credible Interval Widths for S6L6 – Stochastic Temperature.

Figure 4-27 indicates that prior to the stochastic input on Julian Day 381; i.e. 15th January 2007, widths of the three credible intervals is zero which basically means that the values are identical for all the time steps up until Julian Day 381. Minor changes in the interval widths are seen until around beginning of February 2007 before the widths start to approach zero again around beginning of March. Interval widths picks up again for the summer months before the winter decline around November. Same base patterns noticed in Figure 4-26 are still noticed in Figure 4-27.

Figure 4-28 is almost identical to Figure 4-26 and it follows the general pattern with minor local changes. Three paired t-tests, with null hypothesis of zero difference, were done to statistically examine if both runs are identical (Table 4-4). A paired t-test is used to compare each time step, on both runs, individually testing for a zero difference. Additionally, three paired regression analyses are done for each of the three credible intervals. The base stochastic 70% credible interval width was used as an explanatory variable where the stochastic temperature run 70% interval width was used as a response variable in the regression analysis. Similarly, the 95% and 99% credible interval widths were analyzed.

As seen in Table 4-4, the paired t-test results indicate p-Values of 0.03, 0.08 and 0.06 for the 70%, 955 and 99% credible intervals respectively. This suggests that there is no evidence to support rejecting the null hypothesis (base stochastic = stochastic temperature) on all three intervals. Moreover, the regression analysis indicated that there is strong evidence (p-Value < 0.001) that the base stochastic run can explain the stochastic temperature and that both are close to identical (regression coefficient $\cong 1.0$)

Table 4-4: Comparison between Base Stochastic and Stochastic Temperature.

Interval	Parameter	Base Stochastic	Stochastic Temperature
70%	Mean Width	0.05	0.05
	Max Width	0.13	0.13
	p-Value: t-test	0.03	
	p-Value: regression	<0.001	
	Regression coefficient	0.999	
95%	Mean Width	0.10	0.10
	Max Width	0.25	0.25
	p-Value: t-test	0.08	
	p-Value: regression	<0.001	
	Regression coefficient	0.996	
99%	Mean Width	0.13	0.13
	Max Width	0.33	0.33
	p-Value: t-test	0.06	
	p-Value: regression	<0.001	
	Regression coefficient	0.997	

The above three figures, Figure 4-26, Figure 4-27, Figure 4-28 indicate that credible interval widths approach zero around two particular dates; i.e. mid march and mid October. To explain that, CE-QUAL-W2 input data are plotted against time in Figure 4-29 and Figure 4-30.

Figure 4-29 and Figure 4-30 indicate that there is a relatively higher difference between inflow and withdrawals combined with relatively, at least locally, higher temperature. This indicates that the proportional inflow to the reservoir is not of a stochastic nature. This may explain, at least partly, that stochastic input seems to have less effect (interval widths close to zero) around mid October. Following that, are winter months where evapotranspiration is reportedly at lower annual levels which may explain why stochastic input starts to have an effect on the model output. For mid March, the inflow seems to reach a max, and similar to October, this may explain the less effect of stochastic input on the model.

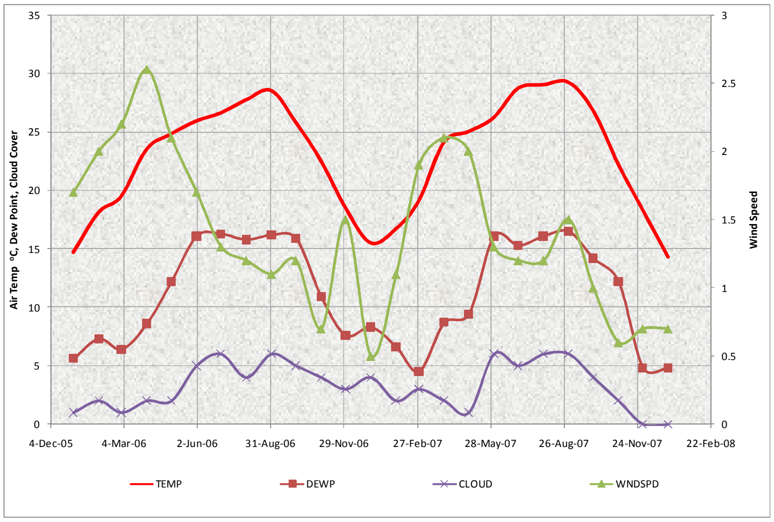


Figure 4-29: Eau Galle Reservoir Meteorological Input.

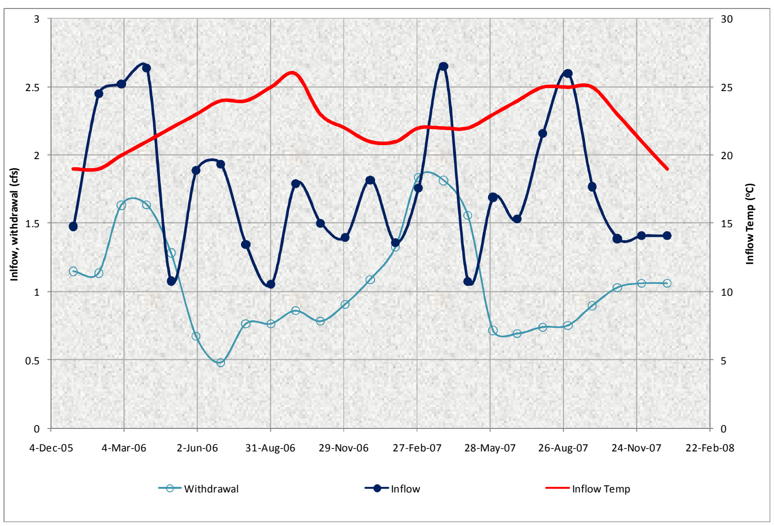


Figure 4-30: Eau Galle Reservoir Inflow and Withdrawal.

As far as the credible intervals are concerned, a repetitive pattern is seen across the three stochastic runs. The basic characteristics of this pattern are listed as follows:

- Wider credible interval in summer months (northern hemisphere).
- Credible interval widths approach zero around mid March and mid October.

- The effect of stochastic input on interval widths peaks decrease annually on the same season. Summer peaks in the second year is less in the first year.

Water surface elevations have also been manually examined. To investigate the stochastic effect on water surface elevations, CE-QUAL-W2 output for segment 6 in the delayed input model was examined (Figure 4-31). Four runs; namely runs, 1, 2, 6 and 13 are considered. Water depth is considered an indicator of water surface elevation, since minimal changes in the reservoir bed morphology are expected.

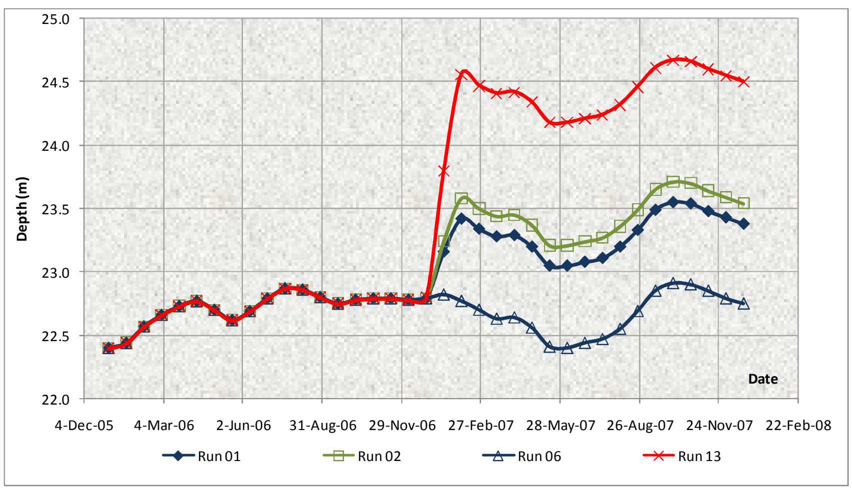


Figure 4-31: Water Depth for Stochastic Runs 1, 2, 6 and 13.

As seen in Figure 4-31 water depth, and hence water surface elevations for all runs are identical up until Julian day 381; i.e. 15th January 2007, when the delayed stochastic input from GSSHA gets into effect. Following this input, a change in the water depth is noticed. Because most stochastic runs are different, the effect that each run has on the depth vary. Run 13 shows the largest gain on water depth (about 1.60 m), while run 6 shows a slight decrease in water depth.

The relative fluctuations in water depths are investigated spatially; i.e. along different reservoir segments, and temporarily; i.e. along time steps. It is noticed that maximum variation in water depth in the reservoir is about 5% which is considerably higher, at least relatively, than the 2% maximum fluctuation in water temperature.

4.3.3 Linkage Findings

In the overall process, it is found that the most time consuming part of the linking is the multiple runs performed by CE-QUAL-W2. Run time varies by the complexity of the model. The most time consuming efforts in the linkage process is running the stochastic version of both GSSHA and CE-QUAL-W2. Pre-processing the linkage, de-aggregating GSSHA output and building CE-QUAL-W2 input takes almost negligible time compared to model runs. As indicated in Figure 4-6, run times can vary significantly for the same model. Varying time steps, total duration of the run and geometric characteristics of the model are among the important factors. A model that runs in less than a minute may take considerably longer time to run with different settings, such as a shorter time step, or a smaller grid cell or additional modeled processes.

The number of stochastic variables is limited to 9 variables for the following reasons.

- A certain limitation in the number of stochastic parameters has to be considered for computational purpose.
- The parameter and value files are harder to manage with more than 9 parameters.
- It is found that 3 - 6 stochastic parameters is a typical number of parameters GSSHA is sensitive for. Any additional parameters modeled stochastically

will just increase the run time without a relative improvement in the final outcome.

4.4 Comparison with EPA's BASINS

The integrated and stochastic water resources modeling framework developed in this research is similar in concept to EPA's BASINS. However, while the approach developed in this research provides linkage of the land and water models, the current version of EPA's BASINS (US-EPA, 2007-a) does not support any automatic linkage between land and water models. HSPF as a model does not have the same power as an integration between a land and a water model. Also the hydrodynamics routines in HSPF are not as rigorous and have known calibration issues.

Another difference is that, BASINS is "manually" stochastic. In a manual stochastic modeling, users have to program in custom tools or integrate with stand-alone packages to obtain stochastic value to generate multiple runs. This should create a set of values for the stochastic runs. These tools can be integrated with statistical packages, e.g. @Risk (Palisade, 2007). As seen in Chapter 3 and earlier in this chapter, this research provides a ready-to-use set of tools for stochastic modeling.

Automated integration is not done in BASINS except between HSPF and AQUATOX (which models aquatic biota in receiving waters) (US-EPA, 2007-a), whereas this research demonstrates how GSSHA and CE-QUAL-W2 are integrated in an automated fashion within the same platform; i.e. WMS, used to pre- and post-process input and output for both models.

As far as uncertainty is concerned, some BASINS-based studies use the implicit margin of safety (MOS). It is to the opinion of the author of this research that the MOS is not the best option to address uncertainty in water resources modeling. MOS is arbitrarily assigned and not based on scientific calculation/theory. Therefore, I think it cannot be considered as an extensive way of addressing, or accounting for, uncertainty. Overestimation of MOS will lead to unnecessary regulation and/or expenses whereas underestimation of it will result in not meeting the desired standards and regulations. This research, as an alternative, proposes the outlined framework to address the uncertainty using sound statistical techniques that are widely used and accepted in various fields. The credible intervals provide a more quantifiable means of addressing the uncertainty.

In summary, the approach developed in this research allows modelers to automatically link two models stochastically and to address uncertainty using quantifiable means; i.e. credible intervals. These tools were developed in an effort to bridge the research gap in stochastic and integrated water resources modeling and at the same time address uncertainty. The developed framework is aimed at helping decision makers in taking better decisions with the available information and resources.

5 Conclusions and Recommendations

The methodology developed in this research utilized GSSHA as a land based watershed model and CE-QUAL-W2 as a receiving water model. Linking a watershed model and a hydrodynamic - water quality model proved to be a comprehensive tool for integrated water resources management. This linkage incorporated stochastic analysis to address uncertainty for integrated water resources modeling.

5.1 Integrated Water Resources

By linking a deterministic GSSHA model to a CE-QUAL-W2 model in this research, GSSHA modelers will have a better understanding of the water quality and hydrodynamics of water bodies in the watershed because of the robust modeling capabilities of CE-QUAL-W2. Similarly, CE-QUAL-W2 modelers will get more accurate input and boundary conditions from GSSHA, including point and non-point source pollutant loadings.

Three main options were discussed in this research for linking the land-based model to the water-based one. These options are:

1. Modeling the water body as a GSSHA lake.
2. Use the new algorithm developed for this research.
3. Use GSSHA index maps.

For ease of application and consistency with other GSSHA modeling process, the GSSHA index map was primarily used in the linkage.

The developed tools in GSSHA's interface in WMS enable modelers to write out the stochastic files (parameter and value files) which are necessary for a stochastic GSSHA run. Modelers can re-load a previously generated value and parameter files into an existing GSSHA model.

As part of this research, the generic version of CE-QUAL-W2 was programmatically modified and a new CE-QUAL-W2 executable was generated and is available to be run within future versions of WMS (beginning with version 8.1) for a deterministic or stochastic run.

It is recommended that modelers implement the linkage on calibrated parent models. Calibrated GSSHA and CE-QUAL-W2 models are important in obtaining the best results from the linkage.

Identical time steps and start time, between GSSHA and CE-QUAL-W2 linked models, were not required to be used in both models. Modelers can use different time steps to link a GSSHA model to a CE-QUAL-W2 model using the framework developed in this research. However, the end time in GSSHA should be set to a maximum of the end time in CE-QUAL-W2. The CE-QUAL-W2 model will not capture any GSSHA output beyond the end time defined in the CE-QUAL-W2 control file unless modelers manually modify the control file accordingly.

The size of both models, i.e. GSSHA and CE-QUAL-W2, in terms of the grid size is typically determined by multiple aspects for both models. However, as far as linkage is concerned, I have found that any size of both models would, theoretically, work. As expected, finer models took longer time to run as opposed to coarser models; i.e. larger grid sizes and/or longer time steps. As an example, GSSHA grid cell size variations will not adversely affect the linkage in and of itself. In fact a larger grid cell size may be better (faster) as far as the linkage is concerned. However, a coarser resolution may not be preferred for a GSSHA model in some cases. It is essential that the modeler maintain a balance between the level of accuracy needed, the available computing resources, the details available and accuracy desired. Although not tested as part of this research, the author believes that network or super-computing might be considered for faster processing.

The tools developed in this research are intended to reduce the time taken for a complete stochastically linked run. However, for practical purposes, 1-D modeling is obviously less time consuming. 2-D modeling, especially integrated models are useful in detailed studies that require more accurate results and better decisions.

5.2 Stochastic Approach

The stochastic GSSHA implementation developed in this research will give modelers more information than a deterministic model. Modelers can investigate stochastic outputs to help in calibration and sensitivity analysis efforts.

Similar to stochastic GSSHA, a stochastically-run CE-QUAL-W2 model provides new insights to modelers and decision makers. This research is thought to enhance reservoir operations and selective withdrawal management schemes by providing a tool to evaluate the uncertainty in model output and show the effect of various results on management decision. These stochastically linked models can determine which parameters the model is more sensitive to. In the test cases, for instance, stochastic temperature was not very important at least relatively (compared to other water quantity related parameters).

Beta distribution was added to the distributions types for stochastic value generation in WMS. Beta-generated values were considerably different, for the case studies and distribution settings used, than the Normal and log-normal distribution-generated and the Log-Normal distribution-generated raw values. There was no strong reason leading to believe that GSSHA flux files resulting from Beta distribution were different from those resulting from normally distributed values. However, it must be noted that this conclusion is unique for the test cases used in this research and for Beta shape factors of 2 and 7. Different values for Beta shape factors may produce different results. Different combinations of these factors give Beta distribution the ability to "approximate" a wide selection of parameter population distribution and hence be a commonly used distribution type.

Without relevant stochastic simulation of model input parameters, uncertainty would be poorly estimated and depend on subjective opinions. Instead of using implicit techniques to address uncertainty (for example: MOS), it is recommended that modelers consider stochastic modeling as appropriate. A robust confidence level estimate cannot be quantified without an appropriate stochastic simulation.

Time series plots are usually generated for a "point of interest" which is a cell in the CE-QUAL-W2 model identified by a segment and a layer. A deterministic line may over-, under-estimate values. The tools developed in this research allow modelers and decision makers to address uncertainty in the overall modeling process. This is accomplished by providing modelers with the option to view output with "the single line" approach or adding a range of values. The "single line" result of this research is a better representation of the mean values, as opposed to the deterministic line. The other option that this research provides is the credible interval. Modelers will be able to view output with a range of potential values associated with a credibility level.

The credible interval time series plot indicate a representative value at each time step along with an "envelope" encompassing a range of possible values. As an example, a credible interval time series, signified by 95% confidence level, were used to infer that there was a 95% chance that temperature would be between the lower and upper bounds of the interval. Credible interval time series plots can also be used to infer that the parameter value would exceed a specific water quality standard 95%, for instance, of the time. The time series plots, in that sense, are useful in evaluating the water quality, or quantity, at the outlet structure of reservoir or a dam.

The developed approach enables modelers to re-plot the time series with different level of confidence. Results shown in Chapter 4, confirmed that, the more confidence we seek, the wider the range and vice versa. A narrower credible interval came at the expense of less confidence. As an example, the 77% credible interval was narrower than the 95% one, but we were less confident in the 77% credible interval.

I recommend using at least 25 model runs to get reliable credible interval. However, this should always be looked at on a case by case basis. The modeler should always try to maintain a balance between the number of runs needed for a reliable inference, the computing resources and time available. The larger the number of simulations used, the less effect on statistical assumptions violations, especially the normality assumption.

Prior to modelers implementing this approach, it is recommended to:

- Run a 5-10 simulations prototype model to get an idea of the width of the interval. This prototype can be used to estimate the number of runs that are needed for a user-defined interval width.
- Perform parameter importance analysis to determine which parameters to model stochastically. With the objective of finding the most critical parameter in any modeling process, sensitivity analysis efforts should be replaced by parameter importance. A model sensitive to a specific parameter does not necessarily mean that this parameter should be looked at more closely for all cases.
- Understand the assumptions and limitations of the linkage. Neither the linkage nor the models are suitable for all conditions and cases. Modelers must confirm that the limitations of this approach do not violate the assumptions made in the modeled system. A full detailed list of the assumptions and limitations of this framework are outlined in Chapter 4.

Stochastic tools developed in this research will help modelers in case of lack of field data. Modelers can simulate parameter values using Monte Carlo Simulation with three sampling methods to choose from in coordination with four types of statistical distributions. Using prior knowledge or depending on previous research or using educated engineering judgment help determine the necessary distribution settings; i.e. minimum, maximum, standard deviation of the missing data.

5.3 Case Studies

The developed integrated and stochastic water resources framework was tested on Lake Zapotlan and Eau Galle Reservoir watersheds. The previously generated, deterministic, GSSHA models were used to build stochastic models for the same test cases. GSSHA output is used to generate input for the previously generated CE-QUAL-W2 model for the receiving water body. The results showed the following:

- Reservoir water depth shows more tangible fluctuations than water temperatures. This finding was expected since all parameters modeled stochastically were water quantity related. The spatio-temporal water depths fluctuations amounting to 5% which is more than double the spatio-temporal temperature which amounts to about 2%.

  Fluctuations in temperature varied greatly by location in the reservoir profile. Generally, the top surface gages showed the lowest fluctuations in temperature. Locations closer to reservoir bottom experienced a little more fluctuation. Mid-depth gages showed the maximum fluctuations in temperature. This stochastic profile analysis is helpful in a reservoir selective withdrawal scheme where a withdrawal column is installed and water is abstracted from the reservoir at a certain depth. This depth may vary across seasons satisfying different criteria. Most selective withdrawal schemes consider water quality parameters other than temperature as well. Also, water surface elevations showed more changes within the stochastic run rather than temperature.
- Stochastically varying inflow water temperature along with the stochastic GSSHA output does not have a significant effect on the overall output for a temperature

time series. A "base" stochastic run, with deterministic water temperature input, had almost the same output as the same model but with stochastic water temperature input.

- Using different statistical distributions for input parameters, had led to little, if any, effect on the final stochastic output. Even though, previous research had indicated that selection of statistical distribution is critical, in the two case studies used for this research, using three different statistical distributions did not prove to pose any effect of the results.
- Using an appropriate statistical distribution in addition to prior knowledge of the range of expected values for some hydro-geological parameters may, in some cases, help in the modeling process in case of the lack of reliably measured parameter values.

As indicated in these test cases, the developed tools can be used to determine parameter importance and model sensitivity. Modelers can utilize the developed tools in initial model reconnaissance, where the main goal is to determine which parameters; the model is more sensitive to. Accordingly, modelers can determine important parameters to invest more time and effort in field measurement.

5.4 Potential Future Research

This research effort tried to bridge part of the gap in the integrated and stochastic water resources arena. Nevertheless, and like any other research, few additional research areas came out as a result, that needed more investigation. These investigations were beyond the original scope and objectives outlined for this framework. Future research efforts related to this research may involve the following:

- Include water quality loadings from GSSHA as soon as GSSHA's nutrient module is in full function. This will generate stochastic constituent time series input files (*_ctr.npt), for CE-QUAL-W2, in a similar fashion to the inflows outlined in this research (*_qtr.npt).
- Research the need to generate grids and maps of probability of exceedance to a certain water quality threshold. In a similar fashion to credible interval time series plots, grids can be generated for a water body that indicate areas that exceed water quality standards at a given confidence level.
- Update the linkage process for a 3-D CE-QUAL-W2 since the model is anticipated to undergo a 3-D update in the future. The linkage should work essentially the same as long as the lateral variations are averaged.
- Incorporate stochastic parameters in CE-QUAL-W2.
- Test the developed integrated and stochastic approach on a number of other watersheds. Each watershed is hydro-geo-morphologically unique and other applications will determine modifications and updates, if applicable, that are necessary. Also, examine if the choice of statistical distributions is critical for these watersheds.
- Investigate distribution of the GSSHA output through the various layers of the neighboring CE-QUAL-W2 segments.
- Research is necessary to determine how to use distributions efficiently. A "best" distribution to represent parameter populations is to be investigated.

- Use the WMS Model Wrapper for the CE-QUAL-W2 multiple runs. Currently, the CE-QUAL-W2 runs stochastically without modeler intervention, but in a DOS window that closes after successful termination of each individual run.
- Investigate linking other models using the same integrating concept. Depending on the models linked, the spatial and temporal domains must be set in accordance. The output of the watershed model has to be processed to generate an input for the receiving water body model. The models may not share the same time steps and the start/end time of the modeling period.

Indeed, integrated water resources modeling helps water resources professionals in the decision making process especially when uncertainty is appropriately addressed. It is recommended that addressing uncertainty and accounting for parameter importance and variations in parameter values be researched more. Implicit approaches in addressing uncertainty should be used in minimum occasions and where quick answers are requested. Stochastic analysis in hydrologic, hydraulic and water quality fields is yet to be explored.

References

Albek, M., Ogutveren, U.B. and Albek, E. (2004). *Hydrological modeling of Seydi Suyu watershed (Turkey) with HSPF.* s.l. : Journal of Hydrology, Elsevier, V.285, P. 260–271.

Ashkar, F. and Arsenault, M. (1998). *Hydrological Frequency Analysis: How critical is the Choice of Statistical Distribution.* Halifax, Nova Scotia : s.n. pp. 215-223.

Bicknell, B. R., et al. (2001). *"Hydrological Simulation Program-Fortran (HSPF), version 12, User's Manual.* s.l. : AQUA TERRA Consultants, Mountain view, California, U.S.A..

Bobba, A. G., Singh, V. P. and Bengtsson L., L. (1996**).** *Application of First-Order and Monte Carlo Analysis in Watershed Water Quality Models.* s.l. : Journal of Water Resources Management, Volume 10, P. 219-240.

Bonta, J. and Cleland, B. (2003). *Incorporating Natural Variability, Uncertainty and Risk into Water Quality Evaluations Using Duration Curves.* s.l. : Journal of the American Water Resources Association (JAWRA), pp. 1481-1296. Vols. 39, Issue 6.

Bury, K. (1999). *Statistical Distribution in Engineering.* s.l. : Cambridge University Press.

Byrd, A., Nelson, E. J. and Downer, C. W. (2005). *Primer: Using Watershed Modeling System (WMS) for Gridded Surface Subsurface Hydrologic Analysis (GSSHA) Data Development—WMS 7.1 and GSSHA 2.0.* s.l. : ERDC/CHL TR-05-XX.

Calder, A. (2005). *Blue Revolution: Integrated Land and Water Resources Management.* s.l. : Earthscan Publications Ltd., Second Edition .

Clarke, R. T. (1994). *Statistical Modeling in Hydrology.* s.l. : John Wiley & Sons.

Cole, T. and Wells, S.A. (2007). *CE-QUAL-W2: A Two-Dimensional, Laterally Averaged, Hydrodynamic and Water Quality Model, Version 3.5, User Manual.* Vicksburg, MS, USA, US Army Corps of Engineers, Instruction Report EL-06-1.

Cotter, A.S., et al. (2003). *Water Quality Model Output Uncertainty as Affected by Spatial Resolution of Input Data.* s.l. : Journal of the American Water Resources Association (JAWRA), pp. 977-986. Vols. 39, Issue 4.

Donkel, D. (2002). *The Aquatic Plant Community in Lake Eau Galle, Dunn County.* http://dnr.wi.gov/fish/pages/reports/final/dunn_lakeeaugalle.pdf : Department of Natural Resources, Wisconsin, MWBC: 2056600.

Downer, C.W. and Ogden, F.L. (2006). *Gridded Surface and Subsurface Hydrological Analysis (GSSHA) User's Manual. Version 1.43 for Watershed Modeling System 6.1". ERDC Technical Report, Engineering Research and Development Center, Vicksburg, MS, USA.* Vicksburg, MS. ERDC/CHL SR-06-1 : US Army Corps of Engineers, Engineering Research and Development Center.

EMS-i. (2008). Watershed Modeling System, Version 8.1. [Online] Environmental Modeling Systems, Inc., http://www.ems-i.com/WMS/WMS_Overview /wms_overview.html.

Gautirrez, S. (2007). *Implementation of Digital Information To Design Hydrologic Models in Mexico.* Provo, Utah, U.S.A. : M.Sc. Project, Department of Civil and environmental Engineering, Brigham Young University.

Good, P. I. and Hardin, J. W. (2006). *Common Erors in Statistics, and How to Avoid Them.* s.l. : Second Edition, Wiley-Interscience.

Hays, W. L. and Winkler, R. L. (1970). *Statististics: Probability, Inference and Decision.* s.l. : Volume 1, Holt, Rinhart and Winston, Inc..

Healey, J. F. (1999). *Statistics: A Tool for Social Research.* s.l. : Fifth Edition, Wadsworth Publishing Company..

Hession, W. C. and Storm, D. E. (2000). *Watershed-Level Uncertainties: Implications for Phosphorous Management and Eutrophication.* s.l. : Journal of Environmental Quality, Volume 29, P. 1172-1179.

Heyman, U., Ryding, S. O. and Forsberg, C. (1984). *Frequency Distribution of Water Quality Variables. RElationships Between Mean and Maximum Values*. pp. 787-794.

Ashkar, F. and Arsenault, M. (1998). *Hydrological Frequency Analysis: How critical is the Choice of Statistical Distribution.* Halifax, Nova Scotia : s.n., Annual Conference of the Canadian Society for Civil Engineering. pp. 215-223.

Jimenez, M. A., De Anda, J. and Shear, H. (2006). *Jimenez M.A., De Anda J., Shear H., "Nutrient/Food Chain Model for Lake Zapotlan (Mexico).* s.l. : International Journal of River Basin Management, Vol. 4, No. 2, P.125-135, IAHR , INBO & IAHS.

Johnson, M. S., et al. (2003). *Application of Two Hydrologic Models with Different Runoff Mechanisms to a Hillslope Dominated Watershed in Northeastern US: A comparison of HSPF and SMR.* s.l. : Journal of Hydrology.

Karamouz, M., Szidarovszky, F. and Zahraie, B. (2003). *Water Resources Systems Analysis.* s.l. : Lewis Publishers.

Krenkel, P. A. and Novotny, V. (1980). *Water Quality Management.* San Diego, CA, U.S.A. : Academic Press.

Landau, D. P. and Binder, K. (2005). *A Guide to Monte Carlo Simulations in Statistical Physics.* s.l. : Cambridge University Press. 2nd Edition.

Law, A. M. and Kelton, W. D. (2000). *Simulation Modeling and Analysis.* s.l. : McGraw Hill. 3rd Edition.

Lindsey, J. K. (2004). *Statistical Analysis of Stochastic Processes in Time.* s.l. : Cambridge University Press.

Marco, J. B., Harboe, R. and Salas, J. D. (1993). *Stochastic Hydrology and its Use in Water Resources Systems Simulation and Optimization.* s.l. : NATO ASI Series..

Mason, D. (2001). *Probability Density Function in Program Analysis.* Toronoto, Canada : 4th ICSE Workshop on Componenet-Based Software Engineering.

McBride, G. B. (2005). *Using Statistical Methods for Water Quality Management, Issues, Problems and Solutions.* s.l. : Wiley-Interscience.

McCuen, R.H. (2005). *Hydrologic Analysis and Design.* 3rd Edition. s.l. : Pearson Prentice Hall.

McCuen, R.H.(2002). *Modeling Hydrologic Change: Statistical Methods* . s.l. : CRC, Lewis Publishers.

Melching, C. S. and Bauwens, W. (2001). *Uncertainty in Coupled Non-Point Source and Stream Water-Quality Models.* s.l. : ASCE, pp. P. 403-412. Vol. 127(6).

Miller, W. G. (2006). *Statistics and Measurement Using the Free OpenStat Packages.* s.l. : URL: http://www.statpages.org/miller/openstat/, Last updated 3rd July 2007.

Nelson, E. J. (2008). *Watershed Modeling System, version 8.1, Tutorial.* s.l. : Environmental Modeling Research Laboratory, Department of Civil and Environmental Engineering, Brigham Young University, Provo, Utah, U.S.A.

NIST. (2007). National Institute for Standards and Technology (NIST). *Engineering statistics Handbook.* [Online] July 2007. URL: http://www.nist.gov.

Olufemi, O. O., Wei, Z. and Beck, M. B. (2003). *Coping with Uncertainty: A Case Study in Sediment Transport and Nutrient Load Analysis.* s.l. : ASCE, Journal of Water Resources Planning and Management, Volume 129, Issue 4, pp. 345-355.

Palisade. (2007). *Guide to Using @RISK. Risk Analysis and Simulation. Add-in for Microsoft Excel. Version 5.0.* New York : Palisade.

Paz, A. (2007). *Two Dimensional Hydrodynamics and Water Quality Model of Lake Zapotlan, Mexico Using CE-QUAL-W2.* s.l. : M.Sc. Thesis, Department of Civil and Environmental Engineering, Brigham Young University, Provo, Utah, U.S.A.

Press, W. H., et al. (2005). *Numerical Recipes in C++. The Art of Scientific Computing.* s.l. : Cambridge University Press, Second edition.

Ramsey, F. L. and Schafer, D. W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis.* s.l. : Duxbury, Second Edition.

Reckhow, K. H. (2003). *On the Need for Uncertainty Assessment n TMDL Modeling and Implementation.* s.l. : Journal Water Resources Planning and Management, Volume 129, Issue 4, P. 245-246, Editorial, American Society of Civil Engineers.

Ricciardi, K. L., Pinder, G. F. and Belitz, K. (2005). *Comparison of the lognormal and beta distribution functions to describe the uncertainty in permeability*. pp. 248-256.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods.* s.l. : Springer, 2nd Edition.

Salah, A. M. and Nelson, E. J. (2005-a). *Water Resources/Quality Modeling using Hydrological Simulation Program-Fortran (HSPF) and Watershed Modeling System (WMS); Technical Concerns Resolved..*

Salah, A. M. (1999). *Productive and Sustainable Use of Water Among Competing Sectors.* Enschede, The Netherlands : The Internatinoal Institute for Aerospace Survey and Earth Sciences (ITC), M.Sc. Thesis.

Salah, A.M., Fields, P.J. and Miller, A.W. (2005-b). *Simulating Uncertainty in Mass Balance Modeling for Fresh Water Reservoirs; Case Study: Deer Creek Reservoir, Utah, U.S.A.* [ed.] Winter Simulation Conference. Orlando, Florida : INFORMS, WSC Foundation.

Shamsi, U. M. (2005). *GIS Applications for Water, Wastewater, and Stormwater Systems.* s.l. : CRC Press.

Shiklomanov, I. A. (1999). *World Water Resources and Their Use.* s.l. : UNESCO/IHP Publications, UNESCO and State Hydrological Institute (SHI).

Shoemaker, C. A. and Benaman, J. (2003). *Methodology for Sensitivity Analysis in Complex distributed Watershed Models.* s.l. : American Society of Civil Engineers, World Water and Environmental Resources Congress.

Smemoe, C. M. (2004). *Floodplain Risk Analysis Using Flood Probability and Annual Exceedance Probability Maps.* Provo, Utah, U.S.A. : Ph.D. Dissertation, Department of Civil and Environmental Engineering, Brigham Young University.

Steeb, W. H. and Solms, F. (2000). *C++ Programming with Applications in Administration, Finance and Statistics.* s.l. : World Scientific.

Takyi, A. K. (1995). *Incorporating Input Data Uncertainty in Water Quality Management Models: Examples for Non-Seasonal and Seasonal Wasteload Allocation.* s.l. : PhD dissertation, University of Manitoba, Winnipeg, Manitoba, Canada.

Taylor, H.M. (1998). *An Introduction to Stochastic Modeling.* Third Edition. s.l. : Academic Press Limited.

US-ACE. (1995). *CE-QUAL-RIV1: A Dynamic, One-Dimensional (Longitudinal) Water Quality Model for Streams: User's Manual.* Vicksburg, MS : Instr. Rpt. EL-95-2, USACE Waterways Experiments Station.

US-ACE. (2007). Eau Galle Recreation Area. *US Army Corps of Engineers; Recreation.* [Online] US Army Corps of Engineers, November 2007. http://www.mvp.usace.army.mil/recreation/default.asp?pageid=153.

US-ACE. (1998). HEC-1. [Online] 6 1998. [Cited: 01 01, 2008.] http://www.hec.usace.army.mil/software/legacysoftware/hec1/hec1.htm.

US-ACE. (2006). HEC-RAS. [Online] 11 2006. http://www.hec.usace.army.mil/software/hec-ras/hecras-hecras.html. Version 4.0.0 Beta.

US-EPA. (2007-a). *Fact Sheet: Better Assessment Science Integrating Point and Non-point Sources (BASINS, version 4).* Washington, D.C. : USEPA office of Water., 2007-a. EPA-823-F-01-021.

US-EPA. (2007-b). Environmental Fluid Dynamics Code (EFDC). [Online] 9 4, 2007-b. http://www.epa.gov/athens/wwqtsc/html/efdc.html.

US-EPA. (2007-c). Watershed and Water Quality Modeling Technical Support Center. [Online] 9 4, 2007-c. http://www.epa.gov/athens/wwqtsc/index.html.

US-EPA. (2005). Introduction to Total Maximum Daily Loads (TMDLs). [Online] EPA, Office of Water, Feb 2005. [Cited: Dec 17, 2005.] http://www.epa.gov/owow/tmdl/intro.html.

US-EPA. (2002). *Status of the TMDL/Watershed Rule.* s.l. : U.S. Environmental Protection Agency, Office of Water, 2002. URL: http://www.epa.gov/owow/tmdl/watershedrule/watershedrulefs.html..

US-EPA. (1999-a). Protocol for developing nutrient TMDLs. [Online] EPA, Office of Water, 1999-a. http://www.epa.gov/owow/tmdl/nutrient/pdf/nutrient.pdf. EPA-841-B-99-007.

US-EPA. (1999-b). *Introduction to Water Quality Standards.* Washington D.C. (: Office of Water, 1999-b. EPA-823-F-99-020.

Valdes, J. B., Diaz-Granados, M. and Bras, R. L. (1990). *A Derived PDF for the initial soil moisture in a catchment.* pp. 163-176.

Walker, W. W. (2003). *Consideration of Variability and Uncertainty in Phosphorus Total Maximum Daily Loads for Lakes.* s.l. : Journal of Water Resources Planning and Management, Volume 129, Issue 4, P. 337-344, American Society of Civil Engineers.

Washington. (2005). *Use Attainability Analysis Guidance to Washington State.* s.l. : Washington State, Department of Ecology. Available online at http://www.ecy.wa.gov/biblio/.

WEF. (1998). *Clean Water Action Plan Announced.* s.l. : Water Environment Federation, Watershed and Wet Weather. 3:4-5..

Wells, S. (2002). *Basis of the CE-QUAL-W2 Version 3 River Basin Hydrodynamic and Water Quality Model.* s.l. : Proceedings, 2nd Federal InterAgency Hydrologic Modeling Conference, Las Vegas, NV.

WHO. (2004). *Guidelines for Drinking-Water Quality.* s.l. : World Health Organization Publications, 3rd Edition, .

WHO/UNICEF. (2005). *Water for Life: Making it Happen.* s.l. : World Health Organization Publications.

Wichura, M. J. (1988). *Algorithm AS 241: The Percentage Points of the Normal Distribution.* s.l. : Applied Statistics, vol. 37, pp. 477-484.

Williams, G. P., Tomasko, D. and Tebes, C. L. (1995). *Parameter Importance for an Analytic Model of LNAPL Transport in the Vadose Zone.* s.l. : Conference Proceeding, Groundwater Management, American Society of Civil Engineers, pp. 343-348.

Wu, J., Zou, R. and Yu, S. L. (2006). *Uncertainty Analysis for Coupled Watershed and Water Quality Modeling Systems.* s.l. : Journal of Water Resources Planning and Management, ASCE, Sep./Oct. 2006, P.351-361.

WWC. (2007). World Water Council (WWC). *Water at a Glance.* [Online] June 2007. [Cited: August 22, 2007.] http://www.worldwatercouncil.org.

Wyss, G. D. and Jorgensen, K. H. (1998). *A User's Guide to LHS: Sandia's Latin Hypercube Sampling Software.* s.l. : Risk Assessment and Systems Modeling Department, Sandia National Laboratories, SAND98-0210.

Zhang, H. X. (2001). *The Critical Flow-Storm Approach and Uncertainty Analaysis for the TMDL Development Process.* s.l. : Ph.D. Dissertation, University of Virginia, Charlottesville, VA.

Appendix A GSSHA/CE-QUAL-W2 Linkage in Watershed Modeling System (WMS): A Primer

This document highlights on the linkage between GSSHA and CE-QUAL-W2. It will walk you through the linkage process using Eau Galle Reservoir dataset.

In order to establish a stochastic integrated link between the two models, a model, each, for the area need to be already set and running. Modelers are highly encouraged to set and calibrate both models individually and deterministically before attempting to link them stochastically.

Before starting to work on this primer, you should have the following fully functional deterministic models for the watershed (Figure A-1):

- Eau Galle Reservoir GSSHA model (EauGalle_GS).
- Eau Galle Reservoir CE-QUAL-W2 model (EauGalle_W2).

Now, we will proceed with the primer.

A. Initial Setup

Let us make sure that WMS is ready to build the link between GSSHA and CE-QUAL-W2. The CE-QUAL-W2 model needs to open first.

1. Open the WMS project for the CE-QUAL-W2 model (EauGalle_W2.wpr) by selecting "*File* | *Open*" from the main WMS menu. When prompted to locate the

control and bathymetry files, go ahead and select “*Ok*” from the respective dialogs to browse for these files (Figure A-2Figure A-1), (Figure A-3).

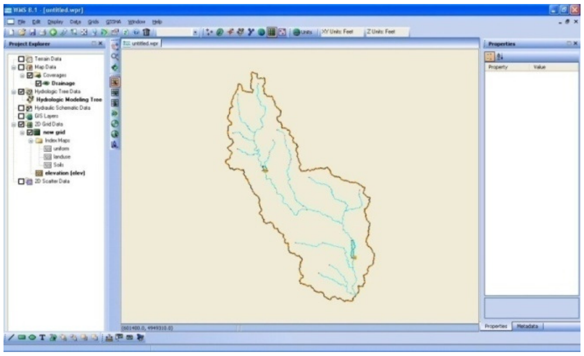


Figure A-1: Eau Galle Watershed.

2. Activate the two dimensional Grid Module.

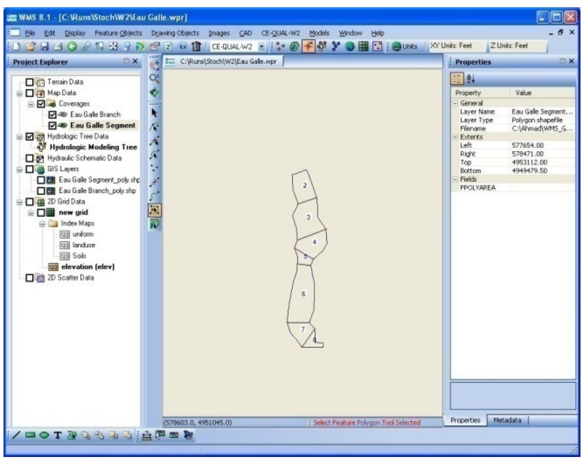


Figure A-2: Eau Galle Reservoir CE-QUAL-W2 Segment IDs.

3. Open the GSSHA project by selecting “*Open Project*” from the “*GSSHA*” menu. This should open the existing GSSHA model (Figure A-4).

N.B.: Make sure that GSSHA’s project full file name including the path and the extension is not more than 80 alpha-numeric characters.

4. Browse and select the file EauGalle_GS.prj to open.

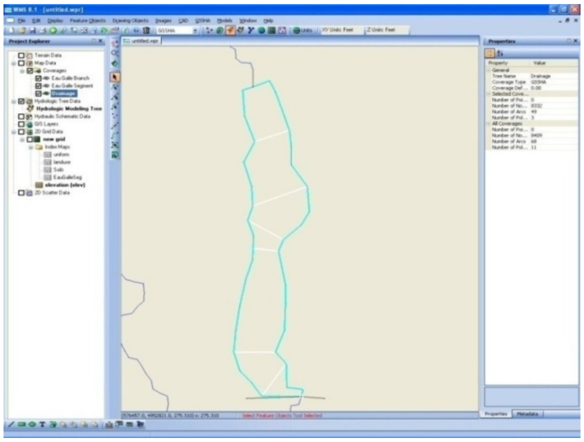


Figure A-3: Eau Galle Branch Coverage.

B. Establishing the Spatial Link

When both models are open, the spatial link can be established as follows:

5. Make sure that you have a valid branch and segment coverages. Segments are mapped to branches and that segments are numbered properly (dummy segments, ascending segment numbers from upstream to downstream).

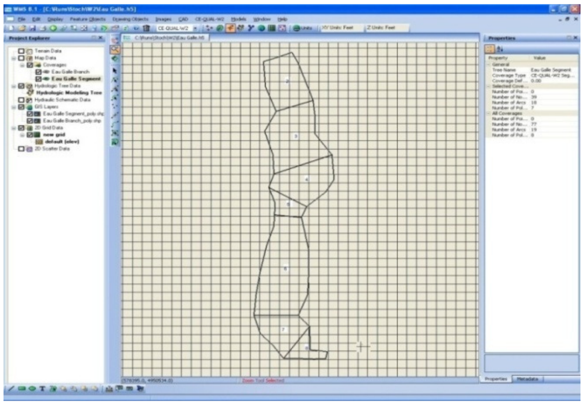


Figure A-4: Eau Galle Segments over a GSSHA Grid.

6. Make sure that the GSSHA model is setup with appropriate cell size. The cell size should be small enough to get each segment in the CE-QUAL-W2 model represented by AT LEAST one cell. In this case, cell size is 100 m.
7. Activate the "*Map Module*". You should have the "*CE-QUAL-W2*" menu available on the main WMS menu. If not, navigate to "*Models*" menu and select "*CE-QUAL-W2*".
8. Select "*GSSHA to CE-QUAL-W2 Link*" from the "*CE-QUAL-W2*" menu. This should generate the segment ID index map that is necessary for the spatial linkage.

N.B.: The "*GSSHA to CE-QUAL-W2 Link*" command (Figure A-5) should be undimmed. If not, you will need to re-initialize the model. Check WMS (Nelson, 2008) tutorial for further details on initializing a CE-QUAL-W2 model.

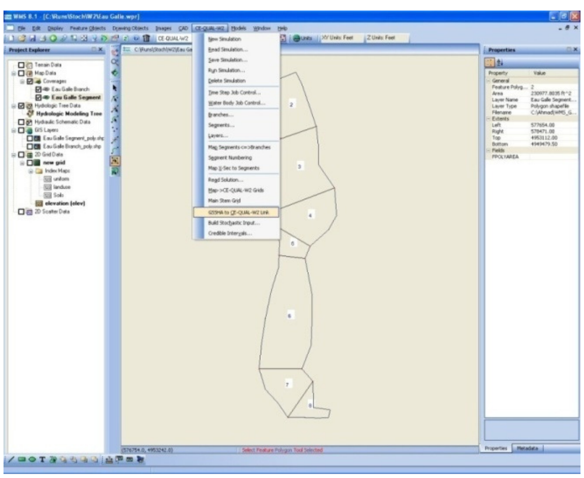


Figure A-5: Link GSSHA to CE-QUAL in WMS.

9. A message will come up (Figure A-6) indicating that the segment ID index map will be generated.

10. Navigate to the "*Grid Module*". Select "*GSSHA | Save Project*". This will bring the "*Save GSSHA Project File*" dialog (Figure A-7) to save the GSSHA project with the index map in the specified location.

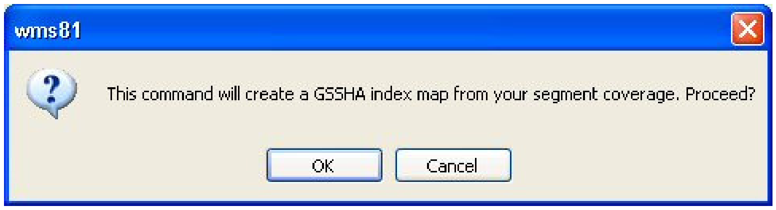


Figure A-6: Segment ID Index Map Confirmation Message.

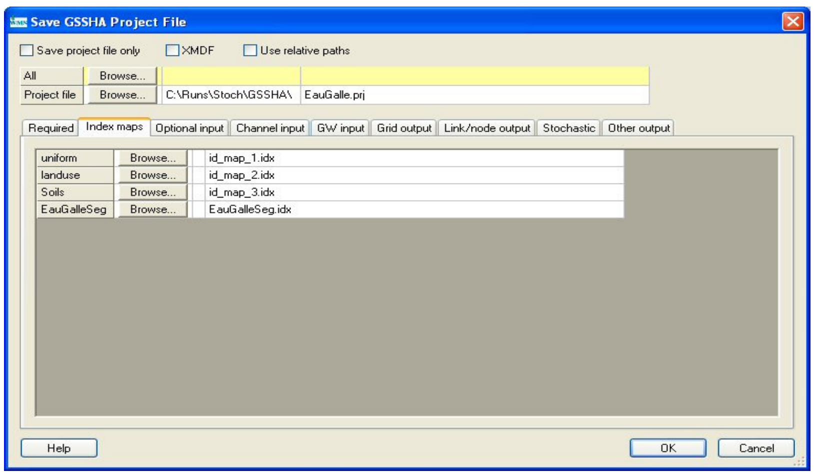


Figure A-7: Save GSSHA Project File Dialog – Index Map Tab.

As noted in this "Save GSSHA Project File" dialog, and in the "Index Maps" tab, the segment ID index map (EauGalleSeg.idx) should be saved to the location indicated.

N.B.: Note that until the GSSHA project is saved, the segment index map is not saved.

The spatial link between the two models is established using a GSSHA index map of the CE-QUAL-W2 segment IDs. The index map should be generated using the same GSSHA grid; i.e. cell size.

Your WMS screen should look like (Figure A-8) with the red cells indicating the GSSHA model and the other colored ones are for the segment ID index map.

N.B.: Once the GSSHA project file (Figure A-9) is saved, make sure a new card "*## GSSHA to W2*" is created and added towards the end of the file.

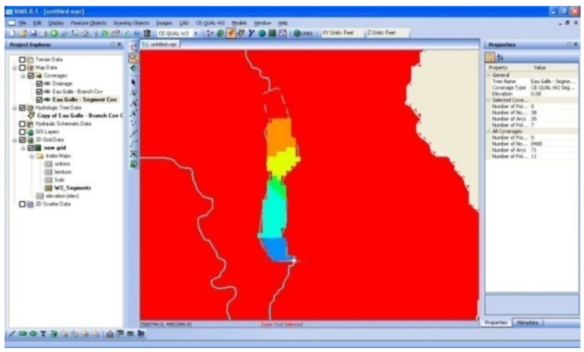
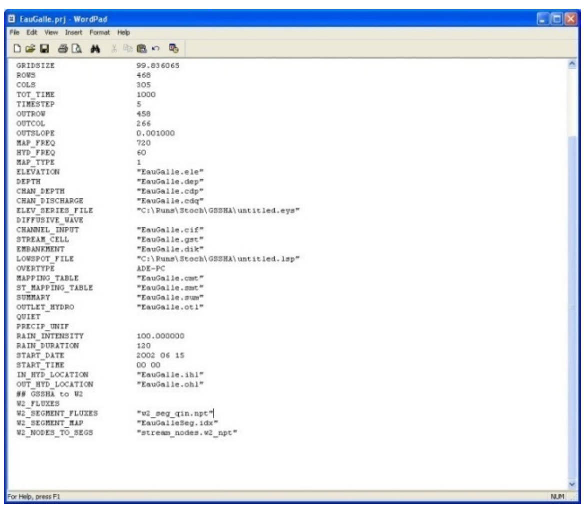


Figure A-8: Eau Galle Reservoir Index Map.



```
EauGalle.prj - WordPad
GRIDSIZE              99.836065
ROWS                  468
COLS                  305
TOT_TIME              1000
TIMESTEP              5
OUTROW                458
OUTCOL                266
OUTSLOPE              0.001000
MAP_FREQ              720
HYD_FREQ              60
MAP_TYPE              1
ELEVATION             "EauGalle.ele"
DEPTH                 "EauGalle.dep"
CHAN_DEPTH            "EauGalle.cdp"
CHAN_DISCHARGE        "EauGalle.cdq"
ELEV_SERIES_FILE      "C:\Runs\Stoch\GSSHA\untitled.eys"
DIFFUSIVE_WAVE
CHANNEL_INPUT         "EauGalle.cif"
STREAM_CELL           "EauGalle.gst"
EMBANKMENT            "EauGalle.dik"
LOWSPOT_FILE          "C:\Runs\Stoch\GSSHA\untitled.lsp"
OVERTYPE              ADE-PC
MAPPING_TABLE         "EauGalle.cmt"
ST_MAPPING_TABLE      "EauGalle.smt"
SUMMARY               "EauGalle.sum"
OUTLET_HYDRO          "EauGalle.otl"
QUIET
PRECIP_UNIF
RAIN_INTENSITY        100.000000
RAIN_DURATION         120
START_DATE            2002 06 15
START_TIME            00 00
IN_HYD_LOCATION       "EauGalle.ihl"
OUT_HYD_LOCATION      "EauGalle.ohl"
## GSSHA to W2
W2_FLUXES
W2_SEGMENT_FLUXES     "w2_seg_qin.npt"
W2_SEGMENT_MAP        "EauGalleSeg.idx"
W2_NODES_TO_SEGS      "stream_nodes.w2_npt"
For Help, press F1
```

Figure A-9: Sample GSSHA Project File.

C. Stochastics in GSSHA

Once the spatial link is established, modelers can modify GSSHA parameter values and re-setup the model for stochastic runs. As an example to illustrate the process, we will only model two parameters stochastically; i.e. capillary head and hydraulic conductivity.

11. Now, re-activate the “two dimensional Grid Module”.
12. Select “*GSSHA | Sotchastic GSSHA*”. This should bring the “*Stochastic GSSHA*” dialog (Figure A-10).

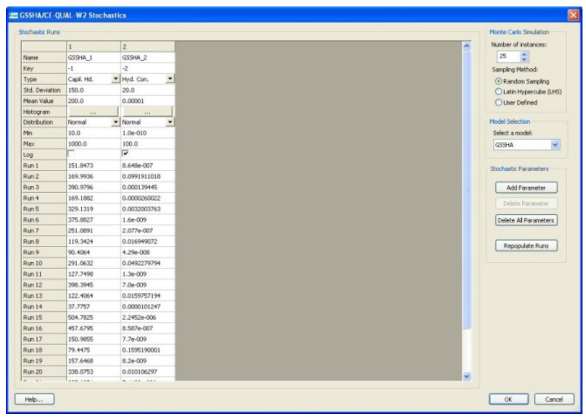


Figure A-10: Stochastic Dialog.

13. Click on “*Add Parameter*” to add capillary head. Select “*Capil. Hd.*” From the “*Type*” field for this parameter.
14. Click on “*Add Parameter*” to add hydraulic conductivity. Select “*Hyd. Con.*” From the “*Type*” field for this parameter.

N.B.: If needed, click on “*Repopulate Runs*” to populate parameter values. This command can be used to generate another set of values for the same set of parameters. Note that the distribution type, standard deviation, minimum and maximum remain the same after repopulating the runs. All what this command does is perturb a new set of values from the distribution selected.

15. In the "*Distribution*" field, select "*Normal*" for both parameters. However, for the hydraulic conductivity, click on the check-box "*Log*" to use a Log-Normal distribution for hydraulic conductivity.

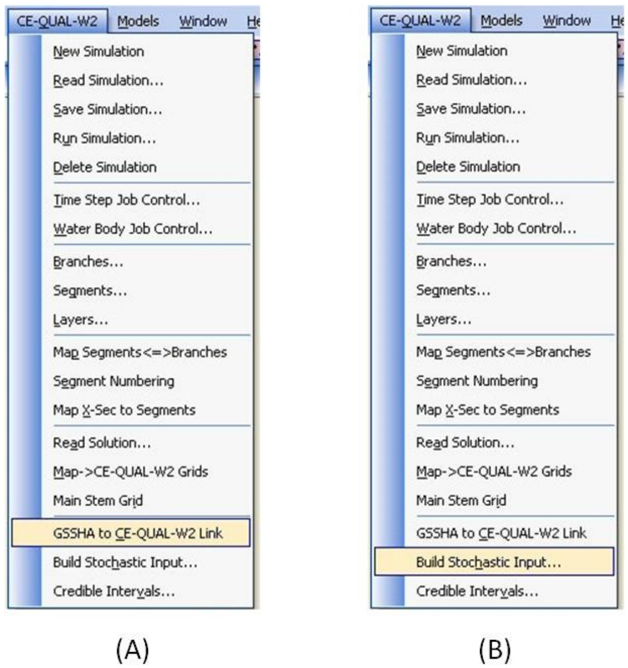


Figure A-11: Menu Commands of the Linkage.

16. Select the "*Number of Instances*" to be 25 instead of 100.

17. Leave the "*Sampling Method*" as default; i.e. "*Random Sampling*".

The Latin Hypercube Sampling (LHS) basically subdivides the PDF into sub-areas and selects only one value for each. A number of instances "a" and a number of variables "b" would result in a total number of simulations equal to a^b.

18. Click on "*Ok*" on the "*GSSHA/CE-QUAL-W2 Stochastics*" dialog.

19. Save the GSSHA project file.

The index map is saved once the GSSHA project is saved. This is designed so to avoid unnecessary used disk space. Modelers typically do not need the index map saved until they actually save a GSSHA project.

N.B.: When a GSSHA project is saved, the parameter and value files are saved. These files are used in the stochastic run.

20. Run a stochastic version of GSSHA. Before you run GSSHA stochastically, make sure that a stochastic card is added to the GSSHA project file.

N.B.: A simulation input file is generated along with the parameter and value files. This File allows users to reload the stochastic parameters and their values used in the previous simulation. Re-populating the runs or selecting other distribution/settings will change these values. However, these changes will not be saved unless the GSSHA project is saved.

D. Generate CE-QUAL-W2 Input

After a successful GSSHA stochastic run, flux files should have been created equal to the number of the simulations/runs specified in the GSSHA stochastic dialog in the previous section. The flux files will be labeled with the number of the run preceding the file name; e.g. 0027_w2_seg_qin.npt. Notice that the file has an "npt" extension as it will be used as "input" to CE-QUAL-W2. All flux files will be generated in the "Stochastic Ouptu" folder generated by WMS.

21. Now, re-activate the "*Map module*".

22. Select "CE-QUAL-W2 | Build Stochastic Input". (Figure A-11, B)

23. Locate the parent folder that you want the run sub-folders to be saved under.

24. After executing this command, you should notice the following:

   a. Flux files broken down to number of files.
   b. Each flux file is broken down into its own folder named with the run number.
   c. CE-QUAL-W2 control file copied and modified to incorporate the broken flux files (Figure A-12).

25. Select “CE-QUAL-W2 | Run Simulation”.

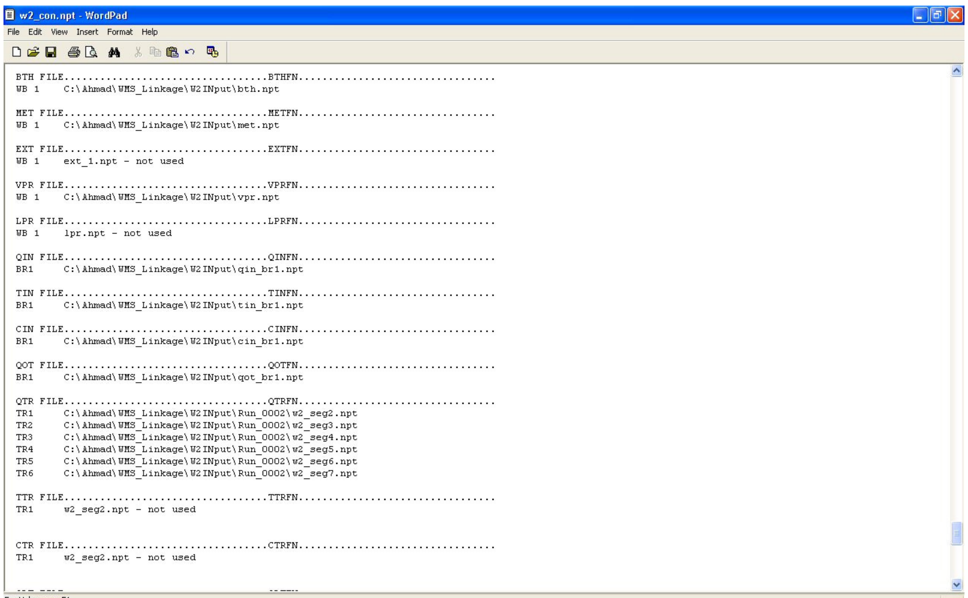


Figure A-12: CE-QUAL-W2 Control File with Flux Files.

26. Select “Stochastic” in the “*Run CE-QUAL-W2 Simulation*” dialog.

27. Navigate to the parent folder that contains all the built runs for CE-QUAL-W2. (from the previous steps)

28. When you click “*Ok*” on the “*Run CE-QUAL-W2 Simulation*” dialog (Figure A-13), you should see the CE-QUAL-W2 runs executed.

E. Read CE-QUAL-W2 Solution

After a successful CE-QUAL-W2 stochastic run, output are saved in the sub-folders under each run number. WMS is ready to read in the various runs.

29. Before we start reading in the solutions, a longitudinal profile of the water body needs to be generated. This profile displays the segments and layers for each branch.

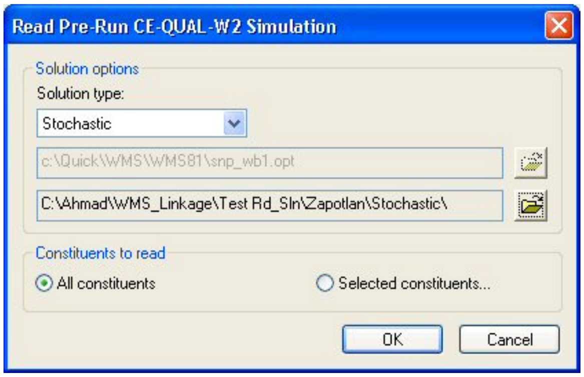


Figure A-13: Read CE-QUAL-W2 Simulation Dialog.

30. Select "*CE-QUAL-W2* | *Map->CE-QUAL-W2 Grids*". This should generate a grid for the selected branch

31. Select "CE-QUAL-W2 | Read Solution". This should bring the "Read Pre-Run CE-QUAL-W2 Simulation" dialog.

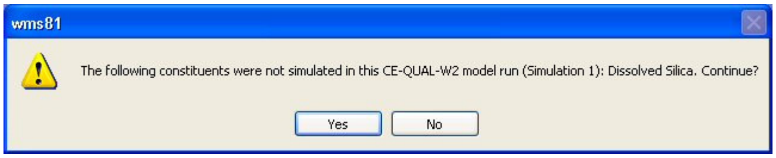


Figure A-14: Simulated Constituents Warning Message.

32. Select "*Stochastic*" from the drop down box on the top left of the dialog.

33. Navigate to the parent folder than contains all the CE-QUAL-W2 runs.

34. After successful reading of the solution, WMS will generate a number of datasets equal to the number of simulations. In Addition to those datasets, minimum, median and maximum datasets will be generated in the following section.

F. View CE-QUAL-W2 Results

To be able to view the stochastic results, a grid must be generated for the branch under investigation as indicated in the previous step. If it is not already setup, WMS will check and ask you if you want to create it as part of the read solution command.

The credible intervals and the level of credibility need to be set to determine the lower and higher bounds of the time series.

35. Highlight the grid desired.
36. Select "*Data | Credible Intervals*". This should bring the "*Select Datasets*" dialog. In this dialog, you should be able to see all the created datasets (i.e. runs) created for the highlighted grid.
37. Select all the datasets (Figure A-15) that you want to include in generating the credible intervals, by using the SHIFT and CTRL keys. More datasets usually indicates higher accuracy and better representation for the credible intervals.
38. Click "*Options*" in the "*Select Datasets*" dialog. This should bring the "*Select Credible Interval"* dialog.
39. Select "User Defined Credibility (e.g. type 95 for 95%)" and type 77 (Figure A-16). Note that this field will only allow users to enter up to two digits.
40. Click "*Ok*" to close the "*Select Credible Dialog*".
41. Click "*Ok*" to close the "*Select Datasets*" dialog.

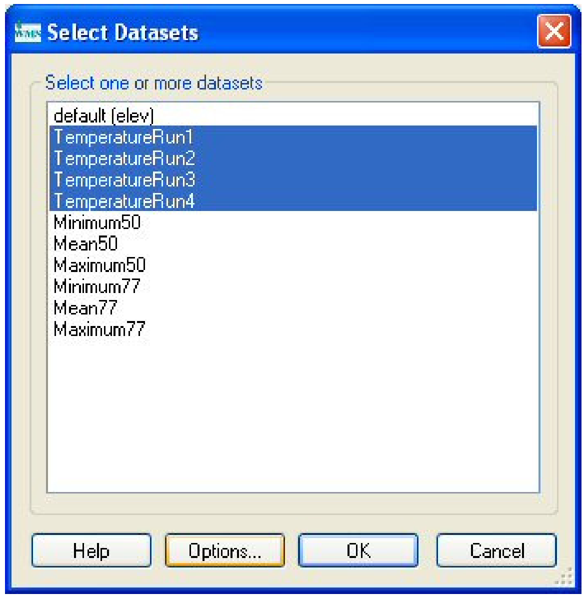


Figure A-15: Selected Datasets Dialog – Select Runs to Create Credible Intervals.

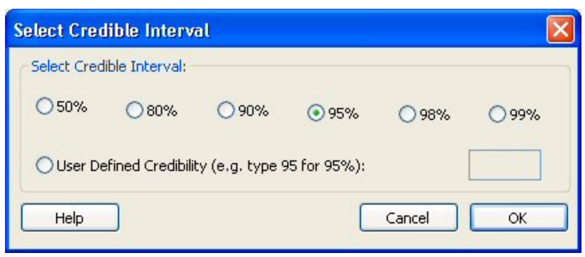


Figure A-16: Select Credible Interval Dialog.

42. This should generate three additional datasets in the main highlighted grid. The three datasets are the 77% credible interval lower bound (minimum77), upper bound (maximum77) and the mean (mean77).

    The interface is ready for the final display of the stochastic time series.

43. Make sure the "*two dimensional Grid Module*" is active.
44. Select the "*Gage*" from the WMS tool palette. Locate a gage (Figure A-17) where necessary (CI_Gage).

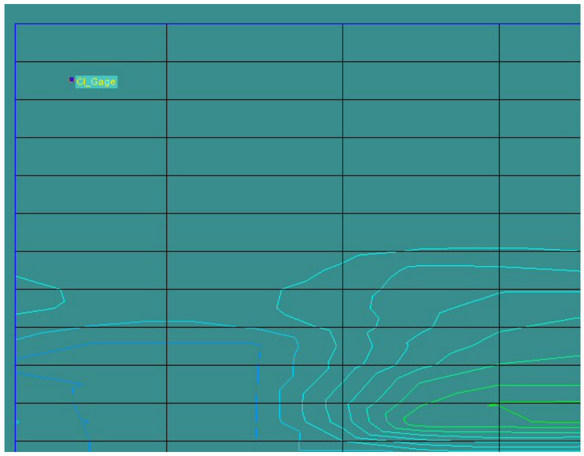


Figure A-17: Gage Location.

45. Select the "*Select Gage*" from the WMS tool palette.
46. Use the mouse to right-click on the gage.
47. Select "*Plot Selected*". This should bring the "*Select Datasets*" dialog.
48. Use the "Shift" key to select the last three created datasets; i.e. Minimum77, Mean77 and Maximum77 (Figure A-18).
49. Click "Ok" on the "*Select Datasets*" dialog.
50. This will create the stochastic time series of the dataset at CI_Gage.

N.B.: Users should label the minimum, maximum and mean datasets as they are generated especially if they are creating multiple credible intervals. As indicated in Figure A-20, there are multiple datasets labeled by the respective credible interval used in creating them.

51. If you want to generate another time series for the same location (gage), right-click on the gage again, and select "*Plot Selected*" (Figure A-19) and select any of the runs that are available in the "*Select Datasets*" dialog (Figure A-20).

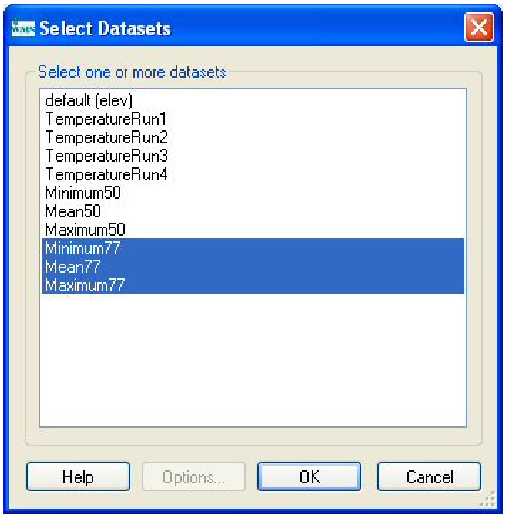


Figure A-18: Selected Datasets Dialog – Select Credible Interval Datasets.

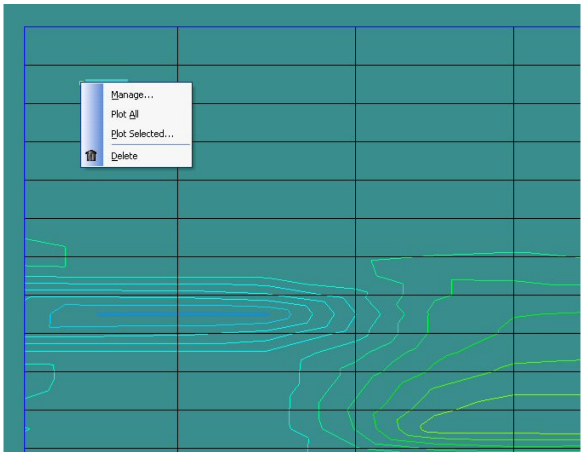


Figure A-19: Gage Location – Plot Selected Datasets.

<u>**N.B.:**</u> Users should notice that reducing the confidence level should always result in a narrower range of plausible values. This means that if users require a higher credibility, that would come at the expense of the width of range of values.

52. Additional analysis can be done by repeating steps 35 through 50 to generate other credible intervals and generating three additional datasets for each credible interval selected to see the effect of the credibility on the interval width.

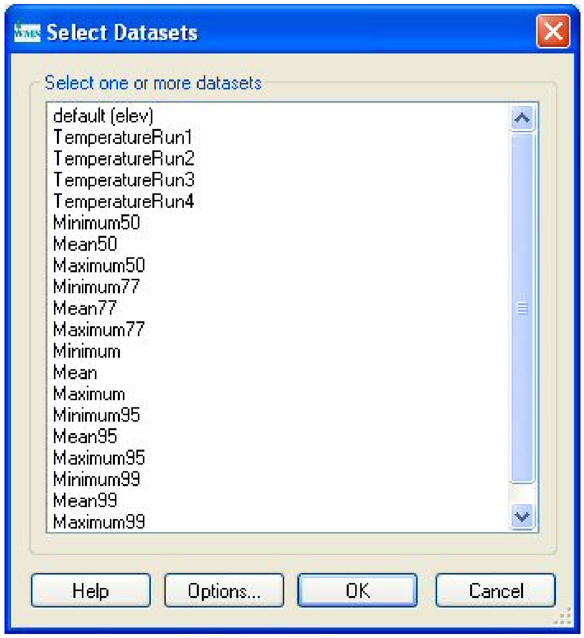


Figure A-20: Select Datasets Dialog.

The general rule is that, the more confident or credible we want the results to be, the wider the range between the lower and upper bounds. The stochastic graph may be interpreted that the temperature ranges between the lower and upper bounds for the designated credibility level (i.e. 95%)

N.B.: Users may rename the created datasets. However, it is always advisable to name the minimum, mean and maximum according to the credible interval chosen.

Among various uses, this application can be utilized in:

- Selective withdrawal analysis for a specific reservoir.
- Integrated Water resources Management.
- Addressing the uncertainty involved in the modeling process.
- Stochastic modeling of water resources/quality.