Using Relevance Vector Machines Approach for Prediction of Total Suspended Solids and Turbidity to Sustain Water Quality and Wildlife in Mud Lake

Hussein Aly Batt

Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/etd

Part of the Civil and Environmental Engineering Commons

Recommended Citation
https://digitalcommons.usu.edu/etd/1315
USING RELEVANCE VECTOR MACHINES APPROACH FOR PREDICTION OF TOTAL SUSPENDED SOLIDS AND TURBIDITY TO SUSTAIN WATER QUALITY AND WILDLIFE IN MUD LAKE

by

Hussein Batt

A dissertation submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Civil and Environmental Engineering

Approved:

David K. Stevens
Major Professor

Thomas Hardy
Committee Member

Gilberto E. Urroz
Committee Member

John Schmidt
Committee Member

Michael J. McFarland
Committee Member

Mark R. McLellan
Vice President for Research and Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah
2012
ABSTRACT

Using Relevance Vector Machines Approach for Prediction of Total Suspended Solids and Turbidity to Sustain Water Quality and Wildlife in Mud Lake

by

Hussein Aly Batt, Doctor of Philosophy
Utah State University, 2012

Major Professor: Dr. David K. Stevens
Department: Civil and Environmental Engineering

Mud Lake is a wildlife refuge located in southeastern Idaho just north of Bear Lake that traps sediment from Bear River water flowing into Bear Lake. Very few water quality and sediment observations, if any, exist spatially in Mud Lake. Spatial patterns of sediment deposition may affect Mud Lake flows and habitat; prediction of those patterns should help refuge managers predict water quality constituents and spatial distribution of fine sediment. This will help sustain the purposes of Mud Lake as a habitat and migratory station for species.

The main objective of the research is the development of Multivariate Relevant Vector Machine (MVRVM) to predict suspended fine sediment and water quality constituents, and to provide an understanding for the practical problem of determining the amount of data required for the MVRVM. MVRVM is a statistical learning algorithm that is based on Bayes theory. It has been widely used to predict patterns in hydrological
systems and other fields. This research represents the first known attempt to use a MVRVM approach to predict transport of very fine sediment and water quality constituents in a complex natural system.

The results demonstrate the ability of the MVRVM to capture and predict the underlying patterns in data. Also, careful construction of the experimental design for data collection can lead the Relevant Vectors (RVs is a subset of training observation which carries significant information that is used for prediction) to show locations of significant patterns.

The predictions of water quality constituents will be of potential value to US Fish and Wildlife refuge managers in making decisions for operation and management in the case of Mud Lake based on their objectives, and will lead the way for scientists to expand the use of the MVRVM for modeling of suspended fine sediment and water quality in complex natural systems.
The main objective of the research is the development of Multivariate Relevant Vector Machine (MVRVM) to predict suspended fine sediment, water quality constituents, and provide an understanding for the practical problem of determining the amount of data required for the MVRVM. MVRVM is a statistical learning algorithm that is based on Bayes theory. It has been widely used to predict patterns in hydrological systems and other fields. This research represents the first known attempt to use a MVRVM approach to predict transport of very fine sediment and water quality constituents in a complex natural system.

The results demonstrate the ability of the MVRVM to capture and predict the underlying patterns in data. Also careful construction of the experimental design for data collection can lead the Relevant Vectors (RVs is a subset of training observation which carry significant information that is used for prediction) to show locations of significant patterns.
To my mother

Leyla,

my father Aly,

my brother Mohamed, my sister May,

and Leila
ACKNOWLEDGMENTS

First of all I would like to thank our GOD for his blessing in my life.

I would like to express my sincere gratitude to my major professor, Dr. David Stevens, for giving me the opportunity to work with him. He believed in my work and spent a lot of time teaching me how to create valuable research, and for the dedicated time to review my work.

I would also like to thank my committee members: Drs. Michael McFarland, for his support and encouraging me during my time as a student; Gilberto Urroz, for helping me to understand the hydrodynamics of the lake in our meeting in 2011; Jack Schmidt, for his experience and valuable advice that changed the methodology of this research in sediment transport; and Thomas Hardy, for his effort during reviewing the research proposal.

I am grateful to Annette Deknijf from US Fish and Wildlife for helping me to access Bear Lake National Wildlife Refuge, the State of Idaho, and PacifiCorp.

Finally, I would like to express special thanks to my family and friends from Egypt and the United States.

Hussein Aly Batt
## CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ABSTRACT</td>
<td>i</td>
</tr>
<tr>
<td></td>
<td>PUBLIC ABSTRACT</td>
<td>iv</td>
</tr>
<tr>
<td></td>
<td>ACKNOWLEDGMENTS</td>
<td>vi</td>
</tr>
<tr>
<td></td>
<td>LIST OF TABLES</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>LIST OF FIGURES</td>
<td>xi</td>
</tr>
<tr>
<td></td>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.1 General Introduction</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.2 Research Motivation</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>1.3 Research Contribution</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>1.4 Organization of the Dissertation</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>6</td>
</tr>
<tr>
<td>2.</td>
<td>RELEVANCE VECTOR MACHINE MODELS OF SEDIMENT TRANSPORT IN A SHALLOW LAKE – I. DATA COLLECTION</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Abstract</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>2.1 Introduction to Problems Facing the Study of Sediment Fate</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2.1.1 Physics-Driven Methods</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2.1.2 Data-Driven Statistical Methods</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2.1.2.1 Artificial Neural Network (ANN)</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>2.1.2.2 Support Vector Machine (SVM)</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>2.1.2.3 Relevance Vector Machine</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>2.2 The Study Area and Challenges</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>2.3 Objectives and Experimental Design</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>2.3.1 Objectives</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>2.3.2 Experimental Design</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>2.4 Field Results and Discussion</td>
<td>18</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Operation scenarios used in Mud Lake as: Inflows and outflows in Mud Lake measured by PacifiCorp in Rainbow Canal, farmers irrigation canal, Causeway, and Lifton. The two colors of cells represent different operation scenario in the Lake.</td>
</tr>
<tr>
<td>2-2</td>
<td>Constituents percentile according to zones in Mud Lake, as a) DO mg/l, (b) pH, (c) Temperature °C, (d) Turbidity NTU, (e) TSS mg/l, (f) Velocity cm/s.</td>
</tr>
<tr>
<td>3-1</td>
<td>Observed, modeled Constituents percentile, and RMSE with respect to Zones in Mud Lake, as a) DO mg/l, (b) pH, (c) Temperature °C, (d) Turbidity NTU, (e) TSS mg/l, (f) Velocity cm/s.</td>
</tr>
<tr>
<td>4-1</td>
<td>Locations of RVs for different constituents, percentage of selected RVs from data, and percentage of required observation for successful modeling based on RVs.</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Mud Lake map as (a) Bathymetry (b) Sampling locations and zones (the dots mark the sample locations of change of hydraulics).</td>
</tr>
<tr>
<td>2-2</td>
<td>Box-whisker plots of collected observations in the thirty locations of Mud Lake during 2009-2010 as a) DO mg/l, (b) pH, Standard Units, (c) Temperature °C, (d) Turbidity NTU, (e) TSS mg/l, (f) Velocity cm/s.</td>
</tr>
<tr>
<td>2-3</td>
<td>Time series collected observations in the thirty locations of Mud Lake during 2009-2010 as a) DO mg/l, (b) pH, Standard Units, (c) Temperature °C, (d) Turbidity NTU, (e) TSS mg/l, (f) Velocity cm/s.</td>
</tr>
<tr>
<td>2-4</td>
<td>Water quality measures relationship as it occurs in Mud Lake. Each curve represents a different sampling date.</td>
</tr>
<tr>
<td>3-1</td>
<td>Box Plots of collected observations in the thirty locations of Mud Lake during 2009-2010 and Predicted MVRVM output as (a) pH, (b) DO mg/l, (c) TSS mg/l, (d) Turbidity NTU, (e)Temperature °C, (f) Velocity cm/s.</td>
</tr>
<tr>
<td>3-2</td>
<td>Residual plots of observations in the thirty locations of Mud Lake during 2009-2010 versus and Predicted MVRVM output as (a) pH, (b) DO mg/l, (c) TSS mg/l, (d) Turbidity NTU, (e)Temperature °C,(f) Velocity cm/s.</td>
</tr>
<tr>
<td>3-3</td>
<td>Scaled plots of observations in the thirty locations of Mud Lake during 2009-2010 versus and Predicted MVRVM output as (a) pH, (b) DO mg/l, (c) TSS mg/l, (d) Turbidity NTU, (e)Temperature °C,(f) Velocity cm/s.</td>
</tr>
<tr>
<td>4-1</td>
<td>Example of selected RVs in DO pattern, where the red points represent the RVs, black lines represent series which contain RVs, and colored lines represent series which contain no RVs.</td>
</tr>
<tr>
<td>4-1</td>
<td>Sample Locations in Mud Lake, and selected Relevant Locations from MVRVM model (the black lines represent time series, the red points represent the selected RV, the colored series represent series which don’t have RVs).</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

1.1 General Introduction

Wetlands are partially or fully covered with water during much of the year. Historically wetlands were thought of as wasted areas that can be utilized for agricultural activities; thus they were commonly drained [US EPA, 2010]. As a consequence of draining wetlands, various species became extinct (e.g. giant Mekong cat fish). As a matter of fact wetlands provide great benefits nationwide, they function as: 1-flood control in coastal areas by forcing water passing through it to slow and thus decrease damage to surrounding areas; 2-when wetlands are flooded with water, their roots and stems act as sediment traps providing a protection to downstream water bodies; 3-wetlands vegetation function as a giant kidney filtering water from impurities, chemicals and nutrients often attached to sediment; 4-wetlands provide vital habitat for many endangered species; about 35% of all animals and plants listed as endangered either depend on wetlands or live on wetlands; 5-land thick vegetation and associated invertebrate provides a source of food for fish and important rearing habitat for many species, as well as migration stations for other species; and 6-wetlands also provide recreational opportunities thus contribute to the economy of the country[US EPA, 2010].

[Deknijf, 2010] mentioned that the Mud Lake Unit is a part of Bear Lake National Wildlife Refuge located in the southern part of Idaho, which is managed by the US Fish and Wildlife Service[see Appendix C]. The refuge was established in 1968 for the purpose of protecting and managing the habitat of migratory birds. The refuge is a
nesting area for white faced ibis, herons, egrets, gulls, terns, grebes, ducks, and geese. Mud Lake consists of wetland and open water; Mud lake functions as: 1-sediment trap to water flowing from the Bear River, thus protecting the Bear Lake; 2-it provides a habitat to migrating species, and habitat for numerous resident species; 3-the vegetation acts as a filter for sediment flowing into Mud lake with nutrients; 4-irrigation source for farmers.

[Deknijf, 2010] raised concerns about issues needing to be addressed in the refuge as: 1-should the public access to Mud Lake be reduced or eliminated? 2-what are the best means to attain productive habitat for wildlife? 3-how can water quality of the refuge be improved?

The concerns of the refuge manager were one of the motives behind this study [Deknijf, 2010]. However these were not the only major challenges that were recognized in MudLake, which functions as a complex natural system with fine sediment flowing into it [see Appendix C]. As a consequence we are trying to identify an appropriate methodology for modeling the pattern (spatial and temporal) of suspended fine sediment transported in Mud Lake.

Well known techniques used to study sediment transport are: 1-indired methods for sediment transport based on grain size distributions (e.g. Einstein method, Yang method, and Toffaleti method) [Garcia, 2002] where the total sediment load is calculated through the bed load and suspended load function. 2-direct methods for determining total sediment load based on the stream power [Garcia, 2002].However as will be detailed below the suspended sediment flowing in our case is characterized by being very fine, which is difficult, both theoretically and practically, to model by the previously mentioned techniques.
In this research, the statistical learning tool MVRVM is considered as an alternative to the previously mentioned techniques. These statistical learning tools are data driven algorithms that rely on patterns in data to create a framework for prediction. However as we will demonstrate in the next chapter ANN, and SVM require large amounts of data. RVM is characterized by the ability to capture the hidden relationships among variables using only a few observations. Thus it is proposed as a modeling algorithm for this study, which will help also in evaluating options and making decisions.

1.2 Research Motivation

This research is motivated by a number of factors:

1- The effects of operation scenarios of Mud Lake should be considered since the habitat of some endangered species relies on the quality of water within MudLake and changes in operation may result in loss of habitat.

2- The spatial distribution of sediment in Mud Lake is not well understood. This uncertainty may contribute to filling MudLake with sediment in the near future [BLWAMM, 2009].

3- Mud Lake is used as a sediment trap for water flowing from the Bear River into Bear Lake; however studying the sediment circulating in different zones of MudLake and its concentration has not been reported.

4- Sediment carries various nutrients (e.g. phosphorous) which create a concern about its accumulation.

5- The concern related to the fluctuations of water quality constituents which can affect the ecosystem.
6- Given the complexity of Mud Lake, there is a need to develop a framework to predict the spatial fine suspended sediment distribution, and water quality constituents that can be used by the refuge managers to address their concerns.

7- A novel machine learning method the Multi Variate Relevance Vector Machine (MVRVM) has not been explored yet to model very fine sediment transport in estuaries and lakes.

1.3 Research Contribution

The proposed research is expected to contribute to the literature of the RVM and suspended fine sediment transport by using the MVRVM algorithm for the first time to model a complex natural system as:

- Model the spatial distribution of suspended fine sediment in a complex natural system using Mud Lake as a case study.
- Model selected water quality constituents in the natural system.
- Demonstrate how the RVs can help decision makers understand the practical problems of how much data are sufficient to support this class of model.

Success in the above mentioned tasks can spawn research that focuses on the spatial distribution and fate of contaminants attached to sediment particles, and the effect of operational scenarios on the dynamics of water quality constituent concentrations.

1.4 Organization of the Dissertation

The dissertation consists of five chapters. The first chapter is a general introduction to the problems in the Mud Lake Unit, as well as the concerns of the refuge
managers. This chapter also contains the motivation and contribution of this research to the literature.

In Chapter 2 we demonstrate the previous techniques that have been used to model suspended fine sediment transport and the limitations of these techniques. It will provide a brief review of the statistical learning tools and why the RVM is considered for the study. We will present the data collected in Mud Lake (for the different constituents e.g. DO, pH, Temperature, Turbidity, TSS) to serve the primary question addressed in the research: what is the driving force of the change of the dynamics of sediments and water quality constituents in Mud Lake, and the methodology that was used to select the sampling locations and how often the data were collected. We will show the different patterns of time series for the observations and their importance on water quality measures and model selection. This chapter also describes the main water quality constituents and the associated range for survival of key aquatic life, and whether the collected observations were outside this range. And thus raises a concern on creating a model to predict these constituents.

The Chapter 3 details development of the statistical learning tools, and basis for the selection of the MVRVM [Tipping, 2001]. We demonstrate in this chapter the results from the MVRVM model, the range of observations for each sample location, and tested water quality constituents, and their errors (prediction errors and not sampling errors) at these locations. We will discuss why the observed patterns in the data exist.

The Chapter 4 addresses whether the arrangement of the observations as a time series helped to resolve the issue of how much data are really needed to carry out in this class of modeling. We demonstrate the number of the selected RVs for each water quality
constituent tested, and the location for each of these RVs in the natural system and over time. We will explain how the significance of these selected locations affects the collection of observations for water quality constituents in Mud Lake.

Chapter 5 is an overall summary of and conclusions from the findings of this research and recommendations for future research in both Mud Lake and the sediment transport field.

This dissertation is formatted as a multi-paper dissertation format. Thus some repetition has occurred in the demonstration of the RVM algorithm structure.

References

Bear Lake Watch Annual Membership Meeting. (BLWAMM) (Last retrieved 31 March 2009)
http://www.bearlakewatch.com/membernotes/blwanmntg07.htm

Deknijf, A.(2010),Bear Lake National Wildlife Refuge Oxford Slough Waterfowl Production AreaPlanning Update Number 1, June 2010

US EPA (2010), Aquatic Biodiversity, Importance of Wetlands
http://www.epa.gov/bioiweb1/aquatic/importance.html,(last retrieved February 2012)


http://mi.eng.cam.ac.uk/~at315/MVRVM.htm
CHAPTER 2

RELEVANCE VECTOR MACHINE MODELS OF VERY FINE SEDIMENT TRANSPORT IN A SHALLOW LAKE – I. DATA COLLECTION

Abstract

Mud Lake is a part of wildlife refuge located in the southern part of Idaho and is operated by PacifiCorp. Mud Lake is used as a sediment trap for the water flowing from the Bear River into Bear Lake; however the model development for predicting different seasonal operations and its effect on sediment circulating in different zones of Mud Lake and its concentration has not been accomplished yet. These reasons combined with the facts: 1-some parts of Mud Lake might fill up with sediments, 2-sediment carries various nutrients (phosphorous) which create a concern about its accumulation, and 3-the system has complicated hydrodynamics and biological characteristics were the motive behind this study which will be based on developing a model to predict suspended fine sediment and its spatial distribution.

In this study we propose the RVM as an approach for predicting the total suspended solids, water quality measures and sediment transport within Mud Lake. In Chapter 2, we describe an experimental design for the collection of turbidity, total suspended solids, hydraulic parameters, hydrodynamic model output as inputs to select the relevant parameters for the RVM model; and determine whether these data have characteristics that lend themselves to modeling with RVM. In Chapter 3, we describe the RVM approach and apply it to the data from Mud Lake.
2.1 Introduction to Problems Facing the Study of Sediment Transport

Management of sediment and its effect on water quality and habitat is an important problem for resource managers. Tools for predicting sediment dynamics with accuracy required for management are limited and often require large amounts of data that are typically not available for most water bodies. New tools are needed for accurate assessment of sediment dynamics with limited data.

Sediment transport and deposition in water bodies depends on factors such as flow, velocity, and sediment characteristics [Garcia, 2002]. Two categories of sediment are transported inflowing water: 1- bed load, consisting of larger particles eroded from the water body’s bed and 2- wash load, consisting of fine material coming from the banks, the watershed, overland flow, and bed [Garcia, 2002]. When a stream approaches a relatively quiescent water body, such as a lake or estuary, flow characteristics generally change as the gradient decreases and the stream widens and deepens [Garcia, 2002]. The increase in cross sectional area and decrease in flow velocity often result in significant amounts of sediment deposition.

2.1.1 Physics Driven Methods

Recent research concerning transport of various grain size classes has focused mainly on the hydrodynamic conditions of the rivers; where the transport potential of sediment sizes is based on momentum balances on one grain size or grain size distributions (e.g. Einstein method, Yang method, and Toffaleti method) [Garcia, 2002]. The use of the physics driven models requires complete information about grain size distribution, sediment density, fluid properties, and hydraulic conditions; however this use
of particle size is unsatisfactory in fine sediment-dominated systems because models generally perform poorly for very fine sediment sizes [Jain, 2001; Nagy, 2002; Sen, 2004].

Other calculation techniques for sediment load include: 1-Reservoir survey: this technique is used by calculating the volumetric sediment deposit in the reservoir and continuous monitoring of the sediment discharged into the system [Odhiambo and Boss, 2004]; 2- Fluvial Data [Guy and Norman, 1970]: A sediment rating curve is computed over a range of discharges to relate the sediment concentration and the flow. The resulting relationship usually exhibits scatter varying over two orders of magnitude at a given discharge. The use of sediment discharge rating curves can be applied if the data used by the curves were collected over years through several flood events, a major limitation in many systems.

2.1.2 Data Driven (Statistical) Methods

Statistical learning tools, developed recently to represent complex patterns in data by assemblies of simple functions, have been used to estimate sediment concentration in water bodies by a number of researchers [Dogan et al., 2007; Jain, 2001; Nagy, 2002], and combining these estimates with flow data to produce estimates of the sediment yield. Three of these methods are artificial neural networks support vector machines, and multivariate relevance vector machines (MVRVM).
2.1.2.1 Artificial Neural Network (ANN)

Neural networks were developed by analogy to the human brain in which complexity is achieved through selective interconnection of larger numbers of neurons that modify inputs to produce desired outputs.

The neural network consists of an input layer, a hidden layer or layers, and an output layer. The input layer, representing the collection of data, simply passes inputs to the hidden layer for processing. The hidden middle layer (or layers) is where the neurons are interconnected and assigned weights to control the passing signal. The ANN algorithm modifies the network by adjusting the weights in the hidden layer to alter the outputs. Finally, the output layer simply collects the modified inputs to produce results for comparison with observations. The ANN is characterized by fast computational time compared with physics based models.

In the past few decades ANNs, such as the multilayer back propagation neural network, have been widely used in various applications concerned with the study of sediment transport, and estimation of sediment load. However there are disadvantages to the use of the ANN algorithm, namely that traditional ANNs can get trapped in local minima, suggesting that the ANN may not be producing unique results. [Doganet al., 2007] mentioned that because of this disadvantage newer training algorithms have been developed. One of these algorithms is support vector machines (SVM).

2.1.2.2 Support Vector Machine (SVM)

A second statistical learning approach, SVMs, is used for classification and regression. The SVM algorithm is based on separation between levels of input values;
thus if the levels are distinct, the SVM selects a model that minimizes the error; by locating data groups or support vectors that maximize the gaps between data levels. Where the data levels are non-distinct the SVM tries to find the plane that maximizes the data level gaps while minimizing the error. This can be achieved by projecting the inputs into a higher dimensional feature space to formulate a linear classification.

2.1.2.3 Relevance Vector Machine

Vapnik[1995] and Tipping[2001] suggested that the SVM suffers from limitations: 1- SVM makes excessive use of the kernel function requiring the number of required observations to grow for training the model; 2- estimation of error/margin parameter using cross-validation is an extra step in the analysis. He introduced the MVRVM as a new algorithm based on a Bayesian approach that does not suffer from the limitations of SVM and requires many fewer kernel functions. The MVRVM algorithm adopts a Bayesian learning technique, which introduces a probabilistic framework for the selection of important information used for training the algorithm.

Developed for pattern recognition, the MVRVM is a relatively new approach that has not been used widely in modeling suspended fine sediment that occurs in many natural environmental systems [Dogan et al., 2007]. Given its inherent ability to recognize patterns and the presence of consistent patterns in space and time for water quality constituents and suspended fine sediment, the MVRVM was considered as an approach to model water quality constituents, and fine suspended sediment patterns in a wetland lake in southeast Idaho, USA.
The MVRVM was considered in this research for using the patterns of velocity, turbidity and water quality as an alternative to physics based models to address a practical problem of designing an efficient monitoring system for suspended fine sediment, and water quality constituents to better serve the management objectives.

2.2 The Study Area and Challenges

The study area, Mud Lake, with a surface area of approximately 20 km$^2$, is located three miles north of the towns of Montpelier and Paris in southeast Idaho [Figure 2-1] and, since 1911, has served as a sediment trap and filter for waters from the Bear River flowing into Bear Lake, immediately to the south, as well as a refuge for migratory birds. Prior to 1911, the flow to the lake consisted mainly of overflows from Bear Lake and surrounding creeks [BLWAMM, 2005]. The quality of nesting habitat for ducks and waterfowl has been observed to be in an inverse relationship with the turbidity in Mud Lake which affects the vegetation growth that is used for nesting and source of food [Bjornn, 1989]. In 1911, Bear Lake was converted to a storage reservoir for flows from the Bear River, turning Mud Lake into a conveyance area between the sediment-laden Bear River and relatively sediment-free Bear Lake. Currently, when water is required for irrigation at the end of the summer, water is pumped from Bear Lake into Mud Lake, and then conveyed via Paris Dike to farmers [see Figure 2-1] [BLWAMM, 2005].

Although Mud Lake has never been assessed for the US EPA 303(d) list, this situation might change in the next few years without proper flow and sedimentation management and control [ADIMLW, 2002]. Mud Lake and Dingle Marsh (just north of Paris Dike, see Figure 2-1) became a part of the U.S. National Wildlife Refuge System in
1986. Accordingly, parts of the marsh were dredged in order to better control water levels, and enhance conveyance of water [BLWAMM, 2005]. The Causeway that separates Mud Lake and Bear Lake was accidentally breached in 1993, resulting in large amounts of sediment and sediment-associated nutrients being transported to Bear Lake, with the potential to affect water quality [BLWAMM, 2005].

The current flow management strategy in Mud Lake is controlled for hydropower production. During spring runoff the flow is diverted from the Bear River into the Rainbow Canal and Mud Lake. At the end of the summer when river flow is not sufficient for irrigation purposes the water is pumped from Bear Lake back into Mud Lake from which it flows into a discharge canal and back into the Bear River or into irrigation canals. This flow management strategy contributes to movement of sediment through the system but details of the fate of sediment as it flows through the parts of Mud Lake are unknown.

The complexity of flow management and the spatial distribution of suspended fine sediments in light of the complex hydrodynamics within Mud Lake, or any similar shallow impoundment, require attention to spatial and temporal patterns in the data. Zicari[2010] notes that “patterns of data modeling are very important. They enable data modeling efforts to be both effective and efficient. Working without patterns is like wandering around in the data wilderness trying to find your way.”

2.3 Objectives and Experimental Design

2.3.1 Objectives

The operation of the lake produces the patterns required for the MVRVM model to predict the proposed modeled constituents. Thus, the objectives of this research are to:

1. Evaluate the use of an MVRVM modeling approach and assess its efficacy in a complex hydraulic system with limited observations to model suspended fine sediment and water quality constituents.
2. Expand the use of the MVRVM for the hydraulics and suspended fine sediment dynamics without the need of developing physics-based models.
3. Verify that the collected data embody patterns that are required for effective use of the MVRVM.
4. Demonstrate that the data from the 30 locations are sufficient to capture the needed patterns resulting from operational time-based management of the lake [see Appendix A].

In this study, we describe an experimental program for the collection of topographic and hydraulic data, turbidity, total suspended solids and other water quality constituent data, and the production of hydrodynamic model output, as inputs to select the relevant MVRVM model parameters. A second chapter provides model details and demonstrates the use of the MVRVM for the Mud Lake case study.

However, it is possible to foresee several challenges in studying Mud Lake and similar systems:

1. The flow in Mud Lake is controlled by a private third party where the flow is estimated based on daily average of recorded gate flow.
2. Absence of velocity data inventory- such data allow identifying ways the sediment is flowing and potentially depositing into different locations of the lake.

3. Flow management - the effects of changing flow management strategy should be considered since the habitat of ducks and waterfowl relies on the quality of water within Mud Lake and changes in operation can result of loss of habitat [Bjornn, 1989].

4. Sediment distribution - the spatial distribution fine sediment deposited in Mud Lake is unknown. This may contribute to filling some parts of the lake with sediment in the near future and limiting its usefulness as a refuge [BLWAMM, 2005].

5. The flow management in the lake by the third party is the driving force for patterns of all constituents in the Lake.

To meet the challenges created by complex natural system and achieve the objectives of our study the following steps were taken:

1. **Select monitoring locations:** Preliminary sampling revealed that the changes of hydraulic properties in the lake could be modeled by creating sampling locations that can track all the constituents’ range of change. Thirty locations are considered for modeling and sampling. These locations are proposed to be sufficient to track the suspended fine sediment circulating in Mud Lake.

2. **Identification of patterns:** Patterns will be sought for the water quality measures and suspended fine sediment at all the selected locations.
3. **Examine flow scenarios**: Study of various flow scenarios combined with the transported suspended sediment will give a better understanding of the amount of sediment and water quality constituents in different locations in Mud Lake.

4. **Create velocity inputs**: A variety of flow scenarios are modeled to understand how the water flow is routed through Mud Lake. The MVRVM model required flow velocities at each of the 30 locations, however, the collection of velocity vector observations was found to time consuming compared to collection of other water quality constituents, and, moreover, most of the observations were below detection limit, thus we used a mechanistic two dimensional hydrodynamic model, CCHE2D [Zhang, 2001], to provide estimates of velocity vectors for each location to use as input for the MVRVM model. The collection for 30 locations over 2 days showed that the observation were above the detection limit only in 7 locations.

### 2.3.2 Experimental design

The study aims to collect and evaluate hydraulic, sediment-related, and water quality data to support the MVRVM modeling approach in the shallow, natural but highly managed Mud Lake. The MVRVM approach has essentially two requirements. First the data must be of a nature to capture the nuances of sediment and other water quality constituents; and second the data must be available in sufficient quantities that the MVRVM algorithm can discover those data that capture model-relevant information (relevant vectors) but exclude data that have little information content or are highly correlated with data that are retained by the MVRVM.
Special metadata (geographic data, and field data) were collected to clarify the nature of Mud Lake, in addition to the sediment and water quality data to construct the MVRVM model.

2.3.2.1 Geographic and field data.

1- GIS coordinates for identifying the sampling locations within the lake. The data from these locations will serve in both the training and the predictions step of the MVRVM model.

2- Mud Lake boundaries using aerial photographs to assist in identifying various features within Mud Lake, such as inflows, vegetation location, and water pathways.

3- Bathymetry and cross-section data using a Garmin Echo 500c ultrasonic depth sounder to be used as input to the development of the hydraulic modeling. (Star Marine Depot Inc, Coral Springs, FL).

4- Flow and velocity profile observation data were collected for the thirty locations twice during summer 2009 [FLO-MATE Model 2000 Flowline Manufacturing Ltd, UK]. These data are used in calibrating the hydrodynamic model to generate velocity vector estimates for each location. The importance of these data is because flow and other hydraulic data were not available in real time or historically.

2.3.2.2 Sediment/ water quality

The study of the fine sediment transport phenomena requires the study of the suspended load data. These data were obtained by weekly sampling during the ice-free periods of 2009 and 2010. Total suspended solids (TSS) samples were collected at mid-depth using a Niskin type sampler (General Oceanic Inc., Miami, FL), and TSS
concentrations were measured at the Utah Water Research Laboratory following the procedure based on the US EPA Method 160.2 at all 30 sampling locations and Standard Methods 2540.D to determine the TSS concentration. A Hydrolab Sonde series MS 5 (Hach Company, Loveland, CO) was used to measure the water quality constituents: turbidity, pH, temperature, and dissolved oxygen [see Appendix B]. All data were recorded in a field book, and then transferred to and saved in Microsoft Excel spreadsheets (Microsoft Inc., Redmond, WA). Statistical analysis was carried out using the R statistics package [Ihaka and Gentleman, 1996]. For quality control purposes we collected duplicate TSS samples for 3 random locations during every field trip.

2.4 Field Results and Discussion

The bathymetry data [Figure 2-1b] (used for developing the hydrodynamic model) revealed that Mud Lake depth ranged between 0.5 to 2 meters. The study area is characterized by complicated hydraulics. The flow data [see Table 2-1] reveals that there are 2 main operating scenarios. The first is during the spring runoff season during which the water flows from the Bear River via Rainbow Canal [see Figure 2-1] through the lake and sometimes continuing into Bear Lake. The second is during the late summer period when water is pumped from Bear Lake back into Mud Lake to satisfy downstream irrigation requirements when the Bear River flow is not high enough to meet irrigation demand.

2.4.1 Assessment of water quality in Mud Lake

**Dissolved Oxygen (DO):** The dissolved oxygen minimum for most aquatic life is 5 to 6 mg/l [DEQ, 2011]. Constant oxygen concentration in water is optimal for survival.
Fluctuations of DO levels can result from aquatic vegetation photosynthesis and respiration. Systems can lose DO due to decomposition of organic matter by bacteria and chemical reactions consuming oxygen, and low levels of DO can stress aquatic organisms and cause mortality [CEES, 2005]. The levels of DO within Mud Lake were generally above the level for aquatic life; however, a small number of observations were recorded below the minimum in station 10 through 13 in Zone 1 where Bear river flows into Mud Lake which raise questions of whether Mud Lake is fully supporting aquatic life.

**pH**: The pH controls many chemical and biological processes in aquatic systems. The survival of aquatic organisms and health of an ecosystem require a specific range of pH (6.5 to 9.0 [DEQ, 2011]) because aquatic organisms are tolerant to a very narrow range around neutral pH. High pH can result from algae and aquatic vegetation using CO$_2$ for photosynthesis. Low pH, caused by respiration of the same organisms, can mobilize many toxic chemicals, particularly heavy metals, to become available for uptake by aquatic plants and animals, creating the potential for toxic conditions for aquatic life [CEES, 2005]. Although the pH in Mud Lake was generally between 6.5 and 9, occasionally the pH fluctuated outside that range especially in Zone 3 where the water is clear and plant density is high.

**Water Temperature**: Aquatic organisms are adapted to specific temperature ranges for growth and survival; when temperature is outside the range for prolonged periods of time it can cause stress or death for aquatic organisms [CEES, 2005]. Water temperature also influences the dissolved oxygen saturation available in water that strongly affects many aquatic organisms. The Bull Trout standard for the state of Idaho is...
9 degree Celsius during spawning time [DEQ, 2011]). The temperature range in Mud Lake varied from freezing to 25 °C [see Figure 2-3] shows that during April the temperature conditions are adequate for the Bull Trout spawning at all stations.

**Turbidity and sediment:** Increased sediment transport into water bodies has an impact on quality of ecosystem required for the survival of inhabiting species [Schubel, 1977]. Turbidity can reduce the amount of light entering the water column thus decrease photosynthesis from aquatic plants. Nutrients, particularly phosphorus, can adsorb onto sediment particles [Bjornn, 1989], thus the spatial distribution of fine sediment is of great importance to determine the fate of nutrients [Schubel, 1977]. Sediment can disrupt food production dynamics through decreased predator success with respect to prey survival. [Moore, 1977; Simenstad, 1990; Coen, 1995]. In the case of Mud Lake the turbidity ranged from 0 NTU to 357 NTU (DEQ guidelines for turbidity are given as the background turbidity + 50 NTU, [DEQ, 2011]). [Figure 2-3] shows that from October through the beginning of the runoff season the turbidity is within the DEQ acceptable range, while during the rest of the year the turbidity was often above 50 NTU. The analysis of the TSS duplicate samples showed that pooled standard deviation was 0.5%.

**Velocity vector magnitude:** Velocity is the driving force responsible for determining the fate and transport of sediment through the lake. Sediment may also carry attached nutrients and could affect the levels of DO by transporting suspended fine sediment that may inhibit photosynthesis; pH, modified by growth of algae caused by nutrients may also be affected. Because of their small magnitude, velocities were not as thoroughly quantified as the water quality constituents, the results from the hydrodynamic model [Figure 2-3] showed that more than 90% of the velocity magnitudes
are below the detection limit However, the velocity magnitudes were higher in Zone 1 and decreased to near zero through Zone 3.

However collection of velocity observations proved difficult due to the small magnitude of the observed velocities, which were usually below the detection limit of the velocity instrumentation. As an alternative, the flow direction was observed indirectly in the field trips by noting the direction of the wake around on the current meter rod inserted in the flow stream.

2.4.2 Suitability of observations to the RVM

On the basis of our preliminary field observations, it was determined [see Figure 2-1] that Mud Lake could be logically divided into three zones, based on location and on observed hydraulic and water quality similarities, as mentioned previously. Observations at some locations may be expected to be different than at other locations, due to ecological conditions observed during data collection that might affect the modeling results. Some locations, especially in Zone 3, had aquatic vegetation and algae that can affect the DO and pH observations and, since the observations reported here were taken during daylight, the DO and pH might be expected to be somewhat higher than the average over the day. Also, at times, local rain during the time surrounding the sampling event at some Mud Lake locations could dilute the constituents in the lake water and change the collected data significantly.

Zone 1 is characterized as a network of canals and provides the source for sediment into Mud Lake. Some parts of this zone are very shallow due to historical sediment deposition; during sampling we observed that channels have been dredged,
presumably to promote flow from the Rainbow Canal toward the irrigation canal and to distribute flow and sediment through all of Mud Lake [Bjornn, 1989]. This zone is clear of vegetation in its water ways but vegetated between channels.

Zone 2 is near the center of Mud Lake with patchy vegetation; transition between turbid and clear water takes place in this zone. The vegetation is low density rooted vegetation (approximately 0.2 meter height), which may affect the pH and DO observations.

Zone 3 is filled with rooted vegetation, often emergent, (>approximately 0.2 m height) at the majority of locations. The DO and pH increase significantly relative to Zones 1 and 2, likely due to photosynthetic activity during the daytime sampling. Zone 3 is characterized by extremely low velocities especially in the east; thus providing an opportunity for the fine sediment to settle, and leading to low TSS and turbidity.

Pattern type I [see Figure 2-2 and 2-3] Temperature is characterized consistent percentile ranges for observations at all the locations: the temperature pattern didn’t vary significantly from Zone 1 to Zone 3.

Pattern type II [see Figure 2-2 and 2-3] pH and DO are characterized by change in the percentile range of observations across locations, especially in the third Zone 3 where there is an increase of vegetation and algae; which increases the pH and DO observations relative to Zone 2. This was expected due to the presence of small amounts of vegetation in Zone 2 compared with Zone 3 where the vegetation was denser.

Pattern type III [see Figure 2-2 and 2-3](described by the Turbidity, TSS, and velocity magnitude) is described by a variable range of change for percentiles in Zones 1 and 2 and then drop to near zero in the Zone 3. The turbidity is related to the TSS; however the
accuracy of measuring turbidity is higher. TSS is heterogeneous and during collection of grab samples we might capture clumps of sediment in either the sample or the smaller portion taken for filtering, which does not represent the whole water. These problems lead to errors in observing TSS as high as 10-20%.

The observations [see Figures 2-2 and 2-3] support the presence of repeated patterns (the observations in every location with respect to time acts in an arrangement of repeated features) for all locations. These patterns follow the water quality measures (an example of this overall pattern is illustrated in [see Figure 2-4]), thus the time series were used to represent the data during the ice-free periods for the three zones [see Figure 2-3]. The general similarity among observations in each zone suggests that the MVRVM model, for which a subset of the observations is selected for model fitting, should have sufficient data for accurate representation of the variables at each sampling location.

2.5 Conclusion

The review of previous modeling efforts used for sediment transport emphasizes the conclusion that these models are less suitable to simulate the spatial sediment distribution because of the fine sediment nature of Mud Lake and low in-lake velocity. It was also demonstrated that the data collected embody patterns that can be useful for the use of the MVRVM, to capture the patterns resulting from time based operation and management of the lake. The operation of the lake caused the presence of these patterns for the proposed modeled constituents.

The collected observations show distinct patterns with respect to the flow conditions at each location and also it reveals that some of the observations were outside
the range designated by the Idaho Department of Environmental Quality that supports aquatic growth \cite{DEQ, 2011}. Even though we have large number of observations, experience has shown they are insufficient to support use of ANN or SVM models. We hypothesize that the data set is more than adequate for modeling using the MVRVM \cite{Batt and Stevens, 2012}. As will be seen in the next chapter, the MVRVM shows promise in modeling systems with complex patterns not only in hydraulics but also for modeling water quality constituents.

References


Center of Earth and Earth Science (2005), Use of long-term research for enhancing water quality in the Great Lakes Region. (Last retrieved February 2012)[http://www.ceeds.iupui.edu/education/Workshops/Project_Seam/water_quality.htm]

Coen, L. D. (1995), A review of the potential impacts of mechanical harvesting on subtidal and intertidal shellfish resources. Marine Resources Research Institute, SC Department of Natural Resources, Charleston, S.C.


Table 2-1. *Operation scenarios used in Mud Lake as:* Inflows and outflows in Mud Lake measured by PacifiCorp in Rainbow Canal, farmers irrigation canal, Causeway, and Lifton. The two colors of cells represent different operation scenario in the Lake.

<table>
<thead>
<tr>
<th>Day</th>
<th>Rainbow</th>
<th>Irrig. farmers</th>
<th>Causeway</th>
<th>Lifton</th>
</tr>
</thead>
<tbody>
<tr>
<td>6/3/2009</td>
<td>919</td>
<td>5</td>
<td>983</td>
<td>0</td>
</tr>
<tr>
<td>7/13/2009</td>
<td>800</td>
<td>986</td>
<td>0</td>
<td>360</td>
</tr>
<tr>
<td>7/30/2009</td>
<td>168</td>
<td>645</td>
<td>0</td>
<td>675</td>
</tr>
<tr>
<td>8/11/2009</td>
<td>144</td>
<td>424</td>
<td>0</td>
<td>360</td>
</tr>
<tr>
<td>8/14/2009</td>
<td>144</td>
<td>553</td>
<td>720</td>
<td>0</td>
</tr>
<tr>
<td>8/19/2009</td>
<td>207</td>
<td>571</td>
<td>0</td>
<td>345</td>
</tr>
<tr>
<td>9/2/2009</td>
<td>152</td>
<td>406</td>
<td>0</td>
<td>340</td>
</tr>
<tr>
<td>9/10/2009</td>
<td>140</td>
<td>328</td>
<td>0</td>
<td>340</td>
</tr>
<tr>
<td>9/19/2009</td>
<td>64</td>
<td>84</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10/23/2009</td>
<td>220</td>
<td>5</td>
<td>131</td>
<td>0</td>
</tr>
<tr>
<td>10/29/2009</td>
<td>185</td>
<td>5</td>
<td>131</td>
<td>0</td>
</tr>
<tr>
<td>11/2/2009</td>
<td>255</td>
<td>744</td>
<td>0</td>
<td>426</td>
</tr>
<tr>
<td>11/5/2009</td>
<td>256</td>
<td>194</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11/12/2009</td>
<td>240</td>
<td>5</td>
<td>219</td>
<td>0</td>
</tr>
<tr>
<td>4/29/2010</td>
<td>200</td>
<td>5</td>
<td>219</td>
<td>0</td>
</tr>
<tr>
<td>5/19/2010</td>
<td>50</td>
<td>5</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td>5/21/2010</td>
<td>135</td>
<td>5</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td>5/28/2010</td>
<td>90</td>
<td>5</td>
<td>132</td>
<td>0</td>
</tr>
<tr>
<td>6/4/2010</td>
<td>160</td>
<td>5</td>
<td>132</td>
<td>0</td>
</tr>
<tr>
<td>6/11/2010</td>
<td>950</td>
<td>5</td>
<td>967</td>
<td>0</td>
</tr>
<tr>
<td>6/18/2010</td>
<td>2050</td>
<td>5</td>
<td>2050</td>
<td>0</td>
</tr>
<tr>
<td>6/28/2010</td>
<td>857</td>
<td>1156</td>
<td>0</td>
<td>378</td>
</tr>
<tr>
<td>7/9/2010</td>
<td>485</td>
<td>708</td>
<td>0</td>
<td>1060</td>
</tr>
</tbody>
</table>

Table 2-2. *Constituents percentile according to zones in Mud Lake, as* (a) DO mg/l, (b) pH, (c) Temperature °C, (d) Turbidity NTU, (e) TSS mg/l, (f) Velocity cm/s

<table>
<thead>
<tr>
<th>Zone</th>
<th>Min</th>
<th>25 Percentile</th>
<th>50 Percentile</th>
<th>75 Percentile</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zone 1</td>
<td>4.92</td>
<td>7.44</td>
<td>8.11</td>
<td>9.33</td>
<td>11.48</td>
</tr>
<tr>
<td>Zone 2</td>
<td>4.48</td>
<td>6.81</td>
<td>7.88</td>
<td>9.35</td>
<td>11.19</td>
</tr>
<tr>
<td>Zone 3</td>
<td>5.95</td>
<td>8.21</td>
<td>9.01</td>
<td>10.37</td>
<td>14.63</td>
</tr>
<tr>
<td>Zone 1</td>
<td>6.35</td>
<td>7.98</td>
<td>8.12</td>
<td>8.26</td>
<td>8.95</td>
</tr>
<tr>
<td>Zone 2</td>
<td>7.74</td>
<td>8.17</td>
<td>8.33</td>
<td>8.56</td>
<td>9.12</td>
</tr>
<tr>
<td>Zone 3</td>
<td>5.22</td>
<td>8.32</td>
<td>8.62</td>
<td>9.07</td>
<td>10.18</td>
</tr>
<tr>
<td>Zone 1</td>
<td>1.46</td>
<td>7.20</td>
<td>14.27</td>
<td>17.93</td>
<td>23.81</td>
</tr>
<tr>
<td>Zone 2</td>
<td>1.41</td>
<td>7.91</td>
<td>14.88</td>
<td>18.82</td>
<td>25.01</td>
</tr>
<tr>
<td>Zone 3</td>
<td>0.97</td>
<td>8.19</td>
<td>15.07</td>
<td>19.28</td>
<td>25.56</td>
</tr>
<tr>
<td>Zone 1</td>
<td>0</td>
<td>14.67</td>
<td>27.67</td>
<td>52.00</td>
<td>170.67</td>
</tr>
<tr>
<td>Zone 2</td>
<td>0</td>
<td>10.67</td>
<td>17.00</td>
<td>34.33</td>
<td>108.67</td>
</tr>
<tr>
<td>Zone 3</td>
<td>0</td>
<td>3.33</td>
<td>6.00</td>
<td>10.67</td>
<td>100.00</td>
</tr>
<tr>
<td>Zone 1</td>
<td>0</td>
<td>19.38</td>
<td>49.75</td>
<td>94.55</td>
<td>357.30</td>
</tr>
<tr>
<td>Zone 2</td>
<td>0</td>
<td>13.50</td>
<td>29.15</td>
<td>60.20</td>
<td>175.30</td>
</tr>
<tr>
<td>Zone 3</td>
<td>0</td>
<td>0</td>
<td>2.95</td>
<td>14.48</td>
<td>78.40</td>
</tr>
<tr>
<td>Zone 1</td>
<td>0</td>
<td>1.72</td>
<td>5.22</td>
<td>10.73</td>
<td>55.11</td>
</tr>
<tr>
<td>Zone 2</td>
<td>0</td>
<td>0.07</td>
<td>0.40</td>
<td>3.14</td>
<td>91.64</td>
</tr>
<tr>
<td>Zone 3</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>2.22</td>
<td>156.66</td>
</tr>
</tbody>
</table>
Figure 2-1. Mud Lake map as (a) Bathymetry (b) Sampling locations and zones (the dots mark the sample locations of change of hydraulics).
Figure 2-2. Box-whisker plots of collected observations in the thirty locations of Mud Lake during 2009-2010 as (a) DO mg/l, (b) pH, Standard Units, (c) Temperature °C, (d) Turbidity NTU, (e) TSS mg/l, (f) Velocity cm/s.
Figure 2-3. Time series observations in the thirty locations of Mud Lake during 2009-2010 as (a) DO mg/l, (b) pH, Standard Units, (c) Temperature °C, (d) Turbidity NTU, (e) TSS mg/l, (f) Velocity cm/s.
Figure 2-4. Water quality measures relationship as it occurs in Mud Lake. Each curve represents a different sampling date.
CHAPTER 3

CAN SUSPENDED FINE SEDIMENT TRANSPORT IN SHALLOW LAKES BE PREDICTED USING MVRVM WITH LIMITED OBSERVATIONS?

Abstract

The study of sediment transport in water natural bodies is a challenging task. There have been several attempts to describe sediment mathematically using hydraulic characteristics of water bodies. Most researchers who developed empirical formulas to describe sediment transport performed laboratory experiments with assumptions that didn’t take into account variation of hydraulic parameters, and fine sediment sizes that is part of this phenomenon. Recently, new approaches for studying sediment transport have been developed involving the use of machine learning algorithms that have proven accuracy and efficiency in predicting sediment transport.

A novel machine learning method the Multivariate Relevance Vector Machine (MVRVM) has yet to be tested to model sediment transport in estuaries and lakes. The selection of the MVRVM is due to the presence of very limited field observations which present a challenge to use other Statistical learning machines. This Chapter tests the success of calibrating the MVRVM model to predict suspended fine sediment transport and other environmental measures in the Mud Lake. We demonstrate the training and prediction results of turbidity, total suspended solids, pH, DO, and water temperature and whether these results would support the use of the MVRVM or not.
3.1 Introduction

The amount of sediment carried by river flow or is deposited in a water body depends on several factors such as flow rate and sediment characteristics. Two types of sediment are transported by flow: 1- bed load eroded from the water body’s bed and 2- the wash load consisting of fine material coming from the banks, the watershed, overload flow, and bed. When a stream approaches a lake or estuary, flow characteristics change. The sudden increase in cross sectional area and decrease in flow velocity often result in a significant amount of sediment deposition. The amount of sediment transport into and out of a lake is related to management requirements and beneficial use of the lake which might not have taken into consideration the dead storage occupied by the sediment.

Over the past few decades research concerning transport of various grain size classes in rivers focused mainly on the hydrodynamic conditions; where the transport potential of sediment sizes is based on various formulas that use one grain size or a distribution of grain sizes (e.g. Einstein method, Yang method, and Toffaleti method) [Garcia, 2002]. Thus the sediment sizes are an important factor in selecting, or creating a model. However this use of particle size is considered difficult in shallow lakes and natural systems because: 1-the recent models developed in the last few decades perform poorly in terms of the very fine sediment sizes that dominate natural systems [Jain, 2001; Nagy et al., 2002; Senand Altunkaynak, 2004]. Physics-based sediment transport models require detailed information about the temporally and spatially variable physical characteristics of the sediment, 2-Alternative modeling approaches using recent advances in statistical learning theory show promise in providing predictive capability in such cases, 3-often sediment water quality criteria are expressed as the more easily measured
turbidity (NTU), models that address sediment particles directly will be less useful for management decision making.

As a result, developing a model to describe the turbidity associated with fine sediments discourages consideration of physics-based models which predict sediment transport based on sediment physical characteristics rather than indirect measures such as turbidity.

3.2 DataDriven (Statistical) Methods

Statistical learning tools algorithms have been used to estimate the sediment concentration in different water bodies, and combining these estimates with flow data produce estimates of the sediment yield. Other studies have examined statistical learning tools to predict the levels of other water quality constituents. Three of these methods are: artificial neural networks (ANN), support vector machines (SVM), and relevance vector machines (RVM).

3.2.1 Artificial Neural Networks (ANN)

An artificial neural network is a mathematical representation of the human brain, and contains billions of neurons that function to recognize patterns, and process data. Like the brain, the ANN algorithm is able to adapt as new data become available and process information which makes it useful in prediction of sedimentation transport.

Recent studies related to using ANN in sediment transport have shown success compared with traditional, physics-based methods[Jain, 2001] mentioned that it is difficult in all cases to find the conventional sediment rating curves sufficiently reliable to correctly estimate the mass of sediment transported by rivers. He proposed the use of
the three layer feed forward ANN to create and study sediment rating curves. He created integrated stage discharge sediment concentration relations for two gauging sites to make a comparison with ANN and conventional curve fitting approaches for predicting suspended sediment concentration. The artificial neural network showed better results for both of the two gauging sites with a one order of magnitude-reduction in the sum of the squared errors compared to the conventional curve fitting approach.

Sen and Altunkaynak, [2004] concluded that the different sediment prediction models in practice, which were developed from rational formulations, suffer from having their parameters estimated using regression methods from a single historical data set. He coupled the ANN with Kalman filtering to model discharge and sediment concentration for the Mississippi River Basin. The resulting statistical analysis of his study showed that this approach improved the prediction, reducing the residual sum of squares by 50% for the loading compared to the regression methods.

Partal [2008] proposed a different ANN approach to accurately predict the suspended sediment loads in streams. His study was divided into two parts: 1- to predict sediment load using past data, and 2- to predict the sediment load using daily river flow measurements. He coupled normal techniques for forming the ANN with wavelet methods (methods that use periodic functions to help capture patterns of data.), and mentioned that the input for this model was selected by applying the wavelet components. These components helped in deciding the parameters which have large effects on the sediment load. From the output it was proven that the coupled wavelet with ANN provided a good fit to the observations for the testing period. The results of this
research were compared to traditional ANN, to show that the wavelet-ANN approach had superior predictions in all cases including the peak estimation of sediment values.

Despite the advantage of using the ANN indicated above there remains a major disadvantage to this algorithm, namely that traditional ANNs can get trapped in local minima, suggesting that the “best” ANN models are not unique. Dogan et al. [2007] mentioned that because of this disadvantage newer statistical learning algorithms have been developed.

3.2.2 Support Vector Machine (SVM)

SVMs are derived from statistical learning theory and have been used for classification and regression. The SVM algorithm is based on separation between input data classes to select subsets of training data that contain important information to be used for testing.

Singh et al. [2008] focused on estimation of discharge, and normal depth in a trapezoidal channel having various bottom slopes using SVM. He used data from the literature collected by Ahmad [2001] and empirical relations suggested by Ahmad [2001] and Gupta et al. [1993]. He found that the correlation coefficient to evaluate the efficiency of settling basins for all the bed slopes was higher than 0.995 for both prediction of discharge and end depth. He suggested the use of SVM to estimate discharge instead of the traditional physics-based approaches.

Lizhong et al. [2007] modeled water quality in lakes using remotely sensed images, and support vector machines, finding that the relationships between remote sensed image data and water quality parameters are nonlinear. He recommended the use
of data driven methods that can accommodate nonlinearity. He attempted the use of ANN but due to the limited number of water monitoring stations to provide enough data for training; the iterations would stop at local minima of the loss function (e.g. sum of squared errors), and fail to find the optimum parameter set. A Support Vector Machine was used for its simple structure and good generalization ability, and its ability to perform well in cases of fewer observations than ANN. The results from his study of Lake Taihu in China supported his hypothesis for using SVM, outperforming the ANN to model the water quality in the lake.

Misra et al.[2009] focused on simulation of runoff and sediment yield using SVM, noting that physical models for computing runoff and sediment yield are complex. He modeled the sediment yield from a 7820 km² watershed in India using data from the monsoon period with SVM, concluding that the SVM predicted the sediment yield and the runoff more accurately than using ANNs.

Goel and Pal[2008] modeled scour, and its effect on a grade control structure using both ANN and SVM. He noted that scour was represented by empirical relationships based on laboratory/field experiments on flow, time, material, type of structure) that were computed from particular situations. These empirical formulas did not offer a general computational prediction capability that can be applied to all cases. He pointed out that many scholars have started adopting the ANN algorithms to model scour; in his research he used the SVM with available data from earlier published studies to model the scour and compared the results to those obtained from ANN algorithm with feed forward/back propagation. He recommended the use of the SVM modeling approach in modeling scour because it performed statistically better in comparison to
both ANN and empirical relationships. Similarly, Singh et al.[2008] studied sediment removal efficiency in settling basins using ANN and SVM and reported that the performance of SVM was found to be better statistically compared to the ANN.

3.2.3 Relevance Vector Machine (RVM)

Tipping[2001] found that the SVM suffers from some disadvantages: first, SVM makes excessive use of the kernel function thus the number of required observations grows with the training set; second, estimation of a trade off error/margin parameter, which is accomplished using a cross-validation process (technique used to partition the sample data into subsets and performing the analysis using one of the sets and then validates the analysis using the other set - this is repeated using different data sets and the validation is averaged over all the sets) that is wasteful of data. He introduced the Relevance Vector Machine (RVM) as an alternative based on a Bayesian approach, which does not suffer from the disadvantages of SVM and requires fewer Kernel functions. Tipping explained that the RVM is a probabilistic Sparse Kernel model similar to SVM, where the sparsity is achieved when the algorithm identifies only those observations that improve the performance of the model. The important difference between SVM and RVM is that the RVM method generally requires many fewer observations than the SVM to achieve the same degree of predictive accuracy.

Dogan et al.[2009] mentioned that the RVM is a new algorithm that has not been used widely in modeling sediment transport in natural environmental systems. Dogan et al.[2009] worked with RVM to estimate sediment concentration time series in streams and rivers. He used the data for building his model from a data pool compiled from river
bed load of various kinds and sizes in the U.S and Europe, without considering spatial
distribution of sediment in lakes or rivers. He divided the data randomly into a training
data set and a model validation data set. He used dimensional analysis to select input
parameters (based on physics of sediment and hydraulics of rivers) for his model to
develop the statistical estimation of sediment concentration. He concluded that the use of
this technique is superior to other methods for sediment concentration prediction;
however, as with ANN and SVM, it should not be used for prediction outside the range of
the training data.

Huang et al.[2008] examined the use of RVM to predict stock indices. Similarly
to Partal[2008] with ANN, Huang et al. combined the RVM algorithm with wavelet
techniques to build his model, using waveletsto extract patterns from the variables’ time
series and the extracted features were used as the RVM input to make predictions. The
RVM/wavelet results were statistically compared with the SVM and other traditional
methods using (standard deviation, measures of Skewness and Kurtosis), and it was
found that the use of RVM gave better prediction results.

Wong et al.[2008] worked on a fully automated emotion recognition system on
the basis of facial analysis using RVM. His research was based on dividing the
recognition system into four components (only the first two components were used in his
research). Using different types of kernels to train his model, his results, using a database
of facial expression data, showed detection rates of over 96% for different kernels used,
while the detection rate of less than 51% using a non facial database[recognition of
objects].
Yuan et al. [2007] used RVM with the cross validation to optimize a seed separation process. He used the cross validation process to minimize the approximate error of the data. He then compared the results of this model to an SVM model with cross validation statistically using the root mean square error (RMSE). The results from the statistical analysis supported his approach for the proposed model.

Silva and Ribeiro [2007] examined the RVM for the purpose of text classification, using different types of kernels to help in defining a higher dimensional space. The results of the model were compared to Reuters benchmark, and showed performance improvement by 10\% when compared with a commonly adopted text classification benchmark.

The assessment of the adequacy for supporting aquatic life in the presence of sediment in water bodies are often given in terms of turbidity (NTU) [e.g. DEQ, 2011], a measure related to the amount of fine sediment in water that aggregates the degree to which particulates reflect light over all particle sizes. Physics-based sediment transport models require detailed information about the temporally and spatially variable physical characteristics of the sediment. Developing a model to describe the turbidity in shallow lakes discourages consideration of physics-based models which predict sediment transport based on sediment physical characteristics rather than indirect measures such as turbidity.

The review of the previous methods, the need to verify the ability of data driven models to introduce an easy framework to the public that can model turbidity and other water quality constituents without dealing with complex data requirement for physics
based models and the water quality criteria requirements, support the exploration of the RVM to study very fine sediment using turbidity as a surrogate.

3.3 RVM Model Structure

The goal of any model is to provide predictions that faithfully represent the target observations with a simple formulation as the observations allow. As discussed previously, RVMs are sparse data driven models that use techniques pioneered in pattern recognition applications. The RVM adopts a Bayesian approach to learn which observations in a data set, $x$, are key to reproducing the patterns represented by those observations, and seeking sparsity by using only those observations that contain independent useful information about the process being modeled.

The RVM model is fitted to a set of target observations of a particular type, $n$, (suspended solids, dissolved oxygen, etc.) by first creating a kernel function, $\Phi^{(n)}(x_i^{(n)}, x_{d,i})$, that represents both the influence of underlying system drivers $x_{d,i}$, and the observations for type $n$, $x_i^{(n)}$, and defining a set of weights, $w$, that multiply the kernel function. These products are then summed to form the vector of predicted values for observation type $n$. The RVM algorithm then modifies the weights, $w$, to minimize the discrepancy between the observed target values and the corresponding predicted values. Sparsity is achieved when one or more of the estimated weights equals zero, indicating that the corresponding observations do not significantly improve the model - represented mathematically by a matrix with most of the elements equal to zero, while the non-zero elements are used for prediction. The importance of sparsity is to minimize the amount of
data required for observations. Once the relevant observation vectors are identified, this information can be used to improve the design of future monitoring campaigns.

Mathematically, the predicted value for the target observations is given by

\[
y^{(n)}(x, w^{(n)}) = \sum_{i=1}^{N} w_i^{(n)} \Phi^{(n)}(x_i^{(n)}, x_{d,i}) + w_0^{(n)}
\]

(3.1)

\[
y^{(n)}(x, w^{(n)}) = \sum_{i=0}^{N} w_i^{(n)} \Phi^{(n)}(x_i^{(n)}, x_{d,i})
\]

(3.2)

\[
y^{(n)}(x, w^{(n)}) = w_i^T \Phi(x_i^{(n)}, x_{d,i})
\]

(3.3)

in which \(y^{(n)}(x, w^{(n)})\) is the vector of predictions for variable of type \(n\) given the observations matrix, \(x\) and the vector of weights for variable of type \(n\), \(w^{(n)}\). The kernel function \(\Phi^{(n)}(x_i^{(n)}, x_{d,i})\) is the inner product of a mapping function for observations that relates the system drivers and the target observations of variable type \(n\). Although the mapping function is general, here we assume a Gaussian kernel, yielding:

\[
\Phi^{(n)}(x_i^{(n)}, x_{d,i}) = \exp(-r^2||x_{d,i} - x^{(n)}||^2)
\]

(3.4)

in which \(r\) is the kernel width (selected and fixed for a particular RVM model) that provides the multi-plane representation of \(x\). The targets (observations matrix used to train the MVRVM model) are samples from the observations, which will contain error after training \(t_n = y(x_n, w) + \epsilon_n\) where \(\epsilon\) is independent zero-mean Gaussian noise with variance \(\sigma^2\), and \(\epsilon \sim N(0, \sigma^2)\);

From this, it can be inferred that the probability distribution of \(t_n\), conditioned on the observations \(x\) is

\[
p(t_n|x) = N(t_n|y(x_N), \sigma^2)
\]

(3.5)
The likelihood of the complete data set is represented by

\[ p(t \mid w, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp \left\{ -\frac{1}{2\sigma^2} ||t - \Phi w||^2 \right\} \]  

(3.6)

where

\[ t = [t_1, t_2, \ldots, t_N]^T, \, N \times 1 \text{ vector} \]  

(3.7)

\[ w = [w_0, w_1, \ldots, w_N]^T, \, (N + 1) \times 1 \text{ vector} \]  

(3.8)

\[ \Phi = [\Phi x_1, \Phi x_2, \ldots, \Phi x_N]^T, \, N \times (N + 1) \text{ matrix} \]  

(3.9)

The Bayesian training algorithm requires the definition of explicit prior distributions for the weights:

\[ p(w \mid \alpha) = \prod_{i=0}^{N} N(w_i \mid 0, \alpha_i^{-1}) \]  

(3.10)

in which \( \alpha \) is a vector of \((N + 1)\) prior parameters. For a given test point \( x_* \), we predict the probability of \( x_* \)

\[ p(t_* \mid t) = \int p(t_* \mid w, \sigma^2) p(w, \alpha, \sigma^2 \mid t) \, dw \, d\alpha \, d\sigma^2 \]  

(3.11)

where \( p(w, \alpha, \sigma^2 \mid t) \) is defined by Bayes rule as

\[ p(w, \alpha, \sigma^2 \mid t) = \frac{p(t \mid w, \alpha, \sigma^2) \cdot p(w, \alpha, \sigma^2)}{p(t)} \]  

(3.12)

Tipping [2001] mentioned that “we cannot perform these computations in full analytically, and must seek an approximation. We cannot compute the posterior \( p(w, \alpha, \sigma^2 \mid t) \) directly instead, we decompose the posterior as:”

\[ p(w, \alpha, \sigma^2 \mid t) = p(w \mid t, \alpha, \sigma^2) \cdot p(\alpha, \sigma^2 \mid t) = \frac{p(t \mid w, \sigma^2) \cdot p(w, \alpha)}{\int p(t \mid w, \sigma^2) \cdot p(w, \alpha) \, dw} \]  

(3.13)

The posterior over weight (constrained over the distribution of weights) is expressed as

\[ p(w \mid t, \alpha, \sigma^2) = \frac{p(t \mid w, \sigma^2) \cdot p(w, \alpha)}{\int p(t \mid w, \sigma^2) \cdot p(w, \alpha) \, dw} \]  

(3.14)
All the probability density functions are Gaussian, thus we can obtain an analytical expression for the posterior probability density function equation over the weight:

\[ p(w|t, \alpha, \sigma^2) = (2\pi)^{-\frac{N+1}{2}} \sum_{i=1}^{N} \left( \sum \frac{1}{2} \exp \left\{ -\frac{(w-\mu)^T\Sigma^{-1}(w-\mu)}{2} \right\} \right) \]  

(3.15)

\[ \sum = (\sigma^{-2} \Phi \Phi^T + A)^{-1} \]  

(3.16)

\[ A = \text{diag}(\alpha_0, \alpha_1, ..., \alpha_N) \]  

(3.17)

\[ \mu = \sigma^{-2} \sum \Phi^T t \]  

(3.18)

Relevance vector (learning) thus becomes the search for the hyper (multi dimension) parameters that maximize

\[ p(\alpha, \sigma^2|t) \propto p(t|\alpha, \sigma^2)p(\alpha)p(\sigma^2) \]  

(3.19)

with respect to \(\alpha\) and \(\sigma^2\).

### 3.4 Objectives and Experimental Design

The objective of this paper is focused on development of and testing the Multi Variate Relevance Vector Machine (or MVRVM) as a mathematical algorithm that can be used to predict patterns in the concentration of suspended fine sediment, and other environmental constituents and, and help to find, how many observations are required to model the complex hydraulics, sediment, and water quality constituents. The MVRVM has not been used in many studies to predict sediment concentration in estuaries and lakes [Dogan et al., 2009].

This paper will examine whether the MVRVM is able to carry out predictions for suspended fine sediment and other water quality constituents.
3.4.1 Experimental design

The study was carried out using Mud Lake, a part of the Bear River National Wildlife refuge in southeastern Idaho [Figure 3-1] that functions as a sediment trap for flows into the adjacent Bear Lake in addition to functioning as a habitat to support migratory species.

The evaluation of environmental quality in Mud Lake to ensure satisfying its beneficial uses can be improved by successful modeling of environmental constituents. However the hydraulic operation, limited observations, and the variable particle size of sediment transported through Mud Lake present challenges to select an appropriate modeling technique or sediment transport function.

To test and validate the MVRVM for the objective, data were collected as detailed in [Batt and Stevens, 2012], and consist of concentrations of fine suspended sediment, turbidity, dissolved oxygen, pH, and temperature at 30 locations, in Mud Lake biweekly over two ice-free seasons in 2009 and 2010.

The investigation of the use of the MVRVM model requires the existence of representative patterns (spatial and temporal) of observations at many locations [Batt and Stevens, 2012]. Consistent patterns in the data were the key requirement to support the use of the MVRVM with limited observations. Initially it was assumed that the flow hydrodynamics (flow velocity, depth, direction) in Mud Lake represented the major driving force for all variables. This assumption was found to be inadequate in the case of modeling the water quality constituents, for which the successful prediction required the collection of additional data which was not considered during the preliminary data collection, namely the effect of vegetation and algae on the observations for DO, and pH.
For the Mud Lake case study, 30 locations were selected for observation, and sediment and water quality observations were made biweekly during the ice-free periods from April - October 2009 and 2010. Details of the data collection efforts are provided in [Batt and Stevens, 2012] and the observation set consists of time series of levels of the six constituents at each of the 30 locations.

The model runs for training, verifying, and prediction required arrangement of the data according to the time and location of the observation. The time series observations consist of 25 days of observation. Based on the preliminary analysis for the RVs of the TSS it was found that 22 days were required for training the MVRVM; while the testing consisted of 3 days. The data are arranged in matrix form for all the constituents and the location of each observation. The input data consists of time series observations matrix for water quality parameters, velocity vectors magnitude and turbidity for the four input locations in the Lake (stations 1, 8, 7, 25 in Figure 3-1), while the MVRVM output data consists of time series observations matrix for the 30 locations for water quality parameters, velocity vector magnitude and turbidity in the Lake. The algorithm is executed while changing the model width, with changing the kernel equation, and number of iterations. The output is then used, with the field observations to plot the RMSE and residuals in order to identify the error in the MVRVM algorithm parameters and what parameters should be changed to minimize the errors.

The MVRVM training and verification were done using a library created for Matlab. The model runs in this study took a range of 5 to 15 minutes to select the required RVS for each constituent using a dual core 2.4 GHZ machine. Jolliffe[1991] mentioned that the quartile analysis is easy to understand and is considered an easy way
to summarize the data; he mentioned also that the quartiles are useful to summarize data and not influenced by the extreme observations. Thus we choose to present our observations and modeled data using the quartiles to eliminate relying on any single extreme observation that can mislead the readers in their understanding the range that prevails in the observation.

3.5 Results and Discussion

The water quality observation matrix described in [Batt and Stevens, 2012] was used for training the MVRVM algorithm. During training, the MVRVM algorithm selects observations that provide relevant information to the model based on Bayesian probability theory. Once an observation is selected as relevant, the remaining observations at the same time are added to the prediction matrix. This process continues until the addition of a new observation supplies no new information for the model; creating a matrix of vector observations containing the string of important vector information and convergence is declared. The Term Relevance Vector can be misleading in some study cases; as the scientists refer to single observations with high probability selected by the model as a vector, forgetting that the model selects more information from the training data set and create the string of Relevant Vector Observations. Sparsity is measured by the fraction of total number of observation that are significant, as an example in the case of velocity vector when the model selected 15 RVs it will mean that 15 out of 660 observations significant. The number of RVs that are a subset of the training data; these sets of RVs are then used for the prediction in the algorithm [Khalil et al., 2005].
Khalil et al.[2005] described the complexity of the model as proportional to the number of the selected RVs, and it was expected that the number of RVs would likely change depending on the complexity of the observed pattern. The number of relevant vectors relative to the total size of the set of observations is the measure of sparsity of the model. Here, we express this measure of sparsity as the percentage of the total number of observations that are included in the set of RVs. As described subsequently and in Batt and Stevens[2012] the physical metadata related to the RVs (i.e. time and location) may provide information for designing future monitoring systems.

The results from the MVRVM model are provided as box-and-whisker plots in Figure 3-2, with each adjacent box and whisker pair representing the distributions of the observations (unshaded box, left) and predictions (shaded box, right). The MVRVM model was capable of predicting the water quality constituents and moreover, to capture the patterns of change in the different locations in our case study. The quality assurance/control for collecting observations (Batt and Stevens, 2012), and design of the experiment minimized any effect of changing the number of iterations that are used to run the MVRVM. For the water quality constituents tested here, the fact that the MVRVM parameter estimate routine readily converged before all data vectors were used suggested that the data collected were sufficient for the MVRVM algorithm.

Residuals plots for the tested constituents [Figure 3-3] that the residuals (observed – predicted levels of the constituents at each time/location) are centered around zero and do not follow a specific pattern (random). Figure 3-4 shows that the observations and predictions do not perfectly fall on the 45 degree line of agreement thus the error exists but spread on all the locations.
The range of percentile observations follows 3 patterns data constituents among
the data collection sites. For discussion purposes locations with same characteristics were
groups in three zones [for more information about zones see Figure 3-1, and Batt and
Stevens [2012].

1. **Pattern type I** [Figure 3-1c, Table 3-1](temperature) is characterized by no
significant change in the percentile range of observations through all the locations.
The RMSE was constant in all three zones, which was expected because the
meteorological conditions that affected temperature were common for all spatial
locations. The temperature pattern didn’t vary significantly from Zone1 to Zone 3.
Since the range of the temperature across location for each time was small, only six
RVs, about 0.9% of the observations, were required to adequately represent
temperature.

2. **Pattern type II**[Figure 3-1( a and b), Table 3-1]pH and DO are characterized by
change in the percentile range of observations for all the locations especially in the
Zone 3, where there is an increase of vegetation and algae that affects the pH and DO
observations during the daytime. The MVRVM algorithm was able to capture the
range of percentile change for the pH and DO observations in all the. The pH and DO
pattern is represented by 16 and 13 RVs respectively. The RMSE for the pH and DO
increased from Zone 1 to Zones 2 and 3, likely due to the increasing amounts of
vegetation from Zone 1 through Zones 2 and 3.

3. **Pattern type III** [Figure 3-1 (d, e and f), Table 3-1](Turbidity, TSS, and velocity
magnitude) is described by random level with larger variability for percentiles in the
first zone, decreased variability in the second zone, and then dropping to near zero for
both level and variability in the third zone. The MVRVVM algorithm reflected this change in the output of the model compared to the field observations. The turbidity is a very interesting constituent because it is related to the TSS, and is measured under the same conditions as the TSS. However the accuracy of measuring turbidity is higher. This accuracy leads to selecting 16 RVs which is less than for the TSS (22 RVs).

In the case of velocity vector the noise in collecting the observations and the small magnitude of velocity in Mud Lake resulted in the number of RVs to be 15, implying similar results as TSS and turbidity. However the selection of the 15 relevant vectors is an artifact of the algorithm which is not robust when the signal-to-noise ratio is low. This number would likely change in case of existence of observations that are above the detection limit of the equipment used.

The turbidity RMSE in zone 1 (2%, which is less than the device error) increased to 10% in zones 2 and 3, which indicates that zones 2 and 3 have the same error. The source of error can result from more complex hydraulics compared to the deeper waters of zone 1, the shallower regions in zones 2 and 3. TSS is heterogeneous and during collection of grab samples we might capture clumps of sediment in either the sample or the smaller portion taken for filtering, which actually does not represent the whole water. These problems lead to an error in observing TSS as high as 10-20% of the mean concentration. The average RMSE was small for the zones 1 and 2 (6%) and zone 3 (14 %) however the error in the 3 zones are within range of the typical measurement error for TSS of 10-20%. The RMSE for velocity seemed high in the first 2 zones but the magnitudes of the velocities were small and close to the error in the device accuracy.
Also the RMSE in zone 3 is (1%) but this error does not mean there is no error because the magnitude of the velocities of this zone close to zero.

The results shown above confirmed that the MVRVM is able to model each water quality constituents. The success of prediction varied depending on the type of constituent tested and the complexity of the hydraulics affecting the sampling location. The MVRVM demonstrated an ability to vary with the minimum and maximum range of the parameters. The MVRVM method found an average number of observations that contribute important information (RVs) between 1 and 3 percent of the total number of observations. However the question remains concerning whether this number of RVs means that 3% of the observations are sufficient to model complex water quality constituents. In the next paper we examine in more detail how much data are required to successfully model the complex water quality constituents and how the RVs are related to this data.

3.6 Conclusions

This paper is the first study to consider the use of MVRVM to model suspended fine sediment transport and other water quality constituents in complex natural systems as in the case of Mud Lake. The MVRVM output demonstrated the capability of the method to capture the spatial and temporal change in patterns in observations for suspended fine sediment and a variety of water quality constituents. The assumption of using the sample location to construct the MVRVM for modeling the selected water quality parameters and suspended fine sediment has proven to work adequately for most of the tested constituents.
The MVRVM results showed a changing RMSE from Zone to Zone based on the type of constituent tested and the sampling location. It is suggested that additional types of observations (e.g. algae, other vegetation) that may influence the selected constituents should be included in future work. For example for DO and pH the amount of algae or other vegetation present near the sampling locations in Mud Lake may improve the predictive ability of the MVRVM model.

The number of MVRVM relevance vectors changed according to the complexity of the modeled pattern. This information could be used to inform design monitoring programs for the purpose of MVRVM.

References


**Table 3-1**: Observed, modeled Constituents percentile, and RMSE percentage with respect to Zones in Mud Lake, as (a) DO mg/l, (b) pH, (c) Temperature °C, (d) Turbidity NTU, (e) TSS mg/l, (f) Velocity cm/s.

<table>
<thead>
<tr>
<th>Zone</th>
<th>DO ± 0.2 mg/l</th>
<th>pH ± 0.2</th>
<th>Temperature ± 0.1 Celsius</th>
<th>TSS ± 15% mg/l</th>
<th>Turbidity ± 3% NTU</th>
<th>Velocity magnitude ± 1.5 cm/s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min 25 Percentile</td>
<td>50 Percentile</td>
<td>75 Percentile</td>
<td>Max</td>
<td>RMSE %</td>
<td>Min 25 Percentile</td>
</tr>
<tr>
<td>Zone 1</td>
<td>4.92 7.44 8.11</td>
<td>9.33</td>
<td>11.48</td>
<td>MVRVM</td>
<td>5.04 7.17 7.90</td>
<td>9.14</td>
</tr>
<tr>
<td>Zone 2</td>
<td>4.48 6.81 7.88</td>
<td>9.35</td>
<td>11.19</td>
<td>MVRVM</td>
<td>6.27 7.54 8.28</td>
<td>9.25</td>
</tr>
<tr>
<td>Zone 3</td>
<td>5.95 8.21 9.01</td>
<td>10.37</td>
<td>14.63</td>
<td>MVRVM</td>
<td>6.31 8.38 9.25</td>
<td>9.76</td>
</tr>
<tr>
<td>Zone 1</td>
<td>6.35 7.98 8.12</td>
<td>8.26</td>
<td>8.95</td>
<td>MVRVM</td>
<td>6.35 7.96 8.11</td>
<td>8.23</td>
</tr>
<tr>
<td>Zone 2</td>
<td>7.74 8.17 8.33</td>
<td>8.56</td>
<td>9.12</td>
<td>MVRVM</td>
<td>7.82 8.20 8.33</td>
<td>8.52</td>
</tr>
<tr>
<td>Zone 3</td>
<td>5.22 8.32 8.62</td>
<td>9.07</td>
<td>10.18</td>
<td>MVRVM</td>
<td>6.74 8.34 8.61</td>
<td>9.00</td>
</tr>
<tr>
<td>Zone 1</td>
<td>1.46 7.20 14.27</td>
<td>17.93</td>
<td>23.81</td>
<td>MVRVM</td>
<td>2.16 7.27 14.42</td>
<td>17.65</td>
</tr>
<tr>
<td>Zone 2</td>
<td>1.41 7.91 14.88</td>
<td>18.82</td>
<td>25.01</td>
<td>MVRVM</td>
<td>2.46 7.72 15.17</td>
<td>19.33</td>
</tr>
<tr>
<td>Zone 3</td>
<td>0.97 8.19 15.07</td>
<td>19.28</td>
<td>25.56</td>
<td>MVRVM</td>
<td>2.21 7.93 15.44</td>
<td>19.69</td>
</tr>
<tr>
<td>Zone 1</td>
<td>0 14.67 27.67</td>
<td>52.00</td>
<td>170.67</td>
<td>MVRVM</td>
<td>0 14.80 25.15</td>
<td>51.14</td>
</tr>
<tr>
<td>Zone 2</td>
<td>0 10.67 17.00</td>
<td>34.33</td>
<td>108.67</td>
<td>MVRVM</td>
<td>0 8.39 14.66</td>
<td>28.08</td>
</tr>
<tr>
<td>Zone 3</td>
<td>0 3.33 6.00</td>
<td>10.67</td>
<td>100.00</td>
<td>MVRVM</td>
<td>0 1.85 3.80</td>
<td>8.47</td>
</tr>
<tr>
<td>Zone 1</td>
<td>0 19.38 49.75</td>
<td>94.55</td>
<td>357.30</td>
<td>MVRVM</td>
<td>0 22.74 50.02</td>
<td>90.04</td>
</tr>
<tr>
<td>Zone 2</td>
<td>0 13.50 29.15</td>
<td>60.20</td>
<td>175.30</td>
<td>MVRVM</td>
<td>1.04 15.97 29.32</td>
<td>61.04</td>
</tr>
<tr>
<td>Zone 3</td>
<td>0 0 2.95</td>
<td>14.48</td>
<td>78.40</td>
<td>MVRVM</td>
<td>0 0.80 3.93</td>
<td>14.32</td>
</tr>
<tr>
<td>Zone 1</td>
<td>0 1.72 5.22</td>
<td>10.73</td>
<td>55.11</td>
<td>MVRVM</td>
<td>0 1.65 4.45</td>
<td>8.48</td>
</tr>
<tr>
<td>Zone 2</td>
<td>0 0.07 0.40</td>
<td>3.14</td>
<td>91.64</td>
<td>MVRVM</td>
<td>0 0.09 0.39</td>
<td>3.24</td>
</tr>
<tr>
<td>Zone 3</td>
<td>0 0 0.01</td>
<td>2.22</td>
<td>156.66</td>
<td>MVRVM</td>
<td>0 0 0.02</td>
<td>2.13</td>
</tr>
</tbody>
</table>
Figure 3-1 Box Plots of collected observations in the thirty locations of Mud Lake during 2009-2010 and Predicted MVRVM output as (a) TSS, mg/L, (b) turbidity (NTU), (c) DO, mg/l, (d) p, std. units, (e) Temperature oC, (f) Velocity cm/s. The shaded bars contain the observation distributions and MVRVM prediction distributions for each location.
Figure 3-2 Residual plots of observations in the thirty locations of Mud Lake during 2009-2010 versus and Predicted MVRVM output as (a) pH, (b) DO mg/l, (c) TSS mg/l, (d) Turbidity NTU, (e)Temperature °C,(f) Velocity cm/s. The different plotting symbols represent different sampling dates.
Figure 3-3 Scaled plots of observations in the thirty locations of Mud Lake during 2009-2010 versus and Predicted MVRVM output as (a) pH, (b) DO mg/l, (c) TSS mg/l, (d) Turbidity NTU, (e)Temperature °C, (f) Velocity cm/s. The different plotting symbols represent different sampling dates.
CHAPTER 4
HOW TO UTILIZE RELEVANCE VECTORS TO COLLECT REQUIRED DATA FOR
MODELING WATER QUALITY CONSTITUENTS, AND FINE SEDIMENT IN
NATURAL SYSTEMS?

Abstract

Relevant Vectors are subsets of the training data set that contain significant information about the modeling state variables. Previous studies that focused on modeling water resources, sediment transport, and water quality measures, using statistical learning tools like MVRVM, have not considered important factors related to the practical use of the MVRVM such as: 1-how much data does the Relevant Vectors correspond to with respect to forming the prediction matrix? and 2-what is the significance of the Relevance Vectors on making decisions? In this chapter we will use the case study of Mud Lake to investigate how careful experimental design and the construction of the MVRVM framework can answer these questions. Results showed that using the MVRVM can eliminate the need of more than 50% of the data collected.

4.1 Background

The multivariate relevant vector machine, (MVRVM), is a statistical learning tool that has proven capability to extract information contained in time series of data [Zaman, 2010]. Khalil et al. [2005] described Relevance Vectors (RVs) as the subset of the observations that are found to contain the most relevant information on which the MVRVM model is built. The MVRVM is an extension to the simple RVM learning tools that allows categorization of data by, for example, time and spatial location.
Tielavilca and McKee [2011] used a MVRVM to develop a model for predictions of required daily irrigation releases from a multiple reservoir system. He also used the selection of Relevance Vectors to determine the significance of the category year on reservoir release.

Each of these efforts can be viewed from the perspective of the design of a monitoring system, with the specific spatial and temporal metadata associated with the RVs specifying observations events that will best inform the MVRVM model. In this paper we will explore how RVs maybe used in different ways to design experiments in research, and how the construction of the model itself can reveal or mask the importance of each observation to the model.

This paper describes the use of the MVRVM for predicting fine sediment distribution and general water quality in a wetland-lake system used as a wildlife refuge and as a sediment trap, subject to flow scheduling related to irrigation requirements. For the case study described below, collection of preliminary data to develop an observation network suggested that a single driving force, system hydraulics, controlled the spatial suspended fine sediment distribution. Although this assumption may not be correct for all the constituents, it was considered practical for situations in which observations may be unavailable. Collecting data at a frequency suitable for management decisions concerning fine sediment transport is expensive, and it is a challenge to determine how much data is sufficient. We propose here that the selection of the RVs from constructing an MVRVM may be used to understand which locations in the aquatic system are the most beneficial and lead to a more effective monitoring system.
4.2 Objectives

This chapter will address the following objectives:

1- Verify whether the selected relevance vectors are linked to important locations in MudLake and understand why they were selected.

2- Address the concern of whether the relevance vectors can help to decrease the data required for modeling.

4.3 Methods

The study area is Mud Lake, a unit of the Bear River Wildlife Refuge in southeastern Idaho [Figure 4-1] and detailed in [Batt and Stevens, 2012a], which functions as a sediment trap for the Bear River, as well as a habitat for migrant species. Mud Lake is characterized by complex hydraulics due to its heterogeneous vegetation patterns and bathymetry, and only small numbers of observations for either water quality measures or spatial sediment distribution. For purposes of this study, the sampling locations in Mud Lake were organized into three zones based on observations similarities in the data, as shown in Figure 4-1 that represent the lake inlet (Zone I), a transitional region (Zone II) and the lake outlet to Bear Lake (Zone III).

To achieve the objectives of exploring the relevance vectors as monitoring system design aids, the environmental system of Mud Lake was idealized as an irregular network grid where the observations at each location potentially interact with those at surrounding locations. Batt and Stevens[2012a] collected 660 observations from 22 time series in 2009 and 2010, for suspended fine sediment and the water quality constituents dissolved oxygen, temperature, pH, and turbidity in Mud Lake. For use in the MVRVM,
observations were organized as time series for each of 30 sampling locations for suspended fine sediment and each water quality constituent. Storing them as time series is one of the keys for making use of the relevance vectors with the MVRVM algorithm structure, as is summarized by [Tipping, 2001].

The RVM model is fitted to a set of target observations of a particular type, $n$, (suspended solids, dissolved oxygen, etc.) by first creating a kernel function, $\Phi^{(n)}(x_{i}^{(n)}, x_{d,i})$, that represents both the influence of underlying system drivers $x_{d,i}$, and the observations for type $n$, $x_{i}^{(n)}$, and defining a set of weights, $w$, that multiply the kernel function. These products are then summed to form the vector of predicted values for observation type $n$. The RVM algorithm then modifies the weights, $w$, to minimize the discrepancy between the observed target values and the corresponding predicted values. Sparsity is achieved when one or more of the estimated weights equals zero, indicating that the corresponding observations do not significantly improve the model - represented mathematically by a matrix with most of the elements equal to zero, while the non-zero elements are used for prediction. The importance of sparsity is to minimize the amount of data required for observations. Once the relevant observation vectors are identified, this information can be used to improve the design of future monitoring campaigns.

Mathematically, the predicted value for the target observations is given by

$$y^{(n)}(x, w^{(n)}) = \sum_{i=1}^{N} w_i^{(n)} \Phi^{(n)}(x_i^{(n)}, x_{d,i}) + w_0^{(n)} \quad (4.1)$$

$$= \sum_{i=0}^{N} w_i^{(n)} \Phi^{(n)}(x_i^{(n)}, x_{d,i}) \quad (4.2)$$

$$= w_i^T \Phi(x_i^{(n)}, x_{d,i}) \quad (4.3)$$
in which \( y^{(n)}(x, w^{(n)}) \) is the vector of predictions for variable of type \( n \) given the observations matrix, \( x \) and the vector of weights for variable of type \( n \), \( w(n) \). The kernel function \( \Phi^{(n)}(x^{(n)}_i, x_{d,i}) \) is the inner product of a mapping function for observations that relates the system drivers and the target observations of variable type \( n \). Although the mapping function is general, here we assume a Gaussian kernel, yielding:

\[
\Phi^{(n)}(x^{(n)}_i, x_{d,i}) = \exp\left(-r^2 \|x_{d,i} - x^{(n)}\|^2\right) \tag{4.4}
\]

in which \( r \) is the kernel width (selected and fixed for a particular RVM model) that provides the multi-plane representation of \( x \). The targets (observations matrix used to train the MVRVM model) are samples from the observations, which will contain error after training \( t_n = y(x_n, w) + \varepsilon_n \) where \( \varepsilon \) is independent zero-mean Gaussian noise with variance \( \sigma^2 \), and \( \varepsilon \sim N(0, \sigma^2) \);

From this, it can be inferred that the probability distribution of \( t_n \), conditioned on the observations \( x \) is

\[
p(t_n|x) = N(t_n|y(x_N), \sigma^2) \tag{4.5}
\]

The algorithm selects an observation as relevant based on the probability mentioned in equation 5.

### 4.4 Results and Discussion

Batt and Stevens[2012b] found that the MVRVM output results for sediment and water quality constituents supported the capability of the MVRVM to model and forecast these constituents, and that the number of selected RVs for each constituent corresponds
to the number of observations used for prediction. They also found that the root-mean-
square error (RMSE), $RMSE = \sqrt{\frac{\sum_{n=1}^{m} (t(n) - y(x_n))^2}{m-1}}$, for each constituent depended on the
sampling location, the sampling and measurement error, and parameters that influence
constituent levels. The results in Table 4-1 and Figure 4-2 suggest that in Mud Lake there
is an effect of the sampling locations selected by the MVRVM, noting that 60% of the
selected relevance vectors for the tested constituents are in Zone 1 (the source of input
flow to Mud Lake), and the remaining 40% of the RVs are divided equally among Zone 2
and Zone 3. If we had grouped the flow input location from Bear Lake with the
remaining input in Zone 1; it would likely result in an increase of the relevant vectors to
80% in Zone 1, leaving the remaining 20% of the relevant vectors to the other two zones.

A description of significant locations for the data pattern in Mud Lake is as
follows [Figure 4-1]: locations (8, 9, 10, and 11) are near the inlet of the lake, the source
of suspended fine sediment input to Mud Lake, sample locations (12, 13, 14, and 15) are
located in the north eastern (canal like) part of Mud Lake which conveys part of the
suspended fine sediment from the source to the rest of Mud Lake. Sample locations (0, 1,
2, 3, 16, 17, and 18) are located where the canal like portions joins together with the open
water in the middle of Mud Lake. On the western side of the lake, sample locations (4, 5,
6, and 7) are in the irrigation canal where the relatively sediment-free water from Bear
Lake is flowing toward the open water of Mud Lake. Sample locations (19, 20 and 21)
are in remote locations in the southeast part of the lake, separated from the open lake by
dense vegetation, and are least likely to be affected by changing hydraulic conditions.
The remaining monitoring locations (23, 24, 25, 26, 27, 28, and 29) were not selected by
the MVRVM because their patterns are correlated in nearby locations that were already selected as RVs. The selection of the amount of data needed to train the MVRVM algorithm was created by using the TSS data because its pattern contained noise compared to the other tested constituents. So we started the training using all the data we have (25 series), but we have noticed that the algorithm in this case required only 22 series of the data; thus we decided that this will be the required number for creating a training data set. The results for all constituents can be found in Table 4-1 and Figure 4-2.

**TSS:** Although closely related, TSS and turbidity are not the same. TSS samples are difficult to collect and analyze because TSS is heterogeneous and, during collection of grab samples, we might capture or exclude clumps of sediment in either the sample or the measurement aliquot, which actually does not represent the whole water. These limitations lead to errors in observed TSS as high as 10-20%, consistent with published norms [US EPA Method 160.2]. Because error clouds information in data, the number of RVs selected by the model is likely to be larger than for other, more easily measured constituents. The MVRVM found that 22 relevance vectors were needed to create the frame work to model the TSS and was the highest among all constituents, likely due to the complex pattern of sediment in Mud Lake, and the larger sampling and measurement errors. The selection of this number of RVs means that 100% of the 660 observations were needed to create the frame work of the MVRVM model. This selection of relevant vectors corresponds to observations that have the highest probability and they occurred in these locations.

**TURBIDITY:** The observed turbidity results closely follow the TSS, however the MVRVM choose only 16 relevance vectors (locations 0, 1, 2, 5, 6, 7, 9, 13, 14, 15, 16,
17, 18, 19, 20, 21), compared with the 22 RVs for the TSS, likely because the turbidity measurements are less prone to the variability inherent in the TSS measurements, but also that turbidity showed a less complex pattern than the TSS. Turbidity is much more consistent in the observations because the turbidity instrumentation averaged the signal over a longer time (approximately 30 seconds) compared to the time required to obtain a grab sample; so many short term variations are averaged out. These factors lead to reduced variability in turbidity measurements and suggest that the complexity observed in the turbidity patterns is more likely to be true complexity of the system and less so random variability. The selection of the 16 RVs means that 72% of the 660 observations are needed to create the frame work of the MVRVM model for turbidity.

**VELOCITY VECTOR MAGNITUDE:** The velocity magnitude is a special case due to the presence of many censored observations, recorded as $< 15$ cm/s (instrument detection limit) at most locations. As described in Batt and Stevens[2012b] the MVRVM algorithm used results from the hydrodynamic model rather than direct observations. During training the MVRVM selected 15 relevance vectors from the hydrodynamic model results, at sampling locations (0, 1, 2, 3, 4, 5, 9, 10, 12, 14, 15, 19, 20, 21, and 22). The RVs were selected in locations of hydraulic change; however the RMSE results should be discounted because the training was generated from mechanistic model rather than using direct observations. The selection of this number of RVs means that 70 % of the 660 observation data is needed to create the frame work of the MVRVM model.

**DO:** The MVRVM selected 13 samples (1, 3, 6, 7, 8, 9, 10, 11, 13, 15, 18, 20, and 21) to be relevant for dissolved oxygen; from the water quality measures there was an inverse relationship between the sediment concentration in the water body and DO, in addition
the presence of algae in any location will increase the observed DO. These factors demonstrated that the RVs were located closely to the RVs for turbidity as in the inflow locations and wherever hydraulics change. Relevant vectors for DO were also located in vegetated areas and where algae were observed. The selection of this number of RVs represented 59% of the 660 observation data as needed to create the framework of the MVRVM model.

**pH:** The MVRVM in the case of pH chose locations (1, 3, 4, 5, 7, 8, 9, 10, 12, 13, 14, 16, 17, 18, 19, and 21) to be significant for the model. Similar to the case of DO, the RVs of the pH were also located where the hydraulics changed in Mud Lake and existence of vegetation or algae. However the observation pattern was somewhat more complex compared to the DO and required more RVs: 70% of the 660 observation data were needed to create the framework of the MVRVM model.

**TEMPERATURE:** The MVRVM chose 6 RVs (27% of observations) at locations (0, 6, 9, 13, 16, and 18). This number of RVs was the smallest among tested constituents, because temperature fluctuations across the locations were small during the daily collection period. The individual RVs represent the different major areas of Mud Lake, but note that locations adjacent to those selected were excluded. It is not surprising that location 9, at Mud Lake inlet, is represented since the major driver of temperature is the inlet flow, as it responds to snowmelt, etc. The remaining locations represent the influence of the flow paths through Mud Lake and the climate conditions that would warm or cool the water. Location 6 reflects the influence of the flow from Bear Lake which will generally have much different temperature fluctuations than in Mud Lake’s interior.
The experimental design for data collection can be improved by understanding the outcome the RVs selection by the MVRVM algorithm for collection of future observations to serve the management objectives. As an example the runoff to Mud Lake in 2011 was considerably higher than in 2009 and 2010, the years of our case study. Because of this change in runoff, the collected constituents’ patterns would likely be different; reflecting the presence of high flows and additional MVRVM training would be beneficial. In this case, to collect new observations we can use the relevant locations selected by the MVRVM and according to the required amount by the relevant vectors.

The experimental design and the data collection can be improved through examining the important objectives of the study. For example, using the results of our study, if the key constituent is turbidity rather than TSS, thus we require only 16 locations (see Table 4-1) from which to collect samples rather than the 22 locations as in the case of TSS. For temperature, only six locations would be required to capture temperature dynamics. A similar assessment would be made for each constituent under study.

The management decision can play a part in changing the experimental design. As an example if the management decision is to monitor the change of DO because of a change in conditions in the lake, such as the onset of low flow conditions, managers can use the locations that were selected by the MVRVM code without missing information relevant to the model.
4.5 Summary and Conclusion

The MVRVM statistical algorithm can be used to model water quality constituents in complex natural systems. The careful planning of field observations, and arrangement of the framework for the MVRVM, will help in creating a series of relevance vectors (RVs) that can be used to better understand the patterns (spatial and temporal) of suspended fine sediment and water quality constituents in the natural system. Keeping the collected observations in time series related to the sampling locations created an advantage to indicate how much data is sufficient to construct the MVRVM framework. This paper has demonstrated the ability of RVs to select subsets of observation that capture important patterns (spatial and temporal) that might be obscured by random variability, and thus can be used as aids to construct experimental designs.

References


Zaman, B. (2010), Remotely sensed data Assimilation Technique to develop machine learning models for use in water management.
**Table 4-1** Locations of RVs for different constituents, percentage of selected RVs from data, and percentage of required observation for successful modeling based on RVs

<table>
<thead>
<tr>
<th></th>
<th>DO</th>
<th>pH</th>
<th>Temp.</th>
<th>TSS</th>
<th>Turbidity</th>
<th>Velocity</th>
</tr>
</thead>
<tbody>
<tr>
<td>RVs %</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
Figure 4-1 Example of selected RVs in DO pattern, where the red points represent the RVS, black lines represent series which contain RVs, and colored lines represent series which contain no RVs and not used to make predictions.

Figure 4-2 Sample Locations in Mud Lake, and selected Relevant Locations from MVRVVM model (the black lines represent time series, the red points represent the selected RV, the colored series represent series which don’t have RVs)
Chapter 5
Summary, Conclusions, and Recommendations

Summary and Conclusions

Wetlands provide great benefits nationwide; adequate food habitat to endangered species, resting stations for migratory birds, function as a kidney to filter nutrients as well as sediment, and in some cases protection to downstream water bodies. Mud Lake is a model case for similar wetlands; it is a unit of the Bear Lake Wildlife Refuge operated by the US Fish and Wildlife Service; MudLake functions as a fine sediment trap and a nutrient filter for flows into the downstream Bear Lake. It also acts as a habitat for endangered species and migratory birds. Mud Lake is characterized by complex hydraulics and water allocation operations; which raises concerns from Refuge managers to the way the Mud Lake is operated.

Natural systems can be complex environmental habitats, which are considered a challenge to modeling the environmental parameters that are affecting these systems. The objective of this dissertation was to: 1- observed and verify the presence of patterns (spatial and temporal) of water quality and suspended fine sediment in the collected data in Mud Lake, 2- test and verify a novel statistical framework, MVRVM, capability to model suspended fine sediment and water quality constituents in Mud Lake, and 3-verify the ability of the statistical tools to select relevance vectors that can solve the practical problem of the quantity, and spatial and temporal distribution of data that will be sufficient to understand and model this natural system. The success in this objective can
serve as a tool for the refuge managers to forecast the water quality constituents and thus inform decision making.

Chapter 2 is an introduction for the case study of Mud Lake and provides general information about the challenges in MudLake, as well as the experimental design to collect data such as GIS data, lake boundaries, bathymetry data, velocity profiles, water quality data and suspended fine sediment concentration. The hydraulic scenarios for operating Mud Lake were presented. In this chapter we introduced the methodology of collecting preliminary observations in MudLake to support decisions on how many sampling locations are to be selected for consideration. We presented the observations for the selected water quality constituents at all the sampling locations, and demonstrated the existence of patterns. We showed that the sampling locations in MudLake can grouped in zones according to the range of observations for each water quality type and the behavior in every parameter. Statistical percentile results (25th, 50th and 75th percentile) for all the parameters in different zones were presented to establish an understanding of the situation in MudLake. The different water quality constituents’ relationships were confirmed among their observations. As an example, the increase of DO is directly proportional to increase of pH and inversely proportional with the turbidity. In this chapter a short demonstration of the statistical learning tools as ANN and SVM was included which clarified that these statistical learning tools require a lot of data; hence our only choice is to use the MVRVM.

Chapter 3 presents of ANN and SVM as statistical learning tools within their historical context, and the efforts of the scientists to use them to model and predict data in hydraulics, and suspended fine sediment transport were mentioned. This provided
evidence to use the MVRVM; however the MVRVM is a relatively new algorithm that has never been used to model suspended fine sediment transport and water quality constituents in natural systems. We presented the framework and experimental design for constructing the MVRVM. We explained in this chapter that there were sources of errors in the modeling output which might have resulted from overlooking constituents that may affect the observations, as in the case of algae and other vegetation growth on pH and DO. We demonstrated the output from the MVRVM model for the water quality constituents against the collected observations, the RMSE, and the residuals in all of the locations. We also presented the selected numbers of RVs for each parameter tested and the percentage of total observation they represent. We have established that complex patterns in observations are responsible for increasing the number of relevance vectors. Examination of the selected water quality residuals and the RMSE supported the assumption that the MVRVM is able to model these constituents in cases of limited observations in a complex natural system.

In Chapter 4 we indicated that the RVM as an algorithm has proven success to model different applications, however few scientists considered the practical use of the selected RVs in selecting the relevant samples and locations. Hence the locations for the selected RVs for each water quality constituent were presented. The RVs showed similarities for significant locations for some of the parameters. In this chapter we explain the significance of arranging the MVRVM training data in time series for each location with respect to the parameter tested. It was found that more than 60% of the RVs are located in Zone 1 which is the upstream zone of the Lake corresponding to the source of sediment flow in the Lake. The locations of RVs supported the experimental design for
collecting observations which was the foundation for observing that patterns exist in the observations. The RVs also showed that the number of observations that can be used to carry out modeling can be reduced by 50 to 70% in case of less complex patterns, or only save less than 5% of the data in cases of very complex patterns.

In general the careful choice of the experimental design can change the meaning of MVRVM output. The output of the model has emphasized the selection of the MVRVM that have the advantage of their simple calibration, and input process which can help US Fish and Wildlife managers benefit from this framework.

**Recommendations for Future Work**

The work presented in this research describes an effort to develop simple statistical tools that can be used to model environmental quality constituents and suspended fine sediment transport in complex natural system, thus provide a simple frame work for managers of US Fish and Wildlife Refuge Agencies to use and make decision to preserve the function of the refuge for species that inhabit it. The spatial observations used in this research were collected especially for the purpose due to lack of historical data in MudLake. There is a need to collect more velocity vector observations because this data helps in improving the understanding of the hydraulics.

Future research can be focused on identifying the amount of nutrient phosphorous that flow into MudLake since nutrients are often attached to sediment particles. The MVRVM can be used in the same way of this research to model the spatial distribution of phosphorous.
Future research can also focus on considering algae and vegetation effects on the observation of pH and DO; observations related to their effect can be used as another parameter and modeled in the same way as other parameters in this research.

More efforts should be focused to: 1-address and verify the success of RVs in decision making with regard to sampling location and understanding the concept; whether it can really provide an estimate of the amount of data required for modeling in complex natural systems. 2- collect field observations for the velocity variable to use in the MVRVM to enhance the understanding of how the contaminants circulate.
CHAPTER 6
ENGINEERING SIGNIFICANCE

Wetlands are areas that can be partially or fully covered with water much of the time. Wetlands are multitasking; they act as flood control, sediment trap, filter for impurities, source of food for growing species, and vital habitat for 35% of endangered species. Mud Lake is a unit from Bear River wildlife refuge, which is considered to be a typical example of a wetland; that acts as sediment trap to Bear River, and habitat for thousands of migrating birds. The refuge managers have been concerned about the best means to attain adequate habitat for wildlife, and improve the quality of water in Mud Lake.

The collection of data concerning several water quality constituents confirmed that the constituents were below the adequate range to support aquatic life on more than one occasion. Development of a simple framework to use by the refuge managers was an objective to help them to prepare and investigate ways to enhance the conditions in Mud Lake. An open source statistical Learning tool MVRVM was considered for this case. The MVRVM has been used to model various applications in hydraulics and hydrology with small number of observations and have proven accuracy.

The MVRVM has proven an ability to accurately model, and capture spatial hidden patterns in water quality constituents, and turbidity in Mud Lake with very limited observations. The careful construction of the experimental design and arrangement of observations for code runs with MVRVM revealed the significant locations to collect
observations. The success of the MVRVM should encourage the refuge managers to use the same methodology to find the best scenarios to preserve the function of MudLake.

When using the MVRVM for modeling the spatial water quality constituents in Mud Lake or any other lake. A few considerations should be sought in case of using the MVRVM as an example: 1- collection of new observations to be added to the training of the algorithm that reflect extreme cases; which were not considered during the development of the model. 2- depending on the accuracy required, collection of metrological observations should be considered to correct for the corresponding unexplained errors. 3- Identification of algae and photosynthesis effects on DO and pH to be included as correction parameters to enhance the modeling results of DO and pH. 4- Collection of more accurate velocity vector magnitude observations to promote the understanding of circulation of different water quality constituents. Economically the MVRVM can be used to select significant locations for sample collections, thus decrease the cost for data collection.

To conclude this part; as an advice for scientists or professionals seeking the use of the MVRVM is to careful construct their experimental design; to be based on the objectives required from the study or the assignment. Because this statistical tools are data driven tools and without proper understanding for the way the parameters affect each other; can lead to unexplainable results.
APPENDICES
APPENDIX A.

Coordinates of observation locations in Mud Lake

<table>
<thead>
<tr>
<th>Sample location</th>
<th>UTMX</th>
<th>UTMY</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>473148.3</td>
<td>4668851</td>
<td>42° 10' 16.854&quot; N</td>
<td>111° 19' 30.409&quot; W</td>
</tr>
<tr>
<td>1</td>
<td>473077.4</td>
<td>4668873</td>
<td>42° 10' 19.407&quot; N</td>
<td>111° 19' 33.918&quot; W</td>
</tr>
<tr>
<td>2</td>
<td>473258.7</td>
<td>4668512</td>
<td>42° 10' 6.955&quot; N</td>
<td>111° 19' 25.799&quot; W</td>
</tr>
<tr>
<td>3</td>
<td>473136.7</td>
<td>4668387</td>
<td>42° 10' 3.805&quot; N</td>
<td>111° 19' 30.900&quot; W</td>
</tr>
<tr>
<td>4</td>
<td>473715.5</td>
<td>4665636</td>
<td>42° 8' 31.854&quot; N</td>
<td>111° 19' 5.052&quot; W</td>
</tr>
<tr>
<td>5</td>
<td>473535.4</td>
<td>4665681</td>
<td>42° 8' 36.241&quot; N</td>
<td>111° 19' 13.177&quot; W</td>
</tr>
<tr>
<td>6</td>
<td>473874.6</td>
<td>4664734</td>
<td>42° 8' 4.150&quot; N</td>
<td>111° 19' 0.250&quot; W</td>
</tr>
<tr>
<td>7</td>
<td>473723.5</td>
<td>466482</td>
<td>42° 7' 57.430&quot; N</td>
<td>111° 19' 4.621&quot; W</td>
</tr>
<tr>
<td>8</td>
<td>475541.2</td>
<td>46703414</td>
<td>42° 11' 5.642&quot; N</td>
<td>111° 17' 46.331&quot; W</td>
</tr>
<tr>
<td>9</td>
<td>475551.5</td>
<td>4670286</td>
<td>42° 10' 50.531&quot; N</td>
<td>111° 17' 59.123&quot; W</td>
</tr>
<tr>
<td>10</td>
<td>475264.3</td>
<td>4669900</td>
<td>42° 11' 3.389&quot; N</td>
<td>111° 17' 45.781&quot; W</td>
</tr>
<tr>
<td>11</td>
<td>475282.4</td>
<td>4669826</td>
<td>42° 10' 48.982&quot; N</td>
<td>111° 17' 57.500&quot; W</td>
</tr>
<tr>
<td>12</td>
<td>475579.1</td>
<td>4670000</td>
<td>42° 10' 55.328&quot; N</td>
<td>111° 17' 44.329&quot; W</td>
</tr>
<tr>
<td>13</td>
<td>475166.1</td>
<td>4668973</td>
<td>42° 10' 21.659&quot; N</td>
<td>111° 18' 2.691&quot; W</td>
</tr>
<tr>
<td>14</td>
<td>476103.1</td>
<td>4667946</td>
<td>42° 9' 48.905&quot; N</td>
<td>111° 17' 21.734&quot; W</td>
</tr>
<tr>
<td>15</td>
<td>475492.4</td>
<td>4667913</td>
<td>42° 9' 47.846&quot; N</td>
<td>111° 17' 48.122&quot; W</td>
</tr>
<tr>
<td>16</td>
<td>475133.6</td>
<td>4667222</td>
<td>42° 9' 25.319&quot; N</td>
<td>111° 18' 3.636&quot; W</td>
</tr>
<tr>
<td>17</td>
<td>474274.4</td>
<td>4667201</td>
<td>42° 9' 23.471&quot; N</td>
<td>111° 18' 41.059&quot; W</td>
</tr>
<tr>
<td>18</td>
<td>475861.1</td>
<td>4666648</td>
<td>42° 9' 5.896&quot; N</td>
<td>111° 17' 31.342&quot; W</td>
</tr>
<tr>
<td>19</td>
<td>477086.5</td>
<td>4665428</td>
<td>42° 8' 25.767&quot; N</td>
<td>111° 16' 38.022&quot; W</td>
</tr>
<tr>
<td>20</td>
<td>477397.1</td>
<td>4664850</td>
<td>42° 8' 7.582&quot; N</td>
<td>111° 16' 24.660&quot; W</td>
</tr>
<tr>
<td>21</td>
<td>477973</td>
<td>4664189</td>
<td>42° 7' 46.221&quot; N</td>
<td>111° 15' 59.480&quot; W</td>
</tr>
<tr>
<td>22</td>
<td>477543.9</td>
<td>4662779</td>
<td>42° 7' 0.479&quot; N</td>
<td>111° 16' 17.977&quot; W</td>
</tr>
<tr>
<td>23</td>
<td>476922.7</td>
<td>4663852</td>
<td>42° 7' 35.200&quot; N</td>
<td>111° 16' 45.180&quot; W</td>
</tr>
<tr>
<td>24</td>
<td>476589.6</td>
<td>4663497</td>
<td>42° 7' 23.630&quot; N</td>
<td>111° 16' 59.639&quot; W</td>
</tr>
<tr>
<td>25</td>
<td>475273.9</td>
<td>4663310</td>
<td>42° 7' 17.442&quot; N</td>
<td>111° 17' 57.158&quot; W</td>
</tr>
<tr>
<td>26</td>
<td>476081.4</td>
<td>4664394</td>
<td>42° 7' 52.685&quot; N</td>
<td>111° 17' 21.905&quot; W</td>
</tr>
<tr>
<td>27</td>
<td>474061.9</td>
<td>4664699</td>
<td>42° 8' 2.404&quot; N</td>
<td>111° 18' 26.395&quot; W</td>
</tr>
<tr>
<td>28</td>
<td>475499.7</td>
<td>4665592</td>
<td>42° 8' 31.434&quot; N</td>
<td>111° 17' 47.420&quot; W</td>
</tr>
<tr>
<td>29</td>
<td>474895.5</td>
<td>4666196</td>
<td>42° 8' 51.138&quot; N</td>
<td>111° 18' 13.837&quot; W</td>
</tr>
</tbody>
</table>
APPENDIX B

Calibration of Hydrolab (Jim Millesan/ Utah Water Research Lab)

1. The Hydrolab is calibrated against known standard solutions and/or values before data collection with maximum 24 hours.

2. If the Surveyor does not have built-in barometric pressure, obtain the local reading from the Utah Climate Center. A value corrected to sea level and reported in inches can be converted by subtracting 4.55 inches, then multiplying by 25.4 to give mm.

3. Dissolved Oxygen sensor is calibrated by drying the sensor and allowing it to measure saturated air (100%) by placing an inch of water in the calibration cup, and correcting for the current barometric pressure.

4. Specific Conductivity probe is calibrated against a known standard of 718 us/cm.

5. pH probe is calibrated using both a 7.0 pH buffer solution and a 10.0 pH buffer solution, which brackets the anticipated pH values of 7-8.5 in local waters.

6. Turbidity probe is calibrated against a known standard of 0.0 NTU (distilled water) and 100.0 NTU.
Appendix C

U.S. Fish & Wildlife Service

Bear Lake National Wildlife Refuge
Oxford Slough Waterfowl Production Area
Planning Update Number 1, June 2010

Comprehensive Conservation Planning Begins at Bear Lake Refuge

Bear Lake National Wildlife Refuge (NWR, Refuge) is initiating a planning process to develop a Comprehensive Conservation Plan (CCP) that will guide management of the Refuge over the next 15 years. This is the first in a series of planning updates we will distribute to keep you informed and invite your participation in the planning process.

When we use the term “Refuge,” we include the Thomas Fork Unit of Bear Lake NWR and Oxford Slough Waterfowl Production Area (WPA). As we work through this two-year CCP process, we have the opportunity to look at the Refuge’s management from fresh perspectives. The purposes of the Refuge will remain the same as when they were established—primarily to provide habitat for waterfowl and other migratory birds.

Through the planning process, however, we will review our management of habitats such as wetlands, meadows, agricultural lands, and riparian areas; and each of our public use programs, including wildlife observation and photography, hunting, environmental education, and interpretation.

Your insights and observations are needed to provide us with a more complete and thoughtful process. We invite you to share your ideas with us by attending a public open house in Montpelier, Idaho, on July 1, 2010, or by submitting written comments by July 21, 2010 (see contact information, see page 6). Your thoughts are important to the success of this effort!

—Ansette de Kuijff, Refuge Manager

In this Update:
Refuge Overview .............................................2
Preliminary Goals .............................................3
Public Use ...................................................3
Management Issues .......................................4
Map ...........................................................5
Open House ..................................................6

Public Open House Meeting
We will hold a public open house on July 1, 2010, in Montpelier. See page 6 for details.

Your participation is critical for a successful planning process!
Refuge Overview

Bear Lake NWR was established in 1968, when management of 17,000 acres of public land in Idaho’s Bear Lake Valley was transferred to the U.S. Fish and Wildlife Service (Service) for the purposes of protecting and managing habitat for migratory birds. This area, the “Dingle Marsh,” was known to be an important nesting area for white-faced ibis, herons, egrets, gulls, terns, grebes, ducks, and geese. In most years, about 75% of the current 19,065-acre refuge is open water and marsh; the remainder consists of grasslands, wet meadows, and steep shrub-covered slopes. Wetlands are managed as a complex of habitats, meeting the needs of birds with very different life history requirements.

Historically the Bear Lake Valley was a hunting ground for the Shoshone, Ute, and Bannock tribes. The valley was settled in 1864 by members of the Church of Jesus Christ of Latter Day Saints. In the early 1900s the Telluride Canal Company developed a diversion system that connected the Bear River to Bear Lake, and allowed a significant portion of the river’s flow to be stored in Bear Lake for future irrigation use. The project significantly altered the hydrology and ecological processes of the Bear Lake watershed.

The 1,015-acre Thomas Fork Unit (TFU) is a satellite of Bear Lake NWR, located near the town of Border, Wyoming. Land for the unit was transferred to the Service from the Farm Home Administration in 1996, to be used for conservation purposes. The TFU consists primarily of wetland habitat, including a 2.75-mile riparian zone along Thomas Fork Creek. The TFU provides important habitat for sandhill cranes, other migratory birds, and Bonneville cutthroat trout. The area was historically used by settlers traveling the Oregon Trail as they attempted to ford the Thomas Fork Creek.

The 1,878-acre Oxford Slough WPA was purchased from the Federal Land Bank in 1985, using Federal Duck Stamp funds for the purpose of preserving small natural wetlands and associated uplands. It is located near the town of Oxford, Idaho, and is the only waterfowl production area in Region 1 of the Service. The Oxford Slough is dominated by a deep bulrush marsh, but also has a diversity of shallow marsh, wet meadows and drier alkali uplands, with some farm units.
Preliminary Goals

Goals are broad statements intended to provide direction for future management of the Refuge. They are based on Refuge purposes, the mission and policies of the National Wildlife Refuge System, input received through this planning process, and key issues identified as most significant to the Refuge. The planning team has developed the following preliminary goals for the Refuge and would like your input on them.

Preliminary Wildlife and Habitat Goals

Goal 1: Wetland Management

Stimulate functional values and recover natural hydrologic regimes using wetland management tools, while providing consistent geographic availability of Refuge wetland habitat on an annual basis.

Goal 2: Riparian Management

Provide high quality riparian habitat within the watershed for the life history requirements of focal wildlife species, while simulating natural environmental processes.

Goal 3: Native Upland Management

Maintain and protect the existing integrity of functional early successional upland habitat and restore the natural range of variability and resilience in late successional upland habitat.

Goal 4: Non-Native Agriculture Management

Provide a supplemental on-Refuge forage base for herbivores and protect the requirements of migratory waterfowl and landbirds within the Bear River watershed corridor.

Goal 5: Invasive Species Management

Prevent ecological resistance and a rapid management response to the invasive threat of invasive species within native Refuge habitats.

Goal 6: Inventory and Monitoring

Utilize inventory, monitoring, surveys, and research to gather scientifically sound information to support adaptive management decisions and management.

Goal 7: Wildlife Dependent Recreation and Public Use

Increase public understanding and appreciation of wildlife, and build support for Bear Lake NWR and adjacent lands. W53 by providing opportunities for all visitors to participate in safe, quality wildlife-dependent recreation and education programs, while minimizing wildlife disturbance.

Goal 8: Historic and Cultural Resources

Increase the understanding and appreciation of the unique historic, archaeological, and cultural resources of the Bear Lake Watershed by Tribes, local communities, and visitors alike.

Goal 9: Land Protection and Acquisition

Secure key habitat areas through cooperative management arrangements and/or purchase of conservation easements on a voluntary basis.

Public Use of the Refuge

The National Wildlife Refuge System Administration Act of 1966, as amended, identified six priority refuge uses: hunting, fishing, wildlife observation and photography, and environmental education and interpretation. These uses receive enhanced consideration in planning and management over all other general public uses on refuges. When compatible, these wildlife-dependent recreational uses are to be strongly encouraged. These uses, as well as other current or proposed uses, receive an extensive compatibility review in the CCP before being allowed. Under Service compatibility policy (403FW2), refuges with limited staffing and funding are required to make efforts to obtain additional resources or outside assistance to provide wildlife-dependent recreational uses, and to document those efforts before determining that any of these uses are not compatible.

Bear Lake Refuge must manage ever-increasing requests for visitation and demand for visitor services programs with a very small staff. Currently, the visitor services and public hunting program at the Refuge is mostly "self serve", with informational kiosks and interpretive displays. To date, the visitor services emphasis is placed on maintaining visitor and hunter facilities, welcoming and orienting visitors, answering information requests, and law enforcement during the hunting season.

Questions to Consider

- Should existing public uses be continued, reduced, or eliminated?
- Should the Refuge improve its visitor services program?
- What actions should be taken to minimize wildlife disturbance issues from public visitation and recreation?
Management Issues, Challenges, and Opportunities

As part of the CCP process, a range of possible alternative management approaches will be explored and evaluated, including current management practices. The effects of the various alternatives on the biological resources and local communities will be evaluated in an Environmental Assessment (EA) that is prepared concurrently with the CCP, in accordance with the National Environmental Policy Act. The planning team has identified some potential issues to be considered. We encourage you to provide us with written comments on these issues and other concerns, and to meet us at our public open house. If you aren’t able to attend the open house on July 1, please submit your comments in writing by July 21, 2010 (see the back page for contact information.)

Habitat Management

Water level management is the overriding factor affecting most Refuge habitat management strategies for nesting birds and wildlife, particularly water birds and muskrats. Management efforts focus on maintaining a given emergent-marsh-to-open-water-habitat ratio using water level manipulations, prescribed fire, and mechanical disturbance.

Riparian habitats comprise a small but important component of Refuge ecosystems. Native fishes historically present within the Refuge waters include Bonneville cutthroat trout. Since these species do not originate on Refuge lands and significant portions of the watersheds lie outside the Refuge, upstream activities have major impacts on Refuge water quality and quantity.

Widespread population and habitat declines have been projected for numerous sagebrush associated species. A growing sense of urgency over the outlook for sagebrush dependent wildlife has spawned sagebrush planning and restoration efforts within Idaho.

Agricultural small grains and short-cover areas at the Refuge provide valuable foraging habitat for key bird species such as cranes, geese, and curlews.

Questions to Consider

- What are the best means to attain productive marsh habitats for Refuge wildlife?
- How can the Service protect and improve the quantity and quality of Refuge water for fish and wildlife resources?
- What can the Service do to prevent the introduction and dispersal of invasive plants and animals and facilitate their removal from the Refuge?
- What should the Refuge’s role be in supporting native fish and riparian habitat restoration?
- What are the most appropriate management techniques for the Refuge’s wet meadow and upland habitats to maximize habitat values for key wildlife species (e.g., sandhill cranes, Canada geese), while assuring other native wildlife cover and forage requirements are still satisfied?
- What is the appropriate role of prescribed fire in habitat management and fuels reduction?
- Should the Refuge attempt cooperative and joint watershed management strategies within Bear Lake and Bear River watershed?
- How can the Refuge engage or adaptively manage in response to predicted and unexpected challenges posed by climate change?
- Given limited budgets and manpower, how can the Refuge most appropriately assess the efficacy of management actions at the appropriate scale?
June 2010

We Are Interested in Your Views about Bear Lake National Wildlife Refuge
We value your input as we work to prepare a Comprehensive Conservation Plan for Bear Lake National Wildlife Refuge. If you have a few minutes to respond, it will help us to identify issues, concerns, and opportunities for the Plan.

Why is Bear Lake National Wildlife Refuge special to you?

Which activities would you like to enjoy at the Refuge over the next fifteen years? Check all that apply.

- Bird watching/observing wildlife
- Photographing wildlife or scenes of nature
- Hiking or walking
- Driving/auto touring
- Canoeing
- Cross-country skiing/snowshoeing
- Waterfowl hunting
- Upland game hunting
- Small game hunting
- Fishing
- Environmental Education
- Interpretation
- Other (please list):

What issues, concerns, or opportunities should the Refuge address in its Comprehensive Conservation Plan?

What suggestions do you have to address your issues of concern?

Thank You! Please respond to us by July 21, 2010.
You may mail this form, fax it to (208) 847-1319, or drop it off at the Refuge office. You may also submit comments by email to FW1PlanningComments@fws.gov. Please type “Bear Lake CCP” in the subject line.

All comments received from individuals, including names and addresses, become part of the official public record and may be released. Requests for such comments will be handled in accordance with the Freedom of Information Act, the Council on Environmental Quality's NEPA regulations (40CFR1506.6) and other Service and Departmental policies and procedures.

Paperwork Reduction Act Statement: The U.S. Fish and Wildlife Service will use this information to better serve the public. There is no requirement to provide a response or to use this form. Response is voluntary, and providing your name, organization and address is optional.
Would you like to remain on the mailing list to receive subsequent information about this project?

_____ Yes  _____ No

If yes, please provide your contact information:

Name:

Street Address:

City, State, Zip:

Additional Comments?

Fold here and tape ends

________________________________________________________

________________________________________________________

________________________________________________________

U.S. Department of the Interior
Fish and Wildlife Service
Bear Lake National Wildlife Refuge
P.O. Box 9
Montpelier, ID 83254
**Refuge Open House**

**July 1 in Montpelier**

---

### How Do I Contact the Service or Provide Comments?

To be included on the mailing list, provide comments, ask questions, or request information, please contact:

<table>
<thead>
<tr>
<th>By Mail:</th>
<th>By Phone:</th>
<th>By Fax:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bear Lake National Wildlife Refuge</td>
<td>208-847-1757</td>
<td>208-847-1319</td>
</tr>
<tr>
<td>370 Webster Box 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Montpelier, Idaho 83254</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

By Email: FWIPlanningComments@fws.gov -or- annette.deknijf@fws.gov

(please place “Bear Lake NWR CCP” in the subject line.)

**Please provide comments by July 21, 2010.**

---

### Planning Schedule

(Schedule dates are tentative and subject to change as the planning process progresses.)

<table>
<thead>
<tr>
<th>Planning Update 1</th>
<th>June 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public Open House</td>
<td>July 1, 2010</td>
</tr>
<tr>
<td>Planning 2</td>
<td>August 2010</td>
</tr>
<tr>
<td>Planning Update 3</td>
<td>November 2010</td>
</tr>
<tr>
<td>Preliminary Alternatives</td>
<td></td>
</tr>
<tr>
<td>Public Review/Comment on Draft CCP/EA</td>
<td>April 2011</td>
</tr>
<tr>
<td>Final CCP</td>
<td>May 2012</td>
</tr>
</tbody>
</table>

---

You are invited to the Open House!

We’d love to see you at our upcoming public meeting. It’s our chance to hear your thoughts about management of the Refuge over the next fifteen years.

**July 1, 2010**

Bear Lake County Senior Citizens Center

115 S. 4th St.

Montpelier, Idaho 83254

6:30 - 8:30 pm
CURRICULUM VITAE

HUSSEIN BATT
Civil Engineer, Environmental Engineer, Water Resources Engineer

Summary
- Civil and Environmental Engineer offering more than 4 years of experience in Civil & Environmental Engineering.
- Participated in hydraulic design and modeling for several water resources systems.
- Participated in Data collection and modeling for water quality constituents, and air quality.
- Focusing on successful assessment of risk management skills including sampling, designing, and decision making of Environmental and hydraulic systems.
- Successful construction design and field work.
- Strong written and verbal communication skills.
- Ability to see tasks through to completion.
- A team player
- Patience, a positive attitude and a willing to learn if required.

Professional Background

Civil Engineer "Design Department", Ministry of Water Resources, Egypt.
- Hydraulic modeling of canals, design of hydraulic structures, and geological stabilization of soils.
- Hydraulic design of 12 miles conveyance system including all the relevant works on the canal system; complete flow measurements north of Delta District, water quality sampling and analysis.
- Flux measurement and air quality modeling in delta districts.
- Investigated, regulatory activities involving domestic water supply and waste water reuse.
- Reviewed and supervised implementation plans for waste treatment plants.

Civil Engineer "Construction Department", Ministry of Water Resources, Egypt.
- Assisted in preparing the implementation vision for Integrated Water Resources Management plan, including participating in the supplemental studies between different water resources department and preparing the detailed visions and strategies.
- Construction and field supervision of hydraulic structures in Upper and Delta of Egypt.
- Water quality sampling and management of various chemicals in the Nile Delta.
- Collected and modeled air quality pollutants.

Graduate Research Assistant “Environmental Engineering Division” Utah Water Research Laboratory, Logan, Utah.
- Create surveys to analyze the impacts of new water resources projects on water quality and residents.
- Statistically analyze new water resources projects and impacts assessment on human, economic, natural resources dimensions.
- Managing Mud Lake as wild life refuge and enhance its function as sediment filter.
- Sample collection for water quality constituents in Mud Lake.
- Analysis of water quality constituent patterns, and create mathematical model to forecast future patterns.
• Create public presentations, and reports to the public.

Teaching Experience

Teaching Assistant, National Water Research Center, Egypt.
Hydrology, Water Quality

Education

• Philosophy Degree, Civil and Environmental Engineering “Environmental Engineering”, Utah State University, Logan, Utah.
• Master of Science, Civil and Environmental Engineering “Irrigation Engineering”, Utah State University, Logan, Utah.
• Bachelor of Science, Civil Engineering, Cairo University, Cairo, Egypt.

Software Applications

<table>
<thead>
<tr>
<th>VB.Net</th>
<th>ResSim</th>
<th>HydroIIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ArcGis</td>
<td>Sap2000</td>
<td>HecHMS</td>
</tr>
<tr>
<td>Drain mode</td>
<td>HecRas</td>
<td>SWMM</td>
</tr>
<tr>
<td>Autocad</td>
<td>Matlab</td>
<td>QUAL2k</td>
</tr>
<tr>
<td>Erdas imagine</td>
<td>MDSWMS</td>
<td>R</td>
</tr>
<tr>
<td>Dambrk</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Related Skills

| Hydraulic Modeling | Sedimentation Engineering and Management |
| Damage Assessment/Analysis | Statistical Analysis |
| CAD Drafting | Environmental Water Quality analysis |
| Policy and Procedures Development | Computer Programming/Modeling |
| Environmental Regulations and Compliance | Structural Design & Analysis |
| Crisis Management and Emergency Response | Crew Supervision |
| Bio-solids Management | Watershed Management |
| Waste Water Management | Irrigation, Drainage systems Design / Management |
| Remote Sensing | |

Publications


Training & Honors

- *Utah Water Research Laboratory (UWRL) Fellowship*, Utah State University, Logan, UT 2007-present.
- *Top Academic Achievement*, Golden Key International Honor Society, Utah State University, 2008.