12-2011

Environmental Sensor Anomaly Detection Using Learning Machines

Erick F. Conde
Utah State University

ENVIRONMENTAL SENSOR ANOMALY DETECTION

USING LEARNING MACHINES

by

Erick F. Conde

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

of

MASTER OF SCIENCE

in

Civil and Environmental Engineering

Approved:

Dr. Mac McKee Major Professor	Dr. David K. Stevens Committee Member
Dr. Jeffery S. Horsburgh Committee Member	Dr. Mark R. McLellan Vice President for Research and Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2011

ABSTRACT

Environmental Sensor Anomaly Detection

Using Learning Machines

by

Erick F. Conde, Master of Science

Utah State University, 2011

Major Professor: Dr. Mac McKee
Department: Civil and Environmental Engineering

The problem of quality assurance/quality control (QA/QC) for real-time measurements of environmental and water quality variables has been a field explored by many in recent years. The use of *in situ* sensors has become a common practice for acquiring real-time measurements that provide the basis for important natural resources management decisions. However, these sensors are susceptible to failure due to such things as human factors, lack of necessary maintenance, flaws on the transmission line or any part of the sensor, and unexpected changes in the sensors' surrounding conditions.

Two types of machine learning techniques were used in this study to assess the detection of anomalous data points on turbidity readings from the Paradise site on the Little Bear River, in northern Utah: Artificial Neural Networks (ANNs) and Relevance Vector Machines (RVMs). ANN and RVM techniques were used to develop regression models capable of predicting upcoming Paradise site turbidity measurements and estimating confidence intervals associated with those predictions, to be later used to determine if a real measurement is an anomaly. Three cases were identified as important to evaluate as possible inputs for the regression models created: (1) only the reported values from the sensor from previous time steps, (2) reported values from the sensor from previous time steps and values of other water types of sensors from the same site as the target sensor, and (3) adding as inputs the previous readings from sensors from upstream sites.

The decision of which of the models performed the best was made based on each model's ability to detect anomalous data points that were identified in a QA/QC analysis that was manually performed by a human technician. False positive and false negative rates for a range of confidence intervals were used as the measure of performance of the models.

The RVM models were able to detect more anomalous points within narrower confidence intervals than the ANN models. At the same time, it was shown that incorporating as inputs measurements from other sensors at the same site as well as measurements from upstream sites can improve the performance of the models.

(94 pages)

PUBLIC ABSTRACT

Environmental Sensor Anomaly Detection

Using Learning Machines

by

Erick F. Conde, Master of Science

Utah State University, 2011

Major Professor: Dr. Mac McKee
Department: Civil and Environmental Engineering

The search for improvements in the quality assurance/quality control (QA/QC) of real-time environmental measurements has been a field well exploited in recent years. These measurements describe actual environmental conditions and processes that provide relevant information upon which water quality management decisions are based. *In situ* sensors (located at the site of interest) are commonly used for such real-time measurement purposes. However, the performance of these types of sensors can be affected by such things as human factors, lack of necessary maintenance, flaws on the transmission line or any part of the sensor, and unexpected changes in the sensors surrounding conditions. These issues have increased the importance of the early detection of anomalous data points within a recorded time series.

This research focuses on the detection of anomalous data points on turbidity readings from the Paradise site on the Little Bear River, in northern Utah. To do so, two machine learning techniques were used: Artificial Neural Networks (ANNs) and

Relevance Vector Machines (RVMs). These techniques were used to develop regression models capable of predicting (with determined confidence intervals) what the next Paradise turbidity time step value should be. The ANNs have displayed good performance for this type of prediction but the RVMs have not been tested yet on the real-time anomaly detection problem. Since for other related applications the RVMs consistently displays better results than the ANNs, there is a motivation for this research to deeply explore that technique.

This research also addressed the possibility of improving results based on evaluating a broader combination of inputs. Three cases were identified as important: (1) only the reported values from the sensor from previous time steps, (2) reported values from the sensor from previous time steps and values of other water types of sensors from the same site as the target sensor, and (3) adding as inputs the previous readings from sensors from upstream sites.

Points detected as anomalous by the models were compared to data points obtained from a QA/QC analysis performed by a human technician. This allowed obtaining the rate of success of the models which was later express on a false positive and false negative basis.

Results determined that the inclusion as input of measurements from other sensors at the same site as well as measurements from upstream sites can improve the models performance. Also, it was shown that RVM models detected more anomalous points within narrower confidence intervals than the ANN models.

(94 pages)

ACKNOWLEDGMENTS

I would like to thank God, for giving me the strength and knowledge to carry on this task. I would also like to dedicate this work to my family, Maria my mother, Fernando my father, and my sisters (Adriana and Desirée) for supporting me with their unconditional love. Special thanks to Maridalia and all my friends for making hard times easier to handle and supporting me trough this journey.

I would like to specially thank my advisor, Dr. Mac McKee, for making it possible to develop this idea and for his constant support and patience throughout this entire process. To my committee members, Dr. Horsburgh and Dr. Stevens, for their guidance, corrections, and important discussions. To Dr. Urroz for his great advising when needed. To my Peruvian friends, Alfonso and Andres, for their important ideas and constant support. I would also like to thank Shelly Hernandez for her support and guidance, the Dominican Republic government, and the Utah State University Water Research Laboratory for providing the funds for this project. Thanks to all!

Erick F. Conde

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 The Problem

Water quality and other types of real-time field data are constantly being acquired in order to address environmental management and research issues. The data obtained from these readings are often used to support decisions about the management of natural resources that affect human welfare and environmental conditions. Measurements such as pH, turbidity, specific conductance, dissolved oxygen, temperature, and stream flow help us understand hydrologic and water quality processes that might be active in the watershed, and they allow us to have accurate information about field conditions and natural processes when decisions must be made.

Procedures for data collection have changed with time. Water samples used to be collected and transported to laboratories where different tests were performed according to the type of information required. While field sampling and laboratory analysis must still be done for a variety of purposes and constituents, with the development of new technologies, *in situ* data collection (through the use of sensors) is becoming a standard practice for measuring many variables in the environmental field. The use of *in situ* measurement technologies allows us to have access to field conditions in near real-time.

The operational characteristics of these sensors are constantly being improved in order to face the adverse conditions found in the field. In spite of these efforts, however, sensors are still subject to failure because of:

1. Errors introduced by installation and calibration procedures.

2. Lack of or failure to apply a maintenance policy: without proper maintenance, unexpected changes in sensor performance can decrease accuracy of the readings.
3. Damage of any part of the sensor system, including battery, wiring, datalogger, telemetry equipment, etc.
4. Changes in the sensor surrounding conditions (e.g., freezing temperatures, debris, and sediment accumulation).

The use of new sensor technologies allows us to collect large amounts of data and to archive historic records of the measured behavior at a site through time. This capability also presents new challenges for data quality assurance and quality control (QA/QC) because human resources are simply not available to track and evaluate all the measurements that are available in real-time. Lack of sufficient QA/QC can present problems especially when the real-time data being collected is also used in real-time to make operational or management decisions about water control facilities. If bad data becomes available in real-time without the availability of appropriate QA/QC, then bad real-time decisions might result. For example, if a water treatment plant is continuously monitoring turbidity levels at a water source and it misses a real spike on the readings, it wouldn’t be possible to make the necessary adjustments in response to that event jeopardizing the quality of the water to be served. These factors and similar cases to the example before described create the need to develop and implement ways of systematically detecting anomalous measurements and improve real-time quality control procedures (Moatar, Miquel and Poirel, 2001).

Several sensor failure detection methods have been developed and implemented, some with significant success. Different technologies have been applied to this problem, with some of the more successful coming from the use of machine learning applications, such as artificial neural networks (ANNs), which were used by Hill and Minsker (2006) in the examination of the time series reported from a single sensor, and by Barron, Mounstapha, and Selmic (2008) for fault detection in wireless sensor networks. Since there are often several sensors at the same site, or a variety of sensors at different locations on a stream or river system, another interesting challenge and potential opportunity for improving the performance of these types of techniques would be to evaluate patterns in the relationships among the readings of multiple sensors located at the same site in order to detect anomalies in one of them. In the same way, relationships or patterns among sensors located at nearby sites might also be exploited in order to improve detection of real-time sensor anomalies and QA/QC of the data generated by any particular sensor in the collection.

1.2 Research Objectives

The main goal of this research is to develop better methods for the detection of anomalous data in environmental sensor data streams in support of real-time application of environmental observations from *in situ* sensors. This will be accomplished by the completion of two particular objectives:

1. Development and evaluation of new regression models based on machine learning (ML) that will allow us to evaluate the likelihood that sensor measurements reported from the field are anomalous based on the time series of sensor measurements recorded from previous periods. ML techniques such as ANN and RVM have been used with success on similar problems and are expected to

provide improvements in the sensor anomaly detection in environmental measurements problem.

2. Evaluation of the potential for exploiting relationships between measurements taken by a single sensor and the measurements from other sensors located at the same site and/or nearby sites, looking forward to improvements in the regression models to be used for assessing the likelihood that sensor anomalies are correctly detected.

1.3 Anticipated Contributions

By developing this research important contributions to the anomaly detection in environmental measurements problem are expected. The most relevant ones are summarized:

1. An assessment on how efficient and reliable the ANN and RVM techniques are for the detection of anomalies in environmental sensor measurements.
2. Evaluation on the relevance to our problem of information obtained from sensors located at the same site and at sites of near that of our target sensor.
3. Recommendations on important steps to follow for future works to be develop on the anomaly detection on environmental measurements problem.

CHPATER 2

LITERATURE REVIEW

This chapter addresses the importance and history of quality assurance and quality control (QA/QC) of environmental measurements (especially water quality measurements) including the use of *in situ* water quality sensors for the measurement of different water quality properties. It provides a review of the most relevant findings in this area and it introduces theories to be implemented in this research modeling approach. Even if some of the cases presented as references for this research addressed the sensor failure/error detection problem and not just the detection of anomalous data points, they were still referenced because they contributed methods and criteria used to detect that a specific data point represented a potential anomaly case.

2.1 In Situ Water Quality Sensors

Since the mid-1980s, the use of electronic sensors to make measurements has become a common way of acquiring data about actual environmental field conditions. Advances in technology have improved sensor characteristics (e.g., size, durability, etc.) allowing them to be continuously deployed in the field for long periods of time as *in situ* sensors. Some important benefits from the use of these types of sensors are:

1. They can measure field conditions on a nearly real-time basis.
2. They are able to track conditions without regard to weather or time of day.
3. Their use can greatly reduce the number of visits to the actual site of study.

In this matter, the EPA has completed an assessment on “Sensor Technology Evaluation, Methodology and Results” regarding water quality distribution systems monitoring (Panguluri et al., 2009). In this report, the EPA recognizes the use of sensors for water quality data collection. Errors in these *in situ* water quality sensor readings, sometimes called anomalies, can mislead efforts to understand water quality conditions and can potentially jeopardize the quality of the water to be delivered. Examples of cases that can be importantly affected by anomalies in sensors readings are presented:

1. In Kansas, water quality sensors have been incorporated into a continuous, real-time monitoring system to provide estimates of constituent concentrations and loads (Christensen et al., 2003). This information is used by water suppliers to modify treatment of water and by local agencies to alert recreational water users of potential health risks. In this case, anomalies in sensor readings can put community health at risk.
2. Water quality sensors are being used in Finland to collect information for improvements on: agricultural management practices such as irrigation and pesticide spraying, monitoring algae blooms; and developing flood and frost warning systems (Kotamäki et al., 2009). Here, anomalous measurements can reduce the accuracy in estimation of the quality of the cultivated crops. More importantly, anomalous measurements can lead to the misuse of pesticides and can compromise the welfare of the population. Economic resources can also be affected by anomalies on the readings.
3. Sensor technologies are used to keep track of rainfall events affecting wastewater treatment plant procedures (Kitaoku, Yoshida, and Aoyama, 2004). Anomalies on these measurements directly affect knowledge of the quality of the water to be supplied for different purposes (e.g., domestic, industrial, recreational).

2.2 QA/QC on Environmental Measurements

The QA/QC of environmental measurements has been an issue for the past several decades. The search for precise information regarding environmental measurements began in the 1970s with the creation of national and international institutions, such as the Environmental Protection Agency (EPA), United Nations Environment Programme (UNEP) and the National Institute for Environmental Studies (NIES), aimed to globally develop and implement environmental protection programs. We can observe how this search increased in importance through the woks of Shirley (1982), who suggested that "A quality assurance program is necessary to ensure the validity and reliability of environmental measurements" while working with equipment used to measure noise and air pollutants for the California Department of Transportation. Keith (1983) defined what they considered important aspects to have in mind to properly define environmental analysis procedures, which included: level and degree of confidence of the measurements, methods of data validation, and degree of quality assurance necessary for the analysis.

As sensor technologies improved, they became an important part of the monitoring of environmental conditions. For this reason it is important that the QA/QC on environmental measurements be translated to QA/QC on environmental sensor measurements. This research will focus on the detection of anomalous data points from environmental measurements collected through the use of *in situ* sensors.

2.3 Early Research on Sensors Anomaly Detection

The use of new sensor technologies can be related to the robotics and electronics fields in the late 1970s, and some of the earliest contributions regarding the detection of anomalous data points within sensors readings came from such fields. Examples of early research relevant to our problem can be found on the following works:

1. Guo and Nurre (1991) developed methods for sensor failure detection in space shuttle main engines. The detection of failure relied on the use of redundant sensors that allowed neural networks check for consistency between the sensors outputs.
2. Xu, Hines, and Uhrig (1999) used neural networks to validate sensor calibrations and to detect sensor failure in the power generation industry. They assessed the fault detection problem by comparing the neural network predictions against the actual measurements through the use of the sequential probability ratio test (SPRT).

2.4 Sensor Anomaly Detection on Environmental Measurements

Several studies have made important contributions to the anomaly detection problems for environmental measurements. These contributions refer to methods, techniques, and approaches that were taken into consideration for developing this research.

Moatar, Fessant, and Poirel (1999) applied artificial neural networks as a new type of model to evaluate daily pH data for the Middle Loire River (France). They used this model for screening of pH measurements, error detection (abnormal values, discontinuities, and recording drifts), and validating the collected data. They compared the measured values of pH with the values estimated by the ANN using the Student t test and the cumulative Page-Hinkley test. Results from this research allowed the Environmental Defense Fund (EDF) to critically evaluate water quality parameters with respect to hydrometeorologic conditions.

Later work by Moatar, Miquel, and Poirel (2001) proposed a quality control method for examining continuous physical and chemical measurements, this time including temperature, dissolved oxygen, and electrical conductivity. This was a complex analysis using deterministic models to examine consistency in the patterns in the structure of internal data series, inter-variable relationships, and relationships with external variables. Outliers were then detected using classical statistical tests (test of mean and gradient using sliding window and Page-Hinkley cumulative tests). With all these considerations for the analysis, this research touches many important points of the approach described in this thesis, such as the evaluation of the performance of regression models when considering measurements from sensors located at the same site and sensors located at nearby sites.

Detection of failure of environmental sensors was the subject of investigation by Hill and Minsker (2006) who addressed the fault detection for *in situ* environmental sensors in an automated fashion. Their fault detection strategies were based on data-driven regression models of the time series data of individual sensors. To create the prediction models, they used four methods: naïve, clustering, perceptron, and artificial neural networks. Anomalous measurements were identified as measurements that fell outside the bounds of an established confidence interval.

Hill and Minsker (2006) developed two approaches: anomaly detection (AD) and anomaly detection and mitigation (ADAM). The ADAM method replaces a measurement detected as anomalous by the value obtained from the prediction model. The performance of the model was evaluated by reference to the rate of false positive and false negative cases identified by each of the methods.

Hill, Minsker, and Amir (2009) expanded the work on anomaly detection, this time including the use of dynamic Bayesian networks with Kalman and Rao-Black particle filtering, considerably reducing false positive/negative rates, in some cases to as low as 2%. The evaluation of new technologies when addressing the anomaly detection problem is one of the main motivations for this research.

Another approach used to detect anomalies in the measurement of stream flow compared sensor readings from upstream to downstream sites while taking into account changes in travel time. Kang et al. (2009) proposed the Smart Window Enumeration and Evaluation of persistence-thresholds (SWEET) method to efficiently explore the search space of all possible travel time window lengths. Torres, Walker, and McKee (2009) demonstrated the use of RVMs for detection and repair of error in forecasted flow rates/water levels in an irrigation canal. This work provides a clear example of the use of new machine learning theories to water management with practical applications, specifically addressing the error detection and mitigation problem.

2.5 Data-Driven Models

Considering the success obtained by previous work on sensor anomaly detection and the characteristics of the data that will be used in our case (Little Bear River Basin time series, discussed later in Chapter 3), data-driven models are appropriate for making the predictions we need for this project. Data-driven models, borrowing heavily from Artificial Intelligence (AI) techniques and statistical learning theory, are not based on an extensive knowledge of the physical process being modeled. Rather, they rely on an ability of the modeling techniques to intuit complex and possibly nonlinear input-output relationships in systems simply from the data describing system inputs and outputs. These methods are able to make abstractions and generalizations of system behaviors. Data-driven modeling uses results from such overlapping fields as data mining, rule-based approaches such as expert systems, fuzzy logic concepts, rule-induction, and machine learning systems (Solomatine, 2002).

2.5.1 Machine learning

Machine learning deals with the design and development of algorithms that allow computers to evolve behaviors based on empirical data, such as from sensor data or databases. A major focus of machine learning research is to develop models that learn to recognize complex patterns and make intelligent decisions based on data. The difficulty lies in the fact that the set of all possible behaviors given all possible inputs is too complex to describe generally in programming languages, so that in effect programs must automatically describe programs. The focus of artificial intelligence (AI) is to make machines intelligent, able to think rationally like humans and solve problems, whereas machine learning is concerned with creating computer systems and algorithms so that machines can “learn” from previous experience. Because intelligence cannot be attained without the ability to learn, machine learning now plays a dominant role in AI (Izenman, 2008).

Machine learning problems are divided into various categories. According to Izenman (2008), the two most relevant to statistics are:

1. Supervised learning: The learning algorithm receives a set of continuous or categorical input variables and a correct output variable and tries to find a function of the input variables to approximate the known output variable. A continuous output variable yields a regression problem, whereas a categorical output variable yields a classification problem.
2. Unsupervised learning: There is no information available to define an appropriate output variable. This is often referred to as "scientific discovery."

2.5.2 Artificial neural networks

Artificial neural networks (ANNs) are one of the most popular and widely used data-driven modeling techniques. ANNs are computational methodologies that perform multifactorial analyses. Inspired by networks of biological neurons, artificial neural network models contain layers of simple computing nodes that operate as nonlinear summing devices (Figure 2.1). These nodes are richly interconnected by weighted and biased links. The weights and biases are obtained when data are presented to the network during a training process. Successful training can result in ANNs that perform tasks such as predicting an output value, classifying an object, approximating a function, recognizing a pattern in multifactorial data, and completing a known pattern (Dayhoff and DeLeo, 2001).

For the regression problem, the ANN relies on the possibility of using multilayer networks which allow improvements on the level of input-output relationships. It takes the inputs, x, and assigns the initial weights and biases for the layers used, W^n, and b^n,

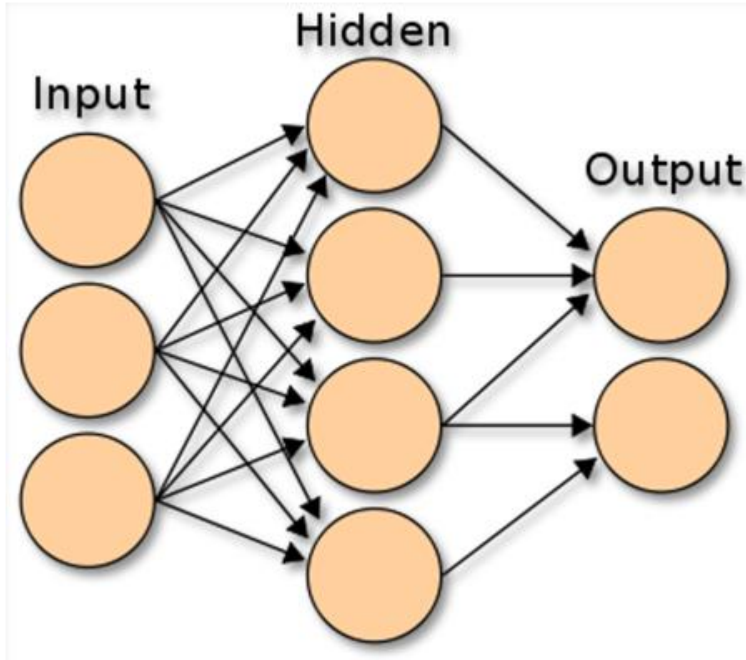

Figure 2.1. A neural network is an interconnected group of nodes, akin to the vast network of neurons in the human brain.

respectively. Through the use of feed-forward and back-propagation techniques to obtain values of the weights and biases in the training process, and employing a selected activation function (Equation 2.1) is able to obtain the predicted values, y.

$$y = W^{n+1} \cdot \tanh\left(W^{n}[x] + b^{n} \right) + b^{n+1} \tag{2.1}$$

2.5.3 The relevance vector machine

The RVM (Tipping, 2001) has shown it can obtain good results when addressing similar problems to those of our interest (Wei et al., 2005; Khalil et al., 2006). This technique relies in the use of Bayesian theory to intuit input (x) and output (y) relationships. Outputs (Equation 2.2) are determined through a linear sum of weights (w_n) and through the use of kernel functions transformations.

$$y = \sum_{n=1}^{N} w_n \cdot \phi \cdot [x] \tag{2.2}$$

The RVM accomplishes good generalization performance with exceedingly sparse predictors. This is achieved by optimizing the hyperparameters α and β shown in the next section.

2.5.4 Bayesian inference

For our case, the Bayesian approach characterizes the unknown parameter vector *w* through a probability distribution *p* (*w*) (Equation 2.3). This distribution is modified by data observation through the use of likelihood functions:

$$p(w \mid \alpha) \in \exp\{-\alpha\Omega(w)\} \tag{2.3}$$

In Equation 3, α can be regarded as a hyperparameter. If a Gaussian distribution is chosen for $p(w)$, for example, it might be chosen a Gaussian distribution for $p(w| \alpha)$ of the form presented on Equation 2.4.

$$p(w| \alpha) = \left(\frac{\alpha}{2\pi}\right)^{N/2} \exp\{-\frac{\alpha}{2} \| w \|^2\} \tag{2.4}$$

Bayes theorem can be used to express the posterior distribution for *w* as the product of the prior distribution and the likelihood function (Equation 2.5) with L(*w*) = $p(t| w, \sigma^2)$ and *t* being the target vectors.

$$p(w \mid t, \alpha, \sigma^2) \in p(w \mid \alpha) L(w) \tag{2.5}$$

In order to obtain an estimate for *w* that maximizes the posterior distribution it would be necessary to minimize the expression in Equation 2.6:

$$\frac{1}{2\sigma^2}\sum_{n=1}^{n} | y(x_n;w) - t_n |^2 + \frac{\alpha}{2}\Omega(w) \tag{2.6}$$

In the Bayesian approach predictions are made by integrating over the distribution of model parameters *w* instead of using a specific value of *w*. These integrations can be analytically intractable and require either sophisticated Markov chain Monte Carlo methods, or more recent deterministic schemes such as variational techniques, to approximate them. At the same time, the integration implied by the Bayesian framework overcomes the issue of overfitting by averaging over many different possible solutions and typically results in improved predictive capability.

Specifically, if we are given a new value of *x* then the predictive distribution for *t* is obtained from the sum and product rules of probability by marginalizing over *w* given in Equation 2.7.

$$p(t \mid t, \alpha, \beta) = \int p(w \mid t, \alpha, \sigma^2) p(t \mid w, \sigma^2) dw \tag{2.7}$$

In most applications, suitable values for the hyperparameters α and σ^2 will not be known in advance (although in some cases the noise level σ^2 may be known) and so a Bayesian treatment will introduce prior distributions over these quantities, and then eliminate them from the problem by marginalization (Bishop and Tipping, 2003).

2.6 Summary

The detection of anomalous data has been a subject of growing interest in recent times. It has been addressed in many fields (e.g., electronics, environmental) over the past few decades. Machine learning techniques have been the preferred tools to deal with these types of problems, especially ANNs. However, RVMs have consistently shown better performance than ANNs for other applications and might achieve a higher level of success with this problem. It is believed that the most promising advances can be made in the environmental sensor anomaly detection area by using more modern and powerful prediction models and by considering in the analysis measurements produced by additional sensors located at the same site and/or surrounding sites.

CHAPTER 3

RESEARCH APPROACH

This chapter presents the research approach used in this investigation, including: research premise, research question, and procedures for data collection and analysis. These elements provide a basis for modeling sensor error detection in environmental measurements.

3.1 Research Question

The main question of this research was: “Can techniques from machine learning theory for regression problems be used to improve the detection of anomalous measurements collected through the use of *in situ* water quality sensors?” In order to satisfactorily answer this question it was necessary to address other questions that helped define the modeling procedures for this research. The most relevant were:

1. Is it possible to make significant improvements from previous work done on the sensor anomaly detection problem through the implementation of recent advances in data-driven modeling?
2. Can anomalous measurements be successfully detected by such models?

3.2 Research Hypothesis

Areas with potential for future research based on the application of modern machine learning techniques were identified in the literature review. This research concentrated on the following hypothesis: “Improvements in environmental sensor quality control and quality assurance are possible through the implementation of recent data-driven methods to the forecasting of sensor measurements." This hypothesis was used as the basis for the algorithms developed and tested using data from the Little Bear River, Utah.

3.3 Research Approach

Based on previous work discussed in the environmental sensor error detection problem presented in the literature review (e.g. Moatar, Fessant, and Poirel, 1999; Hill and Minsker, 2006) and considering possible improvements to be achieved, the approach followed by this research is described:

1. Selection of the case of study (Section 3.5): this includes defining the target measurements to evaluate, training and testing data sets, and definition of the measurements included in the analysis.
2. Development of models (using ANN and RVM techniques) able to predict expected values of the sensor readings for upcoming time steps for our target data series.
3. Based on these predictions, define ranges associated with confidence intervals within which anomalous measurements would fall.
4. Determine if actual measurements are detected as anomalous and, if that is the case, define within which level of confidence they were detected.
5. Define the measure of success of the models based on established parameters (Section 3.7)

3.4 Evaluated Cases

Based on review of the literature, this research proposed improvements to the anomaly detection problem by evaluating three different cases. The idea behind this is to include available data related to our target measurements and determine if this inclusion represents improvements of the performance of the models. These cases are described as follows:

1. **Case I:** Sensor anomaly detection based on historic time series readings. For this case, regression models were built to explore relationships between a target measurement and its historical readings. This was used as a measure of how good the predictions are based only on measurements from previous time steps.
2. **Case II:** Sensor anomaly detection considering the readings of other sensors located at the same site**.** In this approach, in addition to measurements of the variable in question from previous time steps, additional readings from other sensors located at the same site were incorporated as inputs into the analysis. This technique was used to analyze the relationships (if any) of anomalous measurements of a target variable with observations from other sensors at the same site.
3. **Case III:** Sensor anomaly detection considering the readings of sensors located at nearby sites**.** This case was a more complex analysis where measurements from previous time steps and measurements from sensors located at the same site were considered, but at the same time readings from sensors at upstream sites were incorporated into the analysis.

The wide range of cases considered in this approach provided a clear idea of the most relevant cases for predicting measurements of environmental variables to identify possible measurement anomalies in support of real-time, automated QA/QC. Based on the results obtained from these three different approaches it was possible to quantify the probability that a measurement given by an environmental sensor is correct or not.

3.5 Case Study

In order to test the performance of the modeling cases identified above, it was necessary to select a case study area within which the modeling scenarios could be tested. This study used data collected using *in situ* sensors in the Little Bear River (LBR) of northern Utah, USA (Figure 3.1). The LBR watershed encompasses 182,000 acres and includes cropland, pasture, and rangeland. Land use is range/wildlife, irrigated land, dry cropland, and others. Land ownership is approximately 88% private, 10% national forest, and 2% state land. The National Forest and state lands are used primarily for grazing and forest areas (Utah Department of Environmental Quality, 1999).

The LBR database contains continuous water quality monitoring data from the LBR experimental watershed (Horsburgh et al., 2009). This data is managed and published using the Consortium of Universities for the Advancement of Hydrologic Science-Hydrologic Information System (CUAHSI-HIS) server with on-line updates. The data include 14 stations (seven water quality and stream flow measurement sites, four weather stations, one USGS gage, and two repeater sites). The favorable aspects of this database are:

Figure 3.1. Little Bear River Basin map, from Little Bear River WATERS Project (Horsburgh et al., 2009)

1. Good data records (time series): high frequency records of at least two years of measurements of interest are available.
2. Multiple sensors and stations: provide the possibility of evaluating cases contemplating other measurements from the same site and other measurements from nearby sites.
3. Calibration and field records: evaluation of the behavior of sensors at calibration time. This provides the possibility of relating suspicious readings in any particular situation found at the site during the calibration. Field observations can help identify real anomalies.
4. Easy access: can access the data through internet connection anytime.
5. On-line updates: new data is consistently being uploaded to the database.

3.5.1 Sites and variable selection

The sites and variables included in this analysis were strictly dependant on the three cases previously defined. First, the target measurement was defined. This was the variable selected to test our anomaly detection models. After visualizing (Figure 3.2) and analyzing the available data set, turbidity readings from the Paradise site were judged to be a good candidate for the target variable in our experimental procedures. The most relevant considerations for this selection are:

1. Numerous changes from the raw data to the QA/QC data were identified. This provides a data set where errors have previously been manually detected, providing possible targets when testing the models built for each case of the approach.

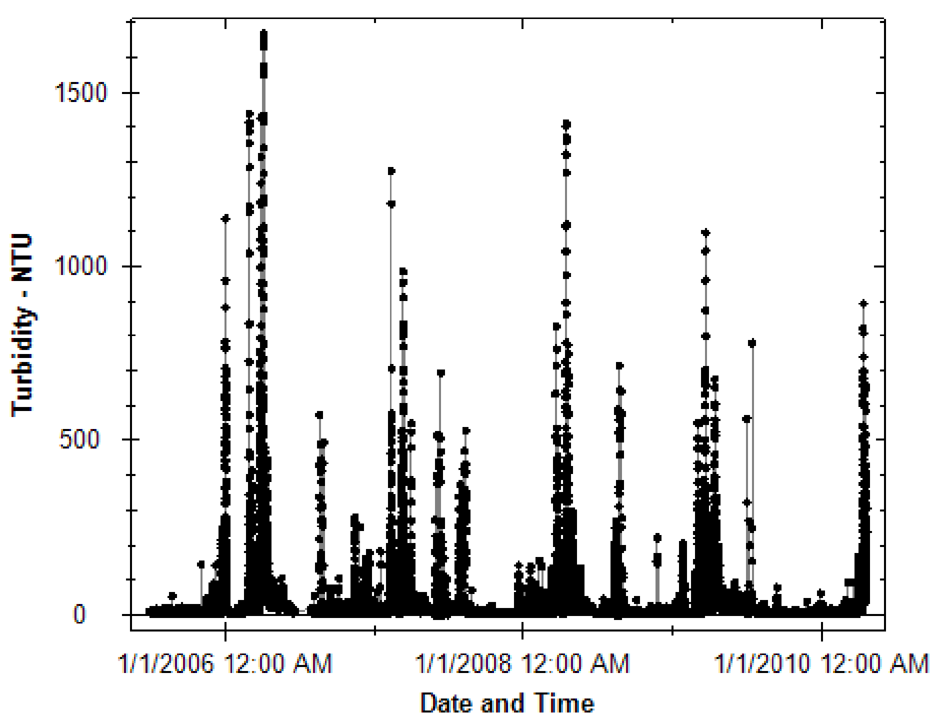

Figure 3.2. Turbidity readings (NTU) at McMurdy Hollow station near Paradise.

2. Unexpected changes in the turbidity readings are easy to see when visualizing the data using simple time series plots.
3. Seasonality changes are easy to visualize in the time series.
4. Turbidity sensors are set to be calibrated every two years. Also, cleaning and maintenance procedures are performed every other week for these sensors making measurements from this probe more reliable than others on the sites (Figure 3.3).
5. Since turbidity provides data on the optical clarity of the water due to suspended solids in it (measured in Nephelometric Turbidity Units, NTU), it is considered one of the principal physical characteristics of water (Panguluri et al., 2009).

For case I defined in the Approach Section, we only need to include turbidity measurements from previous time steps as inputs to the modeling. For case II, we need to include measurements from other sensors located at the same site. Table 3.1 shows other variables monitored and recorded with at least two years of to be included in the analysis. Since we want to evaluate possible relationships between turbidity measurements and other measurements on the site, all of them will be included in the analysis.

When selecting the data for case III, enough information had to be included to enable the evaluation of inter-site relationships. To do so, turbidity data from the Paradise site was set as the target measurement, which will enabled the inclusion of turbidity data from the upstream Confluence and South Fork sites. These sites can be identified in Figure 3.1 with the following numbers: water runs from the South Fork Site (1) to the Confluence site (4) and then goes through Paradise site (5).

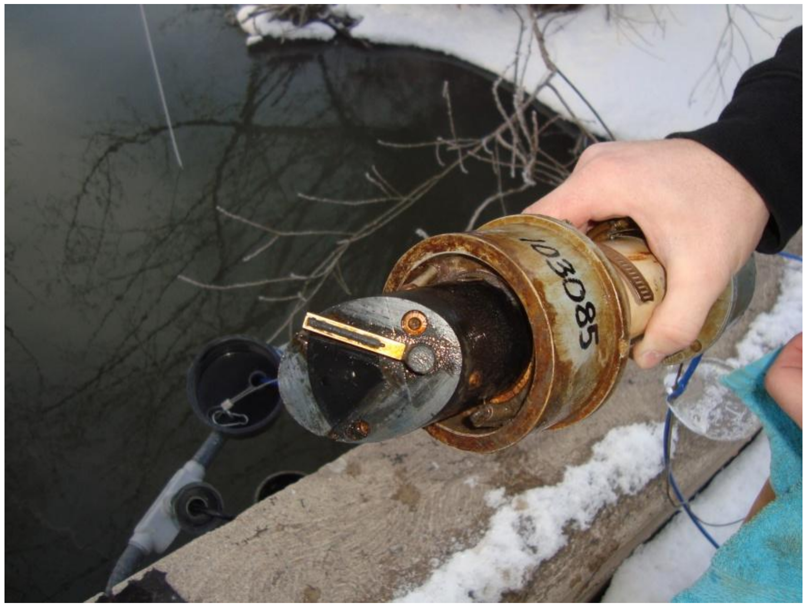

Figure 3.3. DTS-12 turbidity sensor during a calibration procedure.

3.5.2 Sensors description

To provide context to the measurements used in these analyses, a description of the sensors used from the Little Bear River Basin is presented on Table 3.2 (Horsburgh, 2008). One important characteristic to have in mind when analyzing the performance of the ANN and RVM models is the range of uncertainty in the measurements made by the sensors in each of the cases. This, together with the inclusion of several measurements from different sensors at a time into the models can affect their results and confidence

Table 3.1. Measurements included in the modeling for Case II Analysis

Measurement	Units
Turbidity	NTU
Dissolved Oxygen	(mg/L)
Specific Conductance	(µS/cm)
pH	0-14
Discharge	(ft^3/s)
Temperature	(°C)

Table 3.2. Description of the sensors used for variables measurement

Variables	**Type of Sensor/ Manufacturer**	**Accuracy Levels**
Turbidity	DTS-12/ Forest Technology Systems	From 0-499 NTU, ±2% From 500-1600 NTU, ±4%
Dissolved Oxygen Concentration	Hydrolab optical LDO, MiniSonde 5/ Hach Environmental, Inc.	If < 8 mg/L, ±0.1 mg/L If > 8 mg/L, ±0.2 mg/L
Specific Conductance	Hydrolab 4-electrode, MiniSonde 5 Hach Environmental, Inc.	±0.5%
Water Temperature	Hydrolab thermistor , MiniSonde 5/ Hach Environmental, Inc.	±0.1 °C
Stage	SPXD-600 pressure transducer/ KWK Technologies	±0.1 % Span
pH	Hydrolab reference electrode/ Hach Environmental, Inc.	Accuracy: ±0.2 pH units Resolution: 0.01 pH units

intervals within which anomalous points are detected.

The available data in the LBR data base includes raw measurements from the sensors and data from a manually performed QA/QC. This QA/QC analysis was visually performed by a human technician. The decision to categorize a point as an anomaly was based on time series plots of the originally measurements and previous experience in the field. This became relevant because cases detected as anomalies by the models were compared to the QA/QC detection (considered as correct) to assess model performance.

3.5.3 Data selection

Since the modeling approaches used in this study were data driven, it was necessary to have a significant and representative amount of data from the time series of the target measurement for use in the training and testing sets. As described above, a review of the available data was performed to identify the most relevant data for the analyses, and the time series of turbidity data at the Paradise site was selected. Analyzing the database and considering the three desired cases to be modeled, which implies remaining variables analyzed were also available, the data time frame selected for the model training was from 11/15/2007 0:00 to 11/14/2008 23:30, for a total of 17,568 cases (half-hour intervals). For the testing set, the time frame from 11/15/2008 0:00 to 12/14/2008 23:30 was selected, for a total of 1,440 cases (Figure 3.4). This test data set was selected because:

1. Various anomalies were detected.
2. No gaps were found in the time series from this period.
3. Data for other variables (same site and upstream sites) included in the analysis were also available.

3.5.4 Identification of anomalous data in the testing set

The objective of the regression models was the detection of anomalous data in the test dataset. After anomalous data values were detected, they were compared to the data values identified as anomalous in the manual QA/QC exercise. This allowed the identification of which cases were correctly identified as anomalous by the prediction models. This information was crucial for the evaluation of the performance of the models.

To have a clearer idea of the testing set, Table 3.3 shows the data points detected as anomalous by the manual QA/QC analysis, and their corrected values. This will allow visualization of the magnitude of the variation in the sensor measurements that are detected by our model as compared to those in the QA/QC analysis.

3.5.5 Autocorrelation analysis

A correlation analysis was performed to provide information about how many previous time steps must be included in the analysis when predicting a turbidity measurement for the next (future) time step. This analysis was done using MATLAB 2009 software through the use of the parcorr function, part of the GARTH Toolbox. The

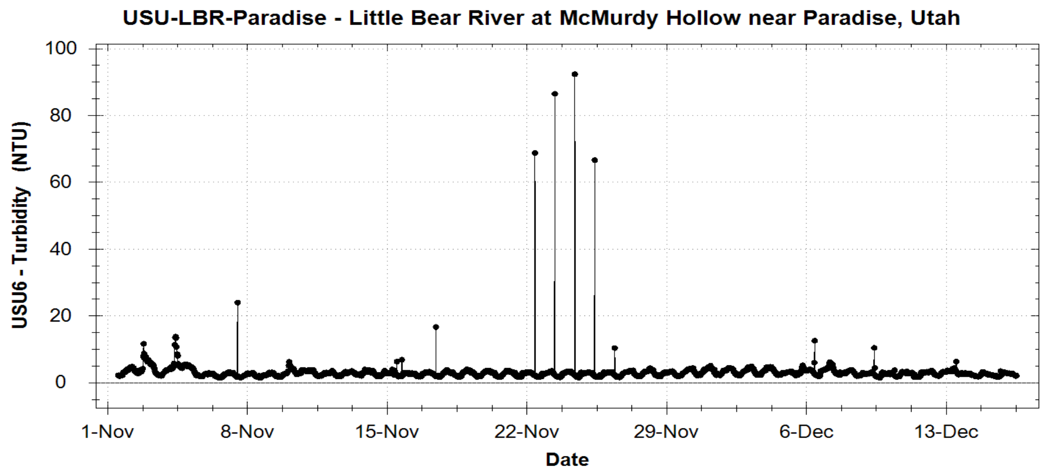

Figure 3.4. Plot of the test set selected for the analysis.

Table 3.3. Anomalous data values detected by the manual QA/QC analysis

Date	Measured Value (NTU)	Corrected value (NTU)
11/15/2008 11:30	6.34	2.28
11/15/2008 17:30	6.86	2.115
11/17/2008 10:30	16.7	2.03
11/20/2008 9:00	3.07	2.515
11/21/2008 9:30	3.07	2.315
11/21/2008 18:30	3.03	2.32
11/22/2008 9:30	68.75	2.195
11/22/2008 18:00	2.74	2.105
11/23/2008 9:30	86.51	2.245
11/24/2008 9:30	92.38	2.305
11/25/2008 9:30	66.66	2.01
11/26/2008 9:30	10.42	2.185
12/1/2008 4:30	5.16	4.17
12/1/2008 9:30	3.95	2.97
12/6/2008 9:30	6.03	2.72
12/6/2008 10:00	12.61	2.57
12/9/2008 9:30	10.44	2.4667
12/9/2008 10:00	4.48	2.3133
12/10/2008 10:00	3.79	2.085
12/13/2008 12:00	6.33	3.45

autocorrelation function is computed by fitting successive autoregressive models by ordinary least squares, retaining the last coefficient of each regression (Box, Jenkins, and Reinsel, 1994).

To obtain results, it was necessary to input the training set data vector for the target measurement (Paradise turbidity). This produced an autocorrelation plot allowing identification of possible statistical meaningful relationships between a current measurement and previous measurements for the same variable.

3.5.6 Travel time calculations

Water quality measurements from upstream sites were used as input for the case III modeling approach. In doing this we were able to analyze the relationships between a measurement upstream and its time-analogous downstream measurement.

Travel times were estimated by calculating the time it takes for a series of remarkable points from an upstream station to get to the Paradise station. Figure 3.5 shows turbidity plots from the Paradise and Confluence sites for similar time frames where similar behaviors for peak events can be observed.

3.6 Modeling

Several steps were followed when using machine learning approaches for sensor anomaly detection. The completion of these steps assured that the modeling process encompassed all required points to develop the most accurate models with the available tools and dataset.

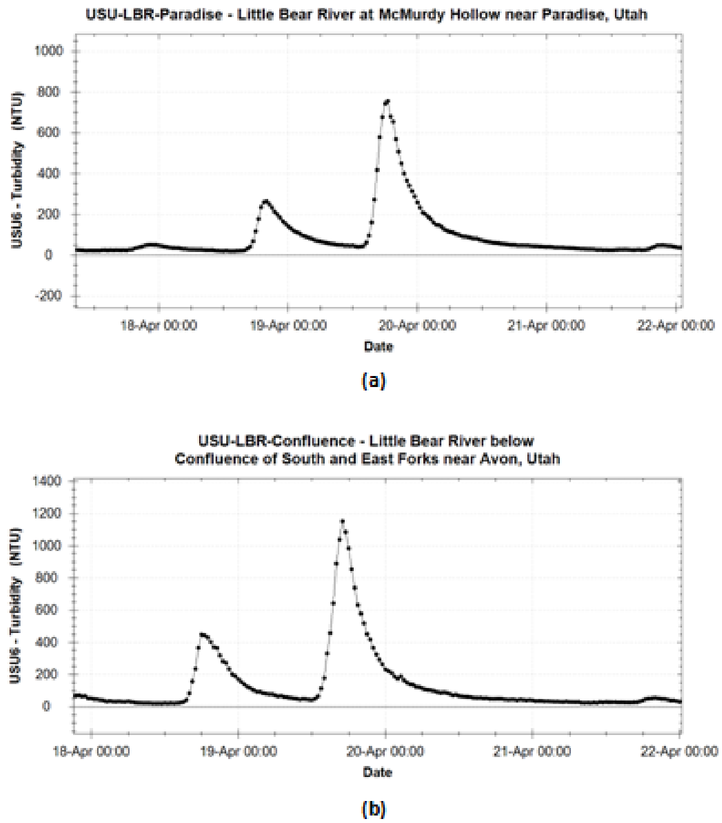

Figure 3.5. Examples of data points used for travel time calculations.

3.6.1 Regression models

Anomalies were identified by comparing actual readings to the values predicted by a regression model. A prediction confidence interval was estimated in order to have a pre-established range within which measurements were allowed to fall without being considered as anomalous. Confidence intervals of 30, 40, 50, 90, 95, and 99% were considered in the analysis. The 90, 95, and 99% intervals were selected because they have been used as reference points to display results on the literature (Hill, Minsker, and Amir, 2009). The 30, 40, and 50% intervals were included in the analysis to provide an extended idea of the capabilities of the models on lower confidence levels and to determine if useful information can be obtained from their application.

In order to determine how well the models were able to make predictions on the test data set, a first run was done predicting the entire test set. This was done for both the ANN and RVM models, taking into account the cases defined in the Approach section. This procedure provided important information in terms of the optimal parameter values needed by the models. It also provided important information in comparing the performance of the various models.

3.6.2 Artificial neural network modeling

This research modeling used a multi layer perceptron network (in this case a 2-layer feed-forward). The modeling was done using MATLAB software and the NETLAB toolbox (Nabney, 2002). By using this toolbox it was possible to create a model that worked in the form described in the approach section. The steps followed to conceive the ANN models were:

1. Determine the inputs and outputs and normalize the data.
2. Define initial parameters: 2-layer feed-forward network selected, linear activation function and conjugate gradient optimization algorithm were selected. The number of hidden units was determined by an iterative process for each of the cases evaluated.
3. Proceed with the prediction and creation of the confidence intervals based on the data and defined parameters.
4. Determine if the cases were anomalous or not based on the results obtained.

3.6.3 Relevance vector machine modeling

Development of the RVM models used a MATLAB sparse Bayesian algorithms implementation developed by Tipping in 2009. The following steps were used for the creation of these models:

1. Data input and normalization.
2. Definition of initial parameters: Gaussian kernel and likelihood selected, with the number of iterations to run set to 500 in this case. The width of the basis function was determined with an iterative process for each of the models.
3. Proceed with the prediction and creation of the confidence intervals based on the data and defined parameters.
4. Determine if the cases were anomalous or not based on the results obtained.

3.7 Measurement of Performance

In order to evaluate the performance of the ANN and RVM models it was necessary to assess them based on statistical characterization of their behavior. For the regression models, the following evaluation parameters were selected: Root Mean Square Error (RMSE), Nash-Sutcliffe Efficiency Coefficient (E), and Noise levels. False positive and false negative rates were used to describe the models accuracy for detecting anomalous values.

The RMSE is a measure of the difference between values predicted by the model or estimator and the values actually observed as given by Equation 3.1:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(x_{1,i} - x_{2,i}\right)^2}{n}} \tag{3.1}$$

where the $x_{1,i}$ are the predicted values from the model, the $x_{2,i}$ are the actual measurements for the time steps, and n is the number of cases in the testing set.

The Nash-Sutcliffe Efficiency Coefficient is used to assess the predictive power of hydrological models with values ranging from $-\infty$ to 1. An efficiency of 1 corresponds to a perfect match of the model to the observed data. Efficiency equal to 0 indicates that the model predictions are no better than the mean of the observed data, and efficiency less than 0 occurs when the observed mean is a better predictor than the model. The Nash-Sutcliffe Efficiency Coefficient is given by Equation 3.2:

$$E = 1 - \frac{\sum_{t=1}^{T} (x_1^t - x_2^t)^2}{\sum_{t=1}^{T} (x_1^t - \overline{x_1})^2} \tag{3.2}$$

The noise levels refer to signals that may be detected by a measurement which are not due to the actual phenomenon being measured, and tend to make the measurement uncertain to a greater or lesser degree. For the ANN case, noise will be calculated by Equation 3.3:

$$\sigma^2 = \frac{1}{\beta} + g^T A^{-1} g \tag{3.3}$$

where σ is the error bar on the prediction, β^{-1} is the noise in the target data, g is the gradient of the prediction matrix and A is the Hessian matrix of the error function (Anthony, 2001). For the case of the RVM models, noise levels were calculated by Equation 3.4:

$$\sigma^2 = \sigma_{MP}^2 + \phi(x)^T \sum \phi(x) \tag{3.4}$$

where σ is the error bar on the prediction, σ^2_{MP} is the noise in the target data, and $\phi(x)$ the basis function.

Since anomalous values were identified within the testing dataset by a manual QA/QC process, a comparison could be made of how many cases were correctly identified as anomalous by each of the models. Two metrics were calculated for this comparison. The false positive rate accounts for the number of cases that the model detects as anomalous but for which the manual QA/QC analysis indicated otherwise. The false negative rate refers to the cases wherein the model predicts that no anomaly has occurred, but for which the manual QA/QC analysis indicated otherwise. The false positive and false negative rates were an important part of the research results because these statistics have been used in previous research described by the literature review as the measure of the performance of models in detecting anomalous sensor readings. They provided an indication of the quality of the behavior of the model when compared to those developed by previous research in this area (e.g., Hill, Minsker, and Amir, 2009).

The last method used to evaluate model performance was a bootstrap analysis. This is a method for assigning measures of accuracy to sample estimates. A bootstrap analysis made it possible to assess the robustness of the ANN and RVM models. In this case, the bootstrap analysis was performed by randomly re-sampling with replacement 100 samples from the training data and calculating the performance parameters previously described (RMSE, NASH Coefficient, and Noise) for each of the random samples. From these results, we extracted an empirical bootstrap distribution and used it to characterize model robustness. This analysis was conducted for both the ANN and RVM models.

CHAPTER 4

RESULTS AND DISCUSSION

Before presenting the individual time step anomaly predictions, the results for data analysis procedures that allowed the use of this data set as useful inputs for the modeling cases are shown.

4.1 Autocorrelation Analysis Results

Results from the autocorrelation analysis (Figure 4.1) show that up to six previous time steps have relevant relationship to a given measurement. With this in mind, the inclusion of the measurements of these previous time steps for the regression analysis was evaluated.

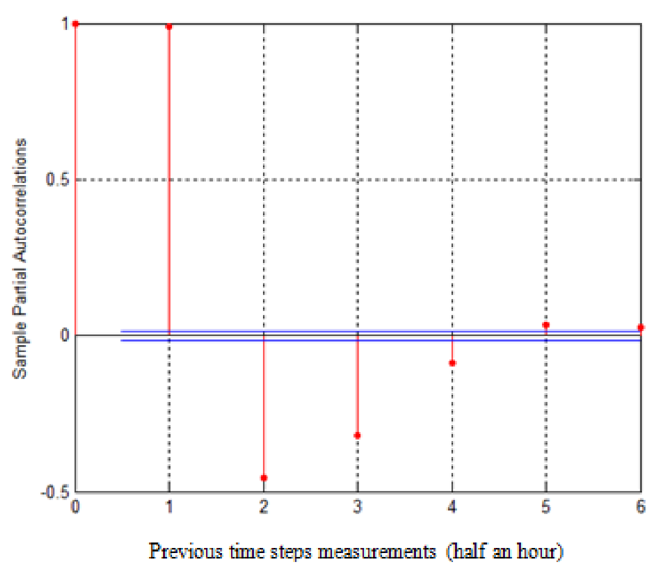

Figure 4.1. Autocorrelation plot for turbidity data from Paradise site (2007-2008). The red lines display the strength of the relationship observed with the current measurement. The blue lines represent the 95% confidence bounds for lags where autocorrelation was observed.

Ultimately, after constructing the ANN and RVM regression models, we found that adding the fifth and sixth previous time step measurements as inputs did not present improvements in any of cases examined (Table 4.1). With this in mind, together with the small autocorrelation obtained (Figure 4.2), the cases analyzed only included up to four previous time steps as inputs.

4.2 Travel Time Calculations Results

As travel time is dependent upon the discharge rate, the travel time analysis was completed for a number of different discharges so that a relationship between discharge and travel time could be extracted. The resulting travel times are shown (Figures 4.2 and 4.3) as a function of the discharge rate. These results allowed the use of appropriately lagged data for the case III analyses.

4.3 Scenarios Evaluated

Three modeling approaches were identified in the previous chapter to evaluate the ability of the various modeling approaches to identify sensor anomalies. Based on the three modeling cases to evaluate and the contents of the available database, a total of 36 scenarios were evaluated (Table 4.2). Each case had two modalities: in one, the data detected as anomalous was corrected before continuing the time-series forecasts; in the other, no corrections were made. In the case where a measurement was determined by the model to be anomalous and a correction was made, raw data values were replaced by the manual QA/QC procedure results. This design enabled the determination of how many upcoming readings were affected by having an anomalous reading as an input to the model for each of the three modeling cases.

Table 4.1. ANN prediction using inputs from previous time steps

Previous Time Steps (half hour)	RMSE	Nash Coefficient	Hidden units	Noise
1	5.45	0.985	6	8.32
2	6.14	0.981	3	6.67
3	6.06	0.981	6	6.44
4	5.98	0.982	20	6.35
5	6.04	0.982	13	6.29
6	6.01	0.982	3	6.32

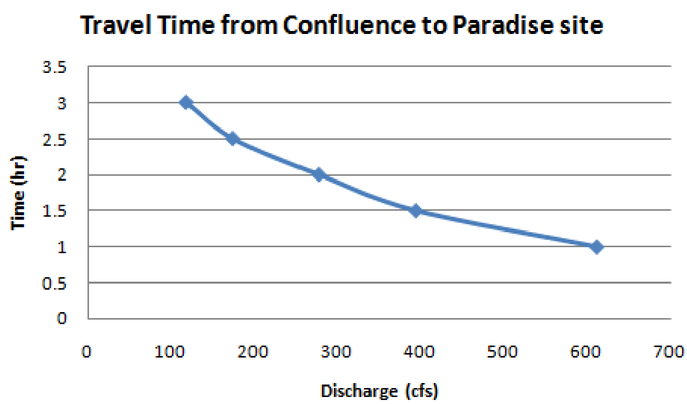

Figure 4.2. Travel time calculation results from Confluence to Paradise site.

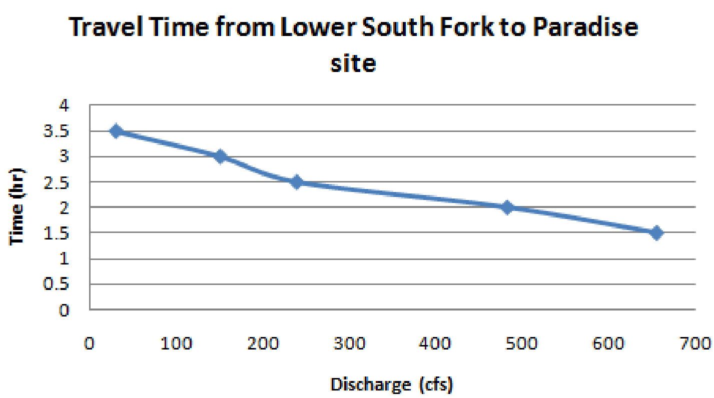

Figure 4.3 Travel time calculation results from South Fork to Paradise site.

Table 4.2. Cases evaluated for the individual time step anomaly detection

No.	Description
1	One previous time step without correcting anomalous data
2	One previous time step correcting anomalous data
3	Four previous time steps without correcting anomalous data
4	Four previous time steps correcting anomalous data
5	Paradise site Dissolved Oxygen data
6	Dissolved oxygen + one previous time step without correcting anomalous data
7	Dissolved oxygen + one previous time step correcting anomalous data
8	Dissolved oxygen + four previous time steps without correcting anomalous data
9	Dissolved oxygen + four previous time steps correcting anomalous data
10	Paradise site Specific Conductance
11	Specific Conductance + one previous time step without correcting anomalous data
12	Specific Conductance + one previous time step correcting anomalous data
13	Specific Conductance + four previous time steps without correcting anomalous data
14	Specific Conductance + four previous time steps correcting anomalous data
15	Paradise site pH
16	pH + one previous time step without correcting anomalous data
17	pH + one previous time step correcting anomalous data
18	pH + four previous time steps without correcting anomalous data
19	pH + four previous time steps correcting anomalous data
20	Paradise site Temperature
21	Temperature+ one previous time step without correcting anomalous data
22	Temperature + one previous time step correcting anomalous data
23	Temperature + four previous time steps without correcting anomalous data
24	Temperature + four previous time steps correcting anomalous data
25	Paradise site Discharge
26	Discharge+ one previous time step without correcting anomalous data
27	Discharge + one previous time step correcting anomalous data
28	Discharge + four previous time steps without correcting anomalous data
29	Discharge + four previous time steps correcting anomalous data
30	Confluence site turbidity
31	Confluence turbidity + 1 previous time step correcting anomalous data
32	Confluence turbidity + 4 previous time steps correcting anomalous data
33	Lower South Fork site turbidity
34	Lower South Fork site turbidity + 1 previous time step correcting anomalous data
35	Lower South Fork site turbidity + 4 previous time step correcting anomalous data
36	Paradise site discharge + 4 previous time steps + Confluence site turbidity

4.4 Artificial Neural Network Modeling Results

Table 4.3 shows a summary of the best results obtained when making the prediction of the test data set using the ANN model. In this analysis it can be observed that using one previous time step displayed one of the best prediction results, but also presented the highest noise level of these cases. This indicates that the model has limited capacity to changes for upcoming measurements when training is based on use of only

Table 4.3. ANN regression results

Case Evaluated	RMSE	Nash Coefficient	Hidden units	Noise
Turbidity one previous time step	5.45	0.985	6	8.32
Turbidity four previous time steps	5.98	0.982	20	6.35
Turbidity four previous time steps + Specific Conductance	6.06	0.981	20	6.02
Turbidity one previous time step + pH	5.65	0.984	8	8.26
Turbidity four previous time steps + Discharge	5.77	0.982	9	6.35
Turbidity four previous time steps + Confluence turbidity	5.64	0.975	8	6.55

one previous time step as input. On the other hand, cases including four previous time steps tended to achieve good prediction with lower levels of noise. Over all, inclusion of four previous time steps and discharge measurements, and four previous time steps and Confluence site turbidity measurements to the analysis displayed the best results. The rest of the results for the cases considered for this prediction are shown in Appendix A.

4.4.1 Artificial neural networks case I results

Case I results refer to the models that only used as inputs turbidity data from the measurements made in previous time steps recorded at the same site as our target site. For this particular case the model was set up to include as inputs the measurements obtained from up to four immediately previous time steps. The model was run varying the number of previous time steps included, from one to four.

The best result for this case was obtained by using substituting the corrected versions of the anomalous data. In this case the most successful input combination included all measurements from the four previous time steps (Table 4.4). The results from other input combinations are shown in the form of false positive/negative rates in

Table 4.4. ANN prediction of Paradise turbidity using as input four previous time step measurements

Cases detected as anomalous:

	RN	Date	4PTS	3PTS	2PTS	1PTS	CTS	PI Detected					
1	24	11/15/2008 11:30	3.01	2.58	2.35	2.31	6.34	30.00%	40.00%	50.00%			
2	36	11/15/2008 17:30	2.04	1.99	1.97	1.93	6.86	30.00%	40.00%	50.00%			
3	118	11/17/2008 10:30	2.37	2.35	2.24	1.99	16.7	30.00%	40.00%	50.00%	90.00%	95.00%	
4	356	11/22/2008 9:30	2.54	2.49	2.27	2.36	68.75	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
5	404	11/23/2008 9:30	2.99	2.81	2.56	2.32	86.51	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
6	452	11/24/2008 9:30	2.68	2.49	2.54	2.39	92.38	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
7	500	11/25/2008 9:30	2.63	2.40	2.11	2.2	66.66	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
8	548	11/26/2008 9:30	2.97	2.58	2.54	2.33	10.42	30.00%	40.00%	50.00%			
9	1028	12/6/2008 9:30	3.53	3.34	3.13	2.87	6.03	30.00%	40.00%				
10	1029	12/6/2008 10:00	3.34	3.13	2.87	6.03	12.61	30.00%	40.00%	50.00%			
11	1172	12/9/2008 9:30	2.89	2.68	2.61	2.62	10.44	30.00%	40.00%	50.00%			
12	1369	12/13/2008 12:00	4.43	4.01	3.79	3.58	6.33	30.00%					

	False Positive Cases	False Negative Cases
30%PI	0	8
40%PI	0	9
50%PI	0	11
90% PI	0	14
95% PI	0	15
99% PI	0	16

PTS	Previous Time Step
CTS	Current Time Step
RN	Reference Number of case
(shaded)	Cases detected as anomalous by manual QA/QC

Appendix B. When including the recorded values from four previous time steps as input, anomalous measurements were detected with a wider prediction interval than when using a single previous time step. Small improvements are shown in the number of data points detected as anomalous for this case.

4.4.2 Artificial neural networks case II results

Case II included as input to the time series model the data measured from previous time steps from both the turbidity probe (which is the measured value in question) and from other measurements from other probes located at the same site (in this case, the Paradise site). The following measurements were examined for this case: dissolved oxygen, pH, specific conductance, temperature, and discharge. In all these cases, we found that the one with the best results was the combination of measured turbidity data from four previous time steps and discharge measurements (Table 4.5). When using this combination, improvements were seen in both the number of measurements detected as anomalous and also in the confidence levels at which anomalous cases were detected. This produced the best combination of all the models evaluated for case II using neural networks as the modeling approach.

4.4.3 Artificial neural networks case III results

When evaluating the inclusion of the measurements made at the Confluence and South Fork sites as inputs to the model, the inclusion of turbidity data from four previous time steps with the turbidity measurements from the Confluence site (lagged for travel time) demonstrates some improvement from ANN case I results, but results from the ANN case II modeling were still better (Table 4.6).

Table 4.5. ANN prediction of Paradise turbidity using as input discharge and four previous time step measurements correcting anomalous data

Cases detected as anomalous:

	RN	Date	Discharge	4pts	3PTS	2PTS	1PTS	CTS	PI Detected					
1	24	11/15/2008 11:30	39.00	3.01	2.58	2.35	2.31	6.34	30.00%	40.00%				
2	36	11/15/2008 17:30	39.00	2.04	1.99	1.97	1.93	6.86	30.00%	40.00%	50.00%			
3	118	11/17/2008 10:30	39.00	2.37	2.35	2.24	1.99	16.7	30.00%	40.00%	50.00%	90.00%	95.00%	
4	356	11/22/2008 9:30	37.00	2.54	2.49	2.27	2.36	68.75	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
5	404	11/23/2008 9:30	38.00	2.99	2.81	2.56	2.32	86.51	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
6	452	11/24/2008 9:30	35.00	2.68	2.49	2.54	2.39	92.38	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
7	500	11/25/2008 9:30	36.00	2.63	2.40	2.11	2.2	66.66	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
8	548	11/26/2008 9:30	37.00	2.97	2.58	2.54	2.33	10.42	30.00%	40.00%	50.00%			
9	1028	12/6/2008 9:30	42.00	3.53	3.34	3.13	2.87	6.03	30.00%	40.00%				
10	1029	12/6/2008 10:00	36.00	3.34	3.13	2.87	6.03	12.61	30.00%	40.00%	50.00%	90.00%		
11	1172	12/9/2008 9:30	33.00	2.89	2.68	2.61	2.62		30.00%	40.00%	50.00%			
12	1369	12/13/2008 12:00	44.00	4.43	4.01	3.79	3.58	6.33	30.00%					

	False Positive Cases	False Negative Cases
30%PI	0	8
40%PI	0	9
50%PI	0	11
90% PI	0	14
95% PI	0	15
99% PI	0	16

PTS	Previous Time Step
CTS	Current Time Step
RN	Reference Number of case
	Cases detected as anomalous by manual QA/QC

4.4.4 ANN false positive and false negative rates

False positive and false negative rates were identified as important results to be obtained by this research. Table 4.7 and 4.8 present a summary of the rates obtained for the most relevant cases evaluated with the ANN models. The small values of the rates presented on Table 4.7 reflect how well the model correctly detects anomalous points within the data set. This table also shows how false positive rates can be reduced to 0% by using the corrected versions of the anomalous data as identified in the manual QA/QC process.

Similarly, Table 4.8 shows results for false negative rates obtained when using ANN models. Further detail on the type of cases correctly/not correctly detected as anomalous by the ANN models is displayed on Section 4.6. The rest of the false positive/negative results for the cases evaluated with the ANN models are presented on Appendix B.

4.4.5 ANN bootstrap analysis results

As previously explained, a bootstrap analysis was performed to assess the model robustness when presented with previously unseen inputs. Results for the ANN bootstrap analyses are shown in Figure 4.4. In the case of the ANN results shown in the figure, low variation can be observed within the statistical parameters analyzed. This implies that the ANN models would maintain their good performance when confronted with unseen data. This is an important part of the project because it will allow analyzing different data cases based on the approach defined in this research and comparison of the robustness of different models. Appendix D shows the rest of the bootstrap results for the ANN models.

Table 4.6. ANN Prediction of Paradise Turbidity using as input four previous time step measurements and Confluence site turbidity measurements

Cases detected as anomalous:

	RN	Date	Discharge	4pts	3PTS	2PTS	1PTS	CTS	PI Detected					
1	24	11/15/2008 11:30	39.00	3.01	2.58	2.35	2.31	6.34	30.00%	40.00%				
2	36	11/15/2008 17:30	39.00	2.04	1.99	1.97	1.93	6.86	30.00%	40.00%	50.00%			
3	118	11/17/2008 10:30	39.00	2.37	2.35	2.24	1.99	16.7	30.00%	40.00%	50.00%	90.00%	95.00%	
4	356	11/22/2008 9:30	37.00	2.54	2.49	2.27	2.36	68.75	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
5	404	11/23/2008 9:30	38.00	2.99	2.81	2.56	2.32	86.51	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
6	452	11/24/2008 9:30	35.00	2.68	2.49	2.54	2.39	92.38	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
7	500	11/25/2008 9:30	36.00	2.63	2.40	2.11	2.2	66.66	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
8	548	11/26/2008 9:30	37.00	2.97	2.58	2.54	2.33	10.42	30.00%	40.00%	50.00%			
9	1028	12/6/2008 9:30	42.00	3.53	3.34	3.13	2.87	6.03	30.00%	40.00%				
10	1029	12/6/2008 10:00	36.00	3.34	3.13	2.87	6.03	12.61	30.00%	40.00%	50.00%	90.00%		
11	1172	12/9/2008 9:30	33.00	2.89	2.68	2.61	2.62		30.00%	40.00%	50.00%			
12	1369	12/13/2008 12:00	44.00	4.43	4.01	3.79	3.58	6.33	30.00%					

	False Positive Cases	False Negative Cases
30%PI	0	8
40%PI	0	9
50%PI	0	10
90% PI	0	15
95% PI	0	15
99% PI	0	16

PTS	Previous Time Step
CTS	Current Time Step
RN	Reference Number of case
	Cases detected as anomalous by manual QA/QC

Table 4.7. False positive rates obtained when using ANN models for the anomaly detection prediction

No.	Description	Confidence Interval					
		30%	40%	50%	90%	95%	99%
1	One previous time step without	0.70%%	0.63%	0.49%%	0.35%%	0.28%%	0.28%%
2	One previous time step correcting	0%	0%	0%	0%	0%	0%
3	Four previous time steps without	1.34%	1.13%	1.06%	0.49%	0.49%	0.35%
4	Four previous time steps	0%	0%	0%	0%	0%	0%
5	Discharge + four previous time	1.55%	1.27%	0.99%	0.49%	0.35%	0.28%
6	Discharge + four previous time	0%	0%	0%	0%	0%	0%
7	Confluence turbidity + 4 previous	0%	0%	0%	0%	0%	0%

Table 4.8. False negative rates obtained when using ANN models for the anomaly detection prediction

No.	Description	Confidence Interval					
		30%	40%	50%	90%	95%	99%
1	One previous time step without correcting anomalous data	35%	50%	55%	80%	80%	80%
2	One previous time step correcting anomalous data	45%	55%	60%	75%	80%	80%
3	Four previous time steps without correcting anomalous data	35%	40%	45%	75%	75%	80%
4	Four previous time steps correcting anomalous data	40%	45%	50%	75%	75%	80%
5	Discharge + four previous time steps without correcting anomalous data	35%	40%	50%	75%	75%	80%
6	Discharge + four previous time steps correcting anomalous data	40%	45%	55%	70%	75%	80%
7	Confluence turbidity + 4 previous time steps correcting anomalous data	40%	55%	60%	75%	75%	80%

4.5 Relevance Vector Machine Results

As defined in the approach section, RVM modeling was also used to evaluate sensor anomaly detection schemes for the previously defined set of cases. As was done for the ANN modeling, a prediction run on the test set was performed using the relevance vector machine models (Table 4.9). This was done to establish the initial parameters for the cases to evaluate and at the same time provide an idea of the most promising cases for the individual time step predictions. This also provided a comparison starting point between the ANN and RVM models.

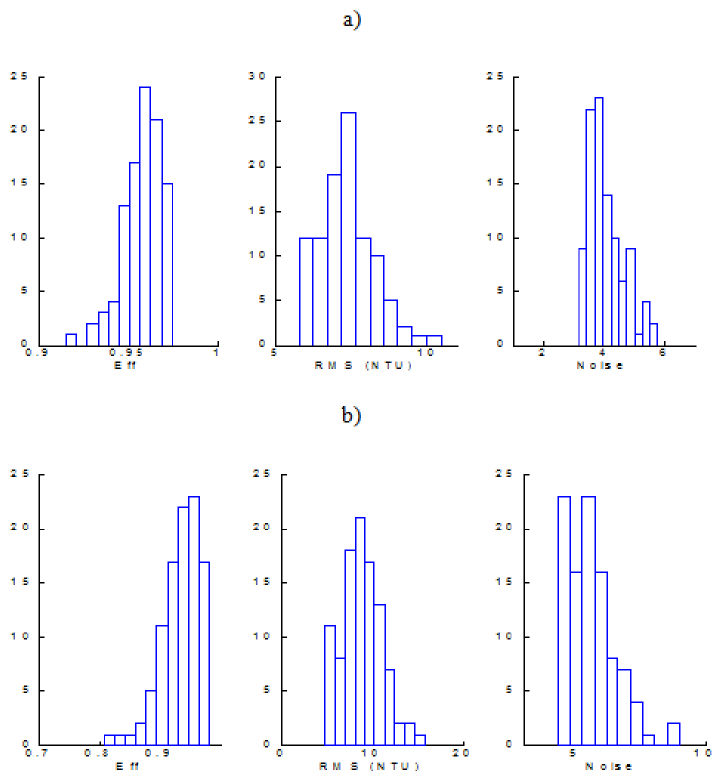

Figure 4.4. ANN bootstrapping analysis results for: a) using as input four previous time steps measurements for the prediction and b) using as input discharge and four previous time steps.

Table 4.9. RVM regression results

Case Evaluated	RMSE	Nash Coefficient	Basis Width	Noise
Turbidity one previous time step	4.03	0.986	0.06	9.48
Turbidity four previous time steps	4.97	0.978	0.08	6.72
Turbidity four previous time steps + Specific Conductance	5.10	0.982	0.02	11.03
Turbidity one previous time step + pH	4.57	0.985	0.05	8.56
Turbidity four previous time steps + Discharge	4.72	0.984	0.07	8.35
Turbidity four previous time steps + Confluence turbidity	4.45	0.983	0.66	8.44

4.5.1 Relevance vector machine case I results

When including data only from previous time steps for detection of individual anomalous data points, the best results (Table 4.10) were obtained when using measurements from four previous time steps as inputs. Using a single previous time step measurement as input for prediction of the sensor value produced good results as well, but when adding measurements from four previous time steps as inputs, confidence levels and the number of cases correctly detected as anomalies were improved.

4.5.2 Relevance vector machine case II results

The best results for the individual time step predictions when including as input other measurements from the same site as the target measurement was obtained when combining measurements from four previous time steps and discharge measurements as inputs (Table 4.11). These results were the best results obtained for the anomaly detection problem addressed in this research.

4.5.3 Relevance vector machine case III results

When including measurements from upstream sites as inputs to the RVM modeling the best results were found when using four previous time steps together with the Confluence site turbidity measurements, again, lagged for travel time (Table 4.12). These results were better than those from the RVM case I but did outperform the best RVM case II model.

Table 4.10. RVM prediction using as inputs data from four previous time steps of Paradise turbidity measurements

Cases detected as anomalous:

	RN	Date	4PTS	3PTS	2PTS	1PTS	CTS	PI Detected					
1	24	11/15/2008 11:30	3.01	2.58	2.35	2.31	6.34	30.00%	40.00%	50.00%			
2	36	11/15/2008 17:30	2.04	1.99	1.97	1.93	6.86	30.00%	40.00%	50.00%			
3	118	11/17/2008 10:30	2.37	2.35	2.24	1.99	16.7	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
4	356	11/22/2008 9:30	2.54	2.49	2.27	2.36	68.75	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
5	404	11/23/2008 9:30	2.99	2.81	2.56	2.32	86.51	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
6	452	11/24/2008 9:30	2.68	2.49	2.54	2.39	92.38	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
7	500	11/25/2008 9:30	2.63	2.40	2.11	2.2	66.66	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
8	548	11/26/2008 9:30	2.97	2.58	2.54	2.33	10.42	30.00%	40.00%	50.00%	90.00%		
9	1028	12/6/2008 9:30	3.53	3.34	3.13	2.87	6.03	30.00%	40.00%	50.00%			
10	1029	12/6/2008 10:00	3.34	3.13	2.87	6.03	12.61	30.00%	40.00%	50.00%	90.00%	95.00%	
11	1172	12/9/2008 9:30	2.89	2.68	2.61	2.62	10.44	30.00%	40.00%	50.00%	90.00%		
12	1173	12/9/2008 10:00	2.68	2.61	2.62	2.47	4.48	30.00%					
13	1369	12/13/2008 12:00	4.43	4.01	3.79	3.58	6.33	30.00%	40.00%				

PTS Previous Time Step
CTS Current Time Step
RN Reference Number of case
Cases detected as anomalous by manual QA/QC

	False Positive Cases	False Negative Cases
30%PI	0	7
40%PI	0	8
50%PI	0	9
90% PI	0	12
95% PI	0	14
99% PI	0	15

Table 4.11. RVM results using as inputs data from four previous time steps of Paradise turbidity and Paradise discharge measurements

Cases detected as anomalous:

	RN	Date	Discharge	4pts	3PTS	2PTS	1PTS	CTS	PI Detected					
1	14	11/15/2008 6:30	39.00	2.99	2.97	3.01	3.98	2.87	30.00%					
2	24	11/15/2008 11:30	39.00	3.01	2.58	2.35	2.31	6.34	30.00%	40.00%	50.00%			
3	36	11/15/2008 17:30	39.00	2.04	1.99	1.97	1.93	6.86	30.00%	40.00%	50.00%			
4	118	11/17/2008 10:30	39.00	2.37	2.35	2.24	1.99	16.7	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
5	356	11/22/2008 9:30	37.00	2.54	2.49	2.27	2.36	68.75	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
6	404	11/23/2008 9:30	38.00	2.99	2.81	2.56	2.32	86.51	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
7	452	11/24/2008 9:30	35.00	2.68	2.49	2.54	2.39	92.38	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
8	500	11/25/2008 9:30	36.00	2.63	2.40	2.11	2.2	66.66	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
9	548	11/26/2008 9:30	37.00	2.97	2.58	2.54	2.33	10.42	30.00%	40.00%	50.00%	90.00%		
10	1028	12/6/2008 9:30	42.00	3.53	3.34	3.13	2.87	6.03	30.00%	40.00%	50.00%			
11	1029	12/6/2008 10:00	36.00	3.34	3.13	2.87	6.03	12.61	30.00%	40.00%	50.00%	90.00%	95.00%	
12	1172	12/9/2008 9:30	33.00	2.89	2.68	2.61	2.62	10.44	30.00%	40.00%	50.00%	90.00%		
13	1073	12/9/2008 10:00	39.00	5.57	5.55	5.58	4.79	4.6	30.00%					
14	1369	12/13/2008 12:00	44.00	4.43	4.01	3.79	3.58	6.33	30.00%	40.00%				

PTS	Previous Time Step
CTS	Current Time Step
RN	Reference Number of case
(shaded)	Cases detected as anomalous by manual QA/QC

	False Positive Cases	False Negative Cases
30%PI	0	7
40%PI	0	8
50%PI	0	9
90% PI	0	12
95% PI	0	14
99% PI	0	15

Table 4.12. RVM results using as inputs data from four previous time steps of Paradise turbidity and Confluence turbidity measurements

Cases detected as anomalous:

	RN	Date	Conf Turb	4PTS	3PTS	2PTS	1PTS	CTS	PI Detected					
1	24	11/15/2008 11:30	2.97	3.01	2.58	2.35	2.31	6.34	30.00%	40.00%	50.00%			
2	36	11/15/2008 17:30	1.99	2.04	1.99	1.97	1.93	6.86	30.00%	40.00%	50.00%			
3	118	11/17/2008 10:30	4.05	2.37	2.35	2.24	1.99	16.7	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
4	356	11/22/2008 9:30	2.36	2.54	2.49	2.27	2.36	68.75	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
5	404	11/23/2008 9:30	2.32	2.99	2.81	2.56	2.32	86.51	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
6	452	11/24/2008 9:30	2.39	2.68	2.49	2.54	2.39	92.38	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
7	500	11/25/2008 9:30	2.2	2.63	2.40	2.11	2.2	66.66	30.00%	40.00%	50.00%	90.00%	95.00%	99.00%
8	548	11/26/2008 9:30	3.03	2.97	2.58	2.54	2.33	10.42	30.00%	40.00%	50.00%	90.00%		
9	1028	12/6/2008 9:30	3.51	3.53	3.34	3.13	2.87	6.03	30.00%	40.00%	50.00%			
10	1029	12/6/2008 10:00	3.44	3.34	3.13	2.87	6.03	12.61	30.00%	40.00%	50.00%	90.00%	95.00%	
11	1060	12/7/2008 1:30	4.28	4.12	4.22	4.24	4.17	4.28	30.00%	40.00%				
12	1172	12/9/2008 9:30	2.67	2.89	2.68	2.61	2.62	10.44	30.00%	40.00%	50.00%	90.00%		
13	1369	12/13/2008 12:00	8.3	4.43	4.01	3.79	3.58	6.33	30.00%	40.00%	50.00%			

	False Positive Cases	False Negative Cases
30%PI	1	8
40%PI	1	8
50%PI	0	8
90% PI	0	12
95% PI	0	14
99% PI	0	15

PTS	Previous Time Step
CTS	Current Time Step
RN	Reference Number of case
	Cases detected as anomalous by manual QA/QC

4.5.4 RVM false positive and false negative rates

The small values of the false positive rates presented in Table 4.13 (similar to the ANN results) reflect how well the model was capable of correctly detecting anomalous points within the data set. It also shows how false positive rates can be reduced to 0% by using the QA/QC corrected versions of the data to predict the sensor value at the next time step.

Table 4.14 shows results for false negative rates obtained for the RVM modeling. The high false negative rates encountered were caused by the elevated number of cases detected as anomalies that showed only small variation between the sensor readings and the adjusted value from the manual QA/QC procedure (Section 4-6). The rest of the results for false positive/negative cases for the RVM models are presented on Appendix C.

Table 4.13. False positive rates obtained using RVM models

No.	Description	Confidence Interval					
		30%	40%	50%	90%	95%	99%
1	One previous time step without correcting anomalous data	0.70%	0.63%	0.49%	0.35%	0.28%	0.28%
2	One previous time step correcting anomalous data	0%	0%	0%	0%	0%	0%
3	Four previous time steps without correcting anomalous data	1.34%	1.13%	1.06%	0.49%	0.49%	0.35%
4	Four previous time steps correcting anomalous data	0%	0%	0%	0%	0%	0%
5	Discharge + four previous time steps without correcting anomalous data	1.55%	1.27%	0.99%	0.49%	0.35%	0.28%
6	Discharge + four previous time steps correcting anomalous data	0%	0%	0%	0%	0%	0%
7	Confluence turbidity + 4 previous time steps correcting anomalous data	0%	0%	0%	0%	0%	0%

Table 4.14. False negative rates obtained using RVM models

No.	Description	Confidence Interval					
		30%	**40%**	**50%**	**90%**	**95%**	**99%**
1	One previous time step without correcting anomalous data	35%	50%	55%	80%	80%	80%
2	One previous time step correcting anomalous data	45%	55%	60%	75%	80%	80%
3	Four previous time steps without correcting anomalous data	35%	40%	45%	75%	75%	80%
4	Four previous time steps correcting anomalous data	40%	45%	50%	75%	75%	80%
5	Discharge + four previous time steps without correcting anomalous data	35%	40%	50%	75%	75%	80%
6	Discharge + four previous time steps correcting anomalous data	40%	45%	55%	70%	75%	80%
7	Confluence turbidity + 4 previous time steps correcting anomalous data	40%	55%	60%	75%	75%	80%

4.5.5 RVM bootstrap analysis results

Figure 4.5 displays the best results obtained for the RVM bootstrap analysis. These results, which are similar to the ones obtained by the ANN models, displayed low variation of the statistical parameters analyzed. These results suggest a good ability of the RVM models to respond to previously unseen data. This is an important characteristic that will allow implementation of this type of models to the detection of anomalies in other types of measurements or problems with a similar focus as the one defined in this research. The rest of the bootstrap analyses evaluated for the RVM models can be found in Appendix E.

4.6 Points Correctly Detected as Anomalous

After observing the results obtained for both types of models, the data values within the test data set that were correctly detected as anomalies can be graphically visualized. This is another indication of the quality of the models. It also illustrates where it is more difficult for the models to detect anomalies. Figure 4.6 shows in a circle the points that were correctly detected as anomalous by the models, and it shows

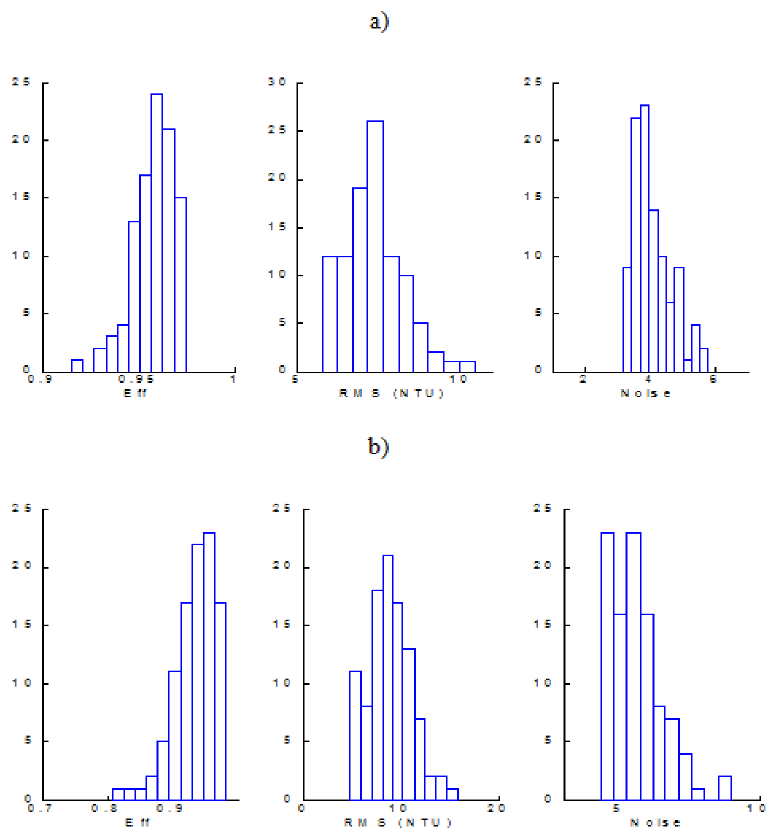

Figure 4.5. RVM bootstrapping analysis results for: a) using as input four previous time steps measurements for the prediction and b) using as input discharge and four previous time steps.

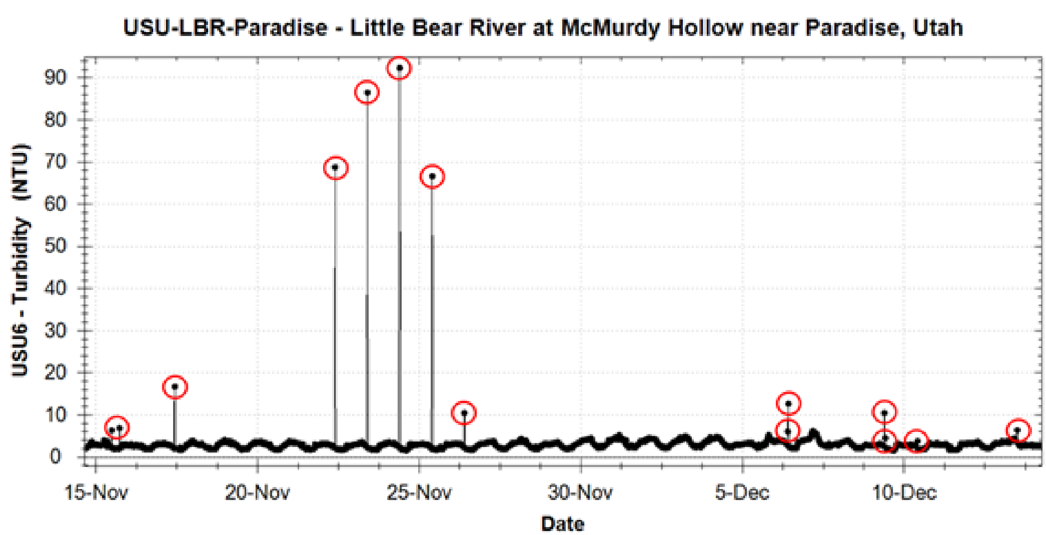

Figure 4.6. Cases correctly detected as anomalous by the ANN and RVM models on the test data set.

how both models are capable of detecting as anomalous cases that would be identified as suspect by a quick view of the test dataset plot.

Figure 4.7 shows some of the cases that were not detected as anomalous by the ANN and RVM models, but were identified by the manually performed QA/QC analysis. These cases showed small variation between the measured value and the value corrected by the manually performed QA/QC analysis. These cases occurred on the 20th and 21st of November 2008. The difference between measurements detected was: from 3.07 to 2.515 NTU, from 3.07 to 2.32 NTU, and from 3.03 to 2.32 NTU. Neither model was able to resolve such small variations between the measured and QA/QC-corrected values. Visually, these points might represent the possibility of anomalies for some experts but potentially not for others. In the end, this decision depends on the operator and whether some particular quantitative requirements are defined for the manual QA/QC procedure.

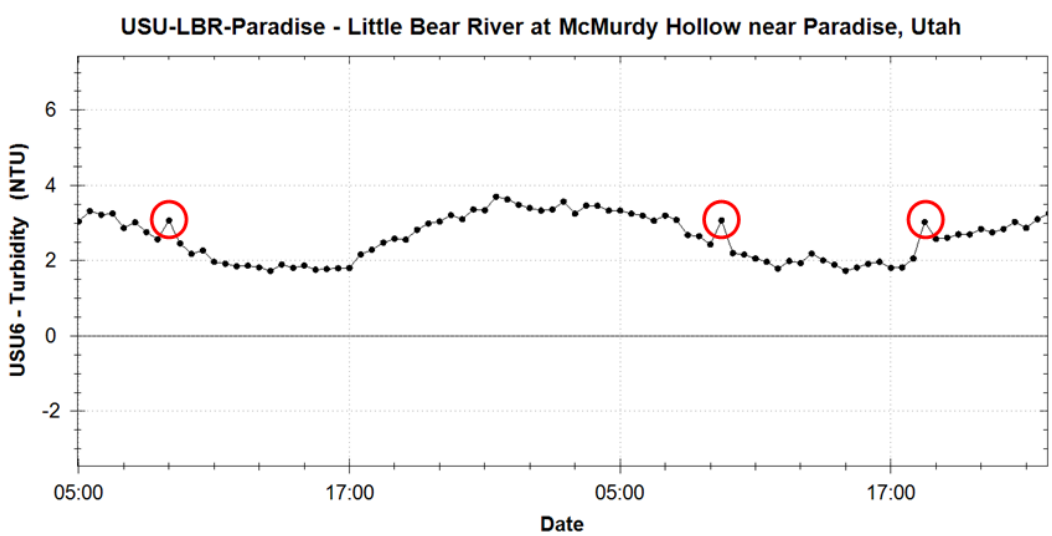

Figure 4.7. Cases not detected as anomalous by the ANN and RVM models, but were identified by the manually performed QA/QC analysis.

CHAPTER 5

CONCLUSION

The results obtained from the ANN and RVM models suggested the following conclusions regarding the detection of anomalous measurements:

1. It is possible to detect anomalous measurements by predicting the value from the sensor for upcoming time steps through the use of learning machine techniques.
2. Measurements from previous time steps proved to provide relevant information for the prediction of anomalous data points for both ANN and RVM models.
3. Information from other sensor measurements located at the same site as the target measurement is useful for the prediction when combined with data from measurements made at previous time steps.
4. Information from upstream stations improved the time series predictions when used together with data from measurements from previous time steps as input for both models.
5. The predictions performed on the test dataset for each model are a good indicator of the most promising cases for detection of individual anomalies.
6. Results obtained with the RVM models displayed some improvements in comparison with its analogous ANN models results. These improvements were observed in the number of correctly detected anomalies and also in the level of confidence within which anomalous measurements were detected.
7. By using the forecasted values for those measurements judged to be anomalous, the false negative rates for the models became zero. This was achieved by

eliminating the effect of anomalous points in the prediction of sensor measurements for upcoming time steps.

8. Both models successfully detected anomalies where there was a notable variation between the reading of the sensor and the corrected value estimated by the manual QA/QC procedure. In cases where only a small difference occurred, the data values were not detected as anomalies by the models.
9. Bootstrap analyses performed on both models displayed promising results when presented with previously unseen data.

CHAPTER 6

RECOMMENDATIONS

The following recommendations are offered:

1. Explore the use of relevance vector machine models such as those described in this research, looking forward to establish anomalous data detection on an online basis. This would allow early identification of suspicious measurements, providing room for improving management procedures with an earlier response to events in the watershed.
2. Conduct further investigations regarding the use of machine learning techniques for error detection in real-time data cases. As shown in this research, in matters of QA/QC data procedures and limitations, general assumptions, and models formulation, there exists room for exploring and better defining these steps for a more precise detection of anomalous points.
3. Evaluate sensor anomaly detection for larger data sets. This would allow more detailed exploration of the information available from other measurements in the data set, especially inter-site relationships between measurements obtained in a basin.
4. Expand on the use of a multi-variate model to predict several measurements at a time. This would facilitate the prediction of anomalous data points in several sensors at a time and might provide better identification of anomalous data points. This would require a more complex and detailed analysis based on the same data series than has been done for simply analyzing one sensor at a time.

5. Perform an analysis on measurements uncertainty and its relation to these types of prediction problem. This could provide a better understanding of the performance of the models when analyzing points in the raw data with small variations from the manual QA/QC analysis.
6. Evaluate the development of a classification model designed to identify the origin of sensor anomalies. Such a model could potentially be used to flag and automatically correct for some types of sensor anomalies.
7. Inquire into the use of maintenance procedures results and reports for determining the origin of anomalies. If different types of anomalies could be related to specific problems corrected when performing maintenance on the sensors, classes referring to sensor anomaly origin might be identifiable.

REFERENCES

Anthony, M. 2001. Discrete mathematics of neural networks: Selected topics. Siam, Philadelphia. 131 p.

Barron, J. W., A. I. Mounstapha, and R. R. Selmic. 2008. Real-time implementation of fault detection in wireless sensor networks using neural networks. Fifth International Conference in Information Technology: New Generations. p. 378-390.

Bishop, C. M., and M. E. Tipping. 2003. Bayesian regression and classification. Advances in learning theory: Methods, models and applications. IOS press. NATO Science Series III: computer and system sciences. 415 p.

Box, G. E., G. M. Jenkins, and G. C. Reinsel. 1994. Time series analysis: Forecasting and control. Third edition. Prentice Hall, Wiley, New Jersey. 592 p.

Christensen, V. G., A. C. Ziegler, P. P.Rasmussen, and X. Jian. 2003. Continous real-time water quality monitoring of kansas streams. AWRA 2003 Spring Specialty Conference.

Dayhoff, J. E., and J. M. DeLeo. 2001. Artificial neural networks, opening the black box. Cancer 91:1615-1635.

Guo, T.-H., and J. Nurre. 1991. Sensor failure detection and recovery by neural networks. Seattle International Joint Conference. Neural Networks 1:221-226.

Hill, D. J., and B. S. Minsker. 2006. Automated fault detection for in-situ environmental sensors. 7th International Conference on Hydroinformatics. Nice. p. 1-8.

Hill, D. J., B. S. Minsker, and E. Amir. 2009. Real-time Bayesian anomaly detection in streaming environmental data. Water Resources Research 45: W00D28.

Horsburgh, J. S. 2008. Hydrologic information systems: Advancing cyberinfrastructure for environmental observatories. Unpublished PhD dissertation. Utah State University, Logan, Utah. 57 p.

Horsburgh, J. S., D. K. Stevens, D. G. Tarboton, N. O. Mesner, A. Spackman Jones, and S. Guerrero. 2009. Monitoring data collected within the Little Bear River Experimental Watershed. Utah, USA. (Available at http://hiscentral.cuahsi.org/pub_network.aspx?n=52).

Izenman, A. J. 2008. Modern multivariate statistical techniques: Regression, classification and manifold learning. Springer, New York. 731 p.

Kang, J. M., S. Shekhar, M. Henjum, P. J. Novak, and W. A. Arnold. 2009. Discovering teleconnected flow anomalies: A relationship analysis of dynamic neighborhoods (RAD) approach. Lecture notes in computer science. p. 44-61.

Keith, L. H. 1983. Principles of environmental analysis. Analytical Chemistry. Transportation Research Board 896: 67-71.

Khalil, A., M. Mckee, M. Kemblowski, T. Asefa, and L. Bastidas. 2006. Multiobjective analysis of chaotic dynamic systems with sparse learning machines. Advances in Water Resources 29:72-88.

Kitaoku, K., T.Yoshida, and Y. Aoyama. 2004. Development of an online autoanalizer for organic acids influents of wastewater treatment plants and verification experiments of the prototype autoanalyzer. Yokogawa Tech Rep. Engl Ed. p. 13-17.

Kotamäki, N., S. Thessler, J. Koskiaho, A. O. Hannukkala, H. Huitu, T. Huttula, J. Havento, and M. Järvenpää. 2009. Wireless in-situ sensor network for agriculture and water monitoring on a river basin scale in Southern Finland: Evaluation from a data user's perspective. Sensors 9(4):2862-2883.

Moatar, F., F. Fessant, and A. Poirel. 1999. ph Modelling by neural networks. Application of Control and Validation data series in the Middle Loire river. Ecological Modeling 120:141-156.

Moatar, F., J. Miquel, and A. Poirel. 2001. A quality control method for physical and chemical monitoring data. Application to dissolved oxigen levels in the river Loire. Journal of Hydrology 252:25-36.

Nabney, I. T. 2002. Netlab, Algorithms for pattern recognition. Springer, Gateshead. 420 p.

Panguluri, S., G. Meiners, J. Hall, and J. G. Szabo. 2009. Distribution system water quality monitoring: Sensor technology evaluation methodology and results. U.S. Environmental Protection Agency, Washington, D. C.

Shirley, E. 1982. Quality control for environmental measurements. Transportation Research Board. p. 67-71.

Solomatine, D. P. 2002. Data-driven modeling: Paradigm, methods, experiences. Procedures of the 5th International Conference in Hydroinformatics, 1. Cardiff, UK. p. 757-763.

Tipping, M. E. 2001. Sparse bayesian learning and the relevance vector machine. Journal of Machine Learning Research 1:211-244.

Torres, A., R. Walker, and M. McKee. 2009. Bayesian analysis of sensor errors in an automated irrigation canal using multivariate relevance vector machines. 2009 Summer specialty conference adaptive management of water resources II. Utah: American Water Resources Association.

Utah Department of Environmental Quality. 1999. Little Bear River watershed total maximun daily loads.

Wei, L., Y. Yang, R. M. Nishikawa, M. N. Wernick, and A. Edwards. 2005. Relevance vector machine for automatic detection of clustered microcalcifications. IEEE Transactions on Medical Imaging 24:1278-1285.

Xu, X., J. Hines, and R. Uhrig. 1999. Sensor validation and fault detection using neural networks. Maintenance and Reliability Conference 2(58):1-9.

APPENDICES

Appendix A

ANN Regression Results

Appendix A display results for regression predictions on the test using ANN modeling. These include all the possible cases considered to have potential to provide important information for the detection of anomalous points and for the set up of the individual time step prediction models.

Table A.1. ANN prediction using specific time steps measurements as inputs

Previous Time Steps (half hour)	RMSE	Nash Coefficient	Hidden units	Noise
1	5.45	0.985	6	8.32
2	6.68	0.977	13	14.52
3	7.69	0.970	13	20.29
4	8.47	0.964	7	25.47
5	9.34	0.956	12	29.84
6	10.16	0.948	12	33.71

Table A.2. ANN prediction using previous year's measurements as input

Previous Years	RMSE	Nash Coefficient	Hidden units	Noise
1	44.33	0.005	8	54.00
2	45.83	0.063	8	52.06

Table A.3. ANN prediction using previous day's measurements as inputs

Previous Days	RMSE	Nash Coefficient	Hidden units	Noise
1	18.45	0.828	12	44.68
2	30.54	0.530	16	51.88

Table A.4. ANN prediction using other sensors on the same site measurements as inputs

Data	RMSE	Nash Coefficient	Hidden units	Noise
Turbidity one previous time step	5.45	0.985	6	8.32
Turbidity four previous time steps	5.98	0.982	20	6.35
Turbidity four previous time steps + Specific Conductance	6.06	0.981	20	6.02
Turbidity one previous time step + pH	5.65	0.984	8	8.26
Turbidity four previous time steps + pH	6.10	0.981	19	6.27
Turbidity four previous time steps + Temperature	6.72	0.977	7	6.93
Turbidity four previous time steps + Discharge	5.77	0.982	9	6.35

Table A.5. ANN prediction using upstream measurements as inputs

Data	RMSE	Nash Coefficient	Hidden units	Noise
Turbidity four previous time steps + Confluence turbidity	5.64	0.975	8	6.55
Turbidity four previous time steps + Confluence turbidity + Specific Conductance	6.14	0.982	18	6.49
Turbidity four previous time steps + Confluence turbidity + Dissolved Oxygen	6.93	0.977	18	8.39
Turbidity four previous time steps + Lower South Fork turbidity	7.85	0.954	17	10.26

Appendix B

ANN Specific Time Steps False Positive and False Negative Rates

Appendix B shows the false positive and false negative rates for all the cases evaluated for case I, case II and case III individual time step prediction using ANN models. Only the best results for each of the cases are displayed in the Results and Discussion section.

Table B.1. ANN false positive rates for specific time steps prediction using previous time steps as inputs

Input Data Description	Confidence Interval					
	30%	40%	50%	90%	95%	99%
Paradise turbidity one PTS without correcting anomalous points	0.70%	0.63%	0.49%	0.35%	0.28%	0.28%
Paradise turbidity one PTS correcting anomalous points	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Paradise turbidity four PTS without correcting anomalous points	1.34%	1.13%	1.06%	0.49%	0.49%	0.35%
Paradise turbidity four PTS correcting anomalous points	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Table B.2. False positive rates for specific time steps predictions using Paradise measurements as inputs

Input Data Description	Confidence Interval					
	30%	40%	50%	90%	95%	99%
Paradise site Dissolved Oxygen	28.45%	0.00%	0.00%	0.00%	0.00%	0.00%
Dissolved Oxygen + one PTS of Paradise turbidity without correcting anomalous points	0.85%	0.63%	0.56%	0.35%	0.35%	0.35%
Dissolved Oxygen + one PTS of Paradise turbidity correcting anomalous points	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Dissolved Oxygen + four PTS of Paradise turbidity without correcting anomalous points	1.48%	1.34%	1.06%	0.35%	0.35%	0.28%
Dissolved Oxygen + four PTS of Paradise turbidity correcting anomalous points	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Paradise site Specific Conductance	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Specific Conductance + one PTS of Paradise turbidity without correcting anomalous points	0.70%	0.63%	0.49%	0.35%	0.35%	0.28%
Specific Conductance + one PTS of Paradise turbidity correcting anomalous points	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Specific Conductance + four PTS of Paradise turbidity without correcting anomalous points	1.33%	1.20%	1.13%	0.56%	0.42%	0.28%
Specific Conductance + four PTS of Paradise turbidity correcting anomalous points	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Paradise site pH	0.35%	0.07%	0.00%	0.00%	0.00%	0.00%
pH + one PTS of Paradise turbidity without correcting anomalous points	0.70%	0.63%	0.49%	0.35%	0.35%	0.28%
pH + one PTS of Paradise turbidity correcting anomalous points	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
pH + four PTS of Paradise turbidity without correcting anomalous points	1.54%	1.40%	1.27%	0.56%	0.56%	0.28%
pH + four PTS of Paradise turbidity correcting anomalous points	21.27%	0.14%	0.07%	0.00%	0.00%	0.00%
Paradise site Temperature	22.81%	0.00%	0.00%	0.00%	0.00%	0.00%
Temperature+ one PTS of Paradise turbidity without correcting anomalous points	1.20%	0.99%	0.77%	0.42%	0.35%	0.28%
Temperature + one PTS of Paradise turbidity correcting anomalous points	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Temperature + four PTS of Paradise turbidity without correcting anomalous points	1.20%	0.92%	0.77%	0.42%	0.35%	0.28%
Temperature + four PTS of Paradise turbidity correcting anomalous points	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Paradise site Discharge	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Discharge + one PTS of Paradise turbidity without correcting anomalous points	0.70%	0.53%	0.42%	0.28%	0.21%	0.21%
Discharge + one PTS of Paradise turbidity correcting anomalous points	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Discharge + four PTS of Paradise turbidity without correcting anomalous points	1.55%	1.27%	0.99%	0.49%	0.35%	0.28%
Discharge + four PTS of Paradise turbidity correcting anomalous points	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Table B.3. ANN false positive rates for specific time steps prediction using upstream measurements as inputs

Input Data Description	Confidence Interval					
	30%	40%	50%	90%	95%	99%
Confluence site turbidity	0.14%	0.00%	0.00%	0.00%	0.00%	0.00%
Confluence turbidity + one PTS of Paradise turbidity correcting anomalous points	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Confluence turbidity + four PTS of Paradise turbidity correcting anomalous points	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Lower South Fork site turbidity	0.92%	0.35%	0.28%	0.14%	0.14%	0.00%
Lower South Fork site turbidity + one PTS of Paradise turbidity correcting anomalous points	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Lower South Fork site turbidity + four PTS of Paradise turbidity correcting anomalous points	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Paradise site Discharge + four PTS of Paradise turbidity + Confluence site turbidity	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Table B.4. ANN false negative rates for specific time steps prediction using previous time steps as inputs

Input Data Description	Confidence Interval					
	30%	40%	50%	90%	95%	99%
Paradise turbidity one PTS without correcting anomalous points	40.00%	50.00%	55.00%	80.00%	80.00%	80.00%
Paradise turbidity one PTS correcting anomalous points	45.00%	55.00%	60.00%	75.00%	80.00%	80.00%
Paradise turbidity four PTS without correcting anomalous points	35.00%	40.00%	45.00%	75.00%	75.00%	80.00%
Paradise turbidity four PTS correcting anomalous points	40.00%	45.00%	50.00%	75.00%	75.00%	80.00%

Table B.5. ANN false negative rates for specific time steps predictions using Paradise measurements as inputs

Input Data Description	Confidence Interval					
	30%	40%	50%	90%	95%	99%
Paradise site Dissolved Oxygen	65%	80%	80%	100%	100%	100%
Dissolved Oxygen + one PTS of Paradise turbidity without correcting anomalous points	45%	50%	50%	75%	80%	80%
Dissolved Oxygen + one PTS of Paradise turbidity correcting anomalous points	50%	55%	60%	75%	80%	80%
Dissolved Oxygen + four PTS of Paradise turbidity without correcting anomalous points	35%	45%	50%	75%	75%	80%
Dissolved Oxygen + four PTS of Paradise turbidity correcting anomalous points	40%	50%	55%	75%	75%	80%
Paradise site Specific Conductance	80%	80%	80%	95%	100%	100%
Specific Conductance + one PTS of Paradise turbidity without correcting anomalous points	40%	50%	55%	75%	80%	80%
Specific Conductance + one PTS of Paradise turbidity correcting anomalous points	45%	55%	60%	70%	80%	80%
Specific Conductance + four PTS of Paradise turbidity without correcting anomalous points	35%	40%	50%	75%	75%	80%
Specific Conductance + four PTS of Paradise turbidity correcting anomalous points	40%	45%	55%	70%	75%	80%
Paradise site pH	80%	80%	80%	100%	100%	100%
pH + one PTS of Paradise turbidity without correcting anomalous points	45%	50%	50%	75%	80%	80%
pH + one PTS of Paradise turbidity correcting anomalous points	50%	55%	60%	75%	80%	80%
pH + four PTS of Paradise turbidity without correcting anomalous points	35%	45%	50%	75%	75%	80%
pH + four PTS of Paradise turbidity correcting anomalous points	40%	50%	55%	75%	75%	80%
Paradise site Temperature	75%	80%	80%	100%	100%	100%
Temperature+ one PTS of Paradise turbidity without correcting anomalous points	40%	45%	50%	75%	75%	80%
Temperature + one PTS of Paradise turbidity correcting anomalous points	50%	50%	55%	75%	75%	80%
Temperature + four PTS of Paradise turbidity without correcting anomalous points	35%	45%	50%	75%	75%	80%
Temperature + four PTS of Paradise turbidity correcting anomalous points	40%	50%	55%	75%	75%	80%
Paradise site Discharge	80%	80%	80%	95%	100%	100%
Discharge + one PTS of Paradise turbidity without correcting anomalous points	35%	50%	55%	75%	80%	80%
Discharge + one PTS of Paradise turbidity correcting anomalous points	45%	55%	60%	75%	80%	80%
Discharge + four PTS of Paradise turbidity without correcting anomalous points	35%	40%	50%	75%	75%	80%
Discharge + four PTS of Paradise turbidity correcting anomalous points	40%	45%	55%	70%	75%	80%

Table B.6. ANN false negative rates for specific time steps prediction using upstream measurements as inputs

Input Data Description	Confidence Interval					
	30%	40%	50%	90%	95%	99%
Confluence site turbidity	80.00%	80.00%	80.00%	90.00%	90.00%	100.00%
Confluence turbidity + one PTS of Paradise turbidity correcting anomalous points	40.00%	55.00%	60.00%	75.00%	75.00%	80.00%
Confluence turbidity + four PTS of Paradise turbidity correcting anomalous points	40.00%	45.00%	50.00%	75.00%	75.00%	80.00%
Lower South Fork site turbidity	80.00%	80.00%	80.00%	90.00%	90.00%	100.00%
Lower South Fork site turbidity + one PTS of Paradise turbidity correcting anomalous points	40.00%	50.00%	55.00%	75.00%	75.00%	80.00%
Lower South Fork site turbidity + four PTS of Paradise turbidity correcting anomalous points	40.00%	45.00%	50.00%	70.00%	75.00%	80.00%
Paradise site Discharge + four PTS of Paradise turbidity + Confluence site turbidity	40.00%	50.00%	55.00%	75.00%	75.00%	80.00%

Appendix C

RVM Specific Time Steps False Positive and False Negative Rates

Appendix C shows the false positive and false negative rates for all the cases evaluated for case I, case II and case III individual time step prediction using RVM models. Only the best results for each of the cases are displayed in the Results and Discussion section.

Table C.1. RVM false positive rates for specific time steps prediction using previous time steps as inputs

Input Data Description	Confidence Interval					
	30%	40%	50%	90%	95%	99%
Paradise turbidity one PTS without correcting anomalous points	0.63%	0.63%	0.56%	0.35%	0.35%	0.28%
Paradise turbidity one PTS correcting anomalous points	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Paradise turbidity four PTS without correcting anomalous points	2.32%	2.32%	2.18%	1.55%	1.48%	1.41%
Paradise turbidity four PTS correcting anomalous points	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Table C.2. RVM false positive rates for specific time steps predictions using Paradise measurements as inputs

Input Data Description	Confidence Interval					
	30%	40%	50%	90%	95%	99%
Paradise site Dissolved Oxygen	39.15%	0.00%	0.00%	0.00%	0.00%	0.00%
Dissolved Oxygen + one PTS of Paradise turbidity without correcting anomalous points	0.70%	0.00%	0.00%	0.00%	0.00%	0.00%
Dissolved Oxygen + one PTS of Paradise turbidity correcting anomalous points	0.70%	0.00%	0.00%	0.00%	0.00%	0.00%
Dissolved Oxygen + four PTS of Paradise turbidity without correcting anomalous points	40.70%	20.92%	6.13%	0.28%	0.28%	0.28%
Dissolved Oxygen + four PTS of Paradise turbidity correcting anomalous points	17.39%	4.86%	0.00%	0.00%	0.00%	0.00%
Paradise site Specific Conductance	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Specific Conductance + one PTS of Paradise turbidity without correcting anomalous points	0.77%	0.70%	0.63%	0.49%	0.35%	0.35%
Specific Conductance + one PTS of Paradise turbidity correcting anomalous points	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Specific Conductance + four PTS of Paradise turbidity without correcting anomalous points	2.67%	2.04%	1.76%	1.27%	1.27%	1.20%
Specific Conductance + four PTS of Paradise turbidity correcting anomalous points	0.63%	0.14%	0.00%	0.00%	0.00%	0.00%
Paradise site pH	0.49%	0.07%	0.07%	0.00%	0.00%	0.00%
pH + one PTS of Paradise turbidity without correcting anomalous points	0.85%	0.85%	0.77%	0.63%	0.63%	0.49%
pH + one PTS of Paradise turbidity correcting anomalous points	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
pH + four PTS of Paradise turbidity without correcting anomalous points	2.54%	2.46%	2.39%	1.48%	1.48%	1.41%
pH + four PTS of Paradise turbidity correcting anomalous points	53.59%	20.35%	0.14%	0.00%	0.00%	0.00%
Paradise site Temperature	45.49%	0.00%	0.00%	0.00%	0.00%	0.00%
Temperature+ one PTS of Paradise turbidity without correcting anomalous points	7.81%	0.00%	0.00%	0.00%	0.00%	0.00%
Temperature + one PTS of Paradise turbidity correcting anomalous points	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Temperature + four PTS of Paradise turbidity without correcting anomalous points	41.41%	14.86%	1.13%	0.00%	0.00%	0.00%
Temperature + four PTS of Paradise turbidity correcting anomalous points	0.70%	0.35%	0.00%	0.00%	0.00%	0.00%
Paradise site Discharge	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Discharge + one PTS of Paradise turbidity without correcting anomalous points	0.77%	0.77%	0.63%	0.49%	0.42%	0.35%
Discharge + one PTS of Paradise turbidity correcting anomalous points	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Discharge + four PTS of Paradise turbidity without correcting anomalous points	1.90%	1.76%	1.69%	1.34%	1.34%	1.34%
Discharge + four PTS of Paradise turbidity correcting anomalous points	0.07%	0.00%	0.00%	0.00%	0.00%	0.00%

Table C.3. RVM false positive rates for specific time steps prediction using upstream measurements as inputs

Input Data Description	Confidence Interval					
	30%	40%	50%	90%	95%	99%
Confluence site turbidity	0.07%	0.07%	0.00%	0.00%	0.00%	0.00%
Confluence turbidity + one PTS of Paradise turbidity correcting anomalous points	0.14%	0.07%	0.07%	0.00%	0.00%	0.00%
Confluence turbidity + four PTS of Paradise turbidity correcting anomalous points	0.07%	0.07%	0.00%	0.00%	0.00%	0.00%
Lower South Fork site turbidity	0.63%	0.21%	0.00%	0.00%	0.00%	0.00%
Lower South Fork site turbidity + one PTS of Paradise turbidity correcting anomalous points	0.07%	0.07%	0.00%	0.00%	0.00%	0.00%
Lower South Fork site turbidity + four PTS of Paradise turbidity correcting anomalous points	2.04%	1.83%	0.08%	0.00%	0.00%	0.00%
Paradise site Discharge + four PTS of Paradise turbidity + Confluence site turbidity	0.28%	0.14%	0.00%	0.00%	0.00%	0.00%

Table C.4. RVM false negative rates for specific time steps prediction using previous time steps as inputs

Input Data Description	Confidence Interval					
	30%	40%	50%	90%	95%	99%
Paradise turbidity one PTS without correcting anomalous points	35.00%	45.00%	55.00%	75.00%	75.00%	80.00%
Paradise turbidity one PTS correcting anomalous points	40.00%	50.00%	60.00%	75.00%	75.00%	80.00%
Paradise turbidity four PTS without correcting anomalous points	35.00%	35.00%	40.00%	60.00%	70.00%	75.00%
Paradise turbidity four PTS correcting anomalous points	35.00%	40.00%	45.00%	60.00%	70.00%	75.00%

Table C.5. RVM false negative rates for specific time steps predictions using Paradise measurements as inputs

Input Data Description	Confidence Interval					
	30%	40%	50%	90%	95%	99%
Paradise site Dissolved Oxygen	65%	80%	80%	100%	100%	100%
Dissolved Oxygen + one PTS of Paradise turbidity without correcting anomalous points	40%	50%	50%	75%	75%	75%
Dissolved Oxygen + one PTS of Paradise turbidity correcting anomalous points	40%	50%	50%	70%	75%	75%
Dissolved Oxygen + four PTS of Paradise turbidity without correcting anomalous points	45%	60%	60%	75%	75%	80%
Dissolved Oxygen + four PTS of Paradise turbidity correcting anomalous points	35%	50%	55%	75%	75%	75%
Paradise site Specific Conductance	80%	80%	80%	90%	100%	100%
Specific Conductance + one PTS of Paradise turbidity without correcting anomalous points	35%	45%	50%	70%	75%	75%
Specific Conductance + one PTS of Paradise turbidity correcting anomalous points	35%	40%	45%	75%	75%	75%
Specific Conductance + four PTS of Paradise turbidity without correcting anomalous points	35%	35%	45%	75%	75%	75%
Specific Conductance + four PTS of Paradise turbidity correcting anomalous points	40%	40%	50%	70%	75%	75%
Paradise site pH	70%	85%	85%	100%	100%	100%
pH + one PTS of Paradise turbidity without correcting anomalous points	45%	50%	50%	75%	80%	80%
pH + one PTS of Paradise turbidity correcting anomalous points	40%	50%	50%	75%	75%	75%
pH + four PTS of Paradise turbidity without correcting anomalous points	35%	45%	50%	75%	75%	80%
pH + four PTS of Paradise turbidity correcting anomalous points	35%	55%	55%	70%	75%	75%
Paradise site Temperature	70%	80%	80%	95%	100%	100%
Temperature+ one PTS of Paradise turbidity without correcting anomalous points	45%	50%	55%	70%	75%	75%
Temperature + one PTS of Paradise turbidity correcting anomalous points	50%	50%	55%	75%	75%	80%
Temperature + four PTS of Paradise turbidity without correcting anomalous points	40%	50%	55%	70%	75%	75%
Temperature + four PTS of Paradise turbidity correcting anomalous points	50%	50%	55%	75%	75%	75%
Paradise site Discharge	80%	80%	80%	90%	100%	100%
Discharge + one PTS of Paradise turbidity without correcting anomalous points	35%	35%	35%	70%	75%	75%
Discharge + one PTS of Paradise turbidity correcting anomalous points	40%	40%	45%	65%	70%	75%
Discharge + four PTS of Paradise turbidity without correcting anomalous points	35%	35%	35%	55%	70%	75%
Discharge + four PTS of Paradise turbidity correcting anomalous points	35%	40%	45%	60%	70%	75%

Table C.6. RVM false negative rates for specific time steps prediction using upstream measurements as inputs

Input Data Description	Confidence Interval					
	30%	40%	50%	90%	95%	99%
Confluence site turbidity	80.00%	80.00%	80.00%	85.00%	90.00%	100.00%
Confluence turbidity + one PTS of Paradise turbidity correcting anomalous points	45.00%	55.00%	60.00%	75.00%	75.00%	80.00%
Confluence turbidity + four PTS of Paradise turbidity correcting anomalous points	40.00%	40.00%	40.00%	60.00%	70.00%	75.00%
Lower South Fork site turbidity	80.00%	80.00%	80.00%	90.00%	90.00%	100.00%
Lower South Fork site turbidity + one PTS of Paradise turbidity correcting anomalous points	40.00%	40.00%	50.00%	65.00%	70.00%	75.00%
Lower South Fork site turbidity + four PTS of Paradise turbidity correcting anomalous points	40.00%	40.00%	50.00%	70.00%	70.00%	75.00%
Paradise site Discharge + four PTS of Paradise turbidity + Confluence site turbidity	45.00%	55.00%	60.00%	75.00%	75.00%	80.00%

Appendix D

ANN Bootstrap Analysis Results

Appendix D shows results for bootstrap analyses performed on the cases which obtained the best results for test set prediction when using the ANN model.

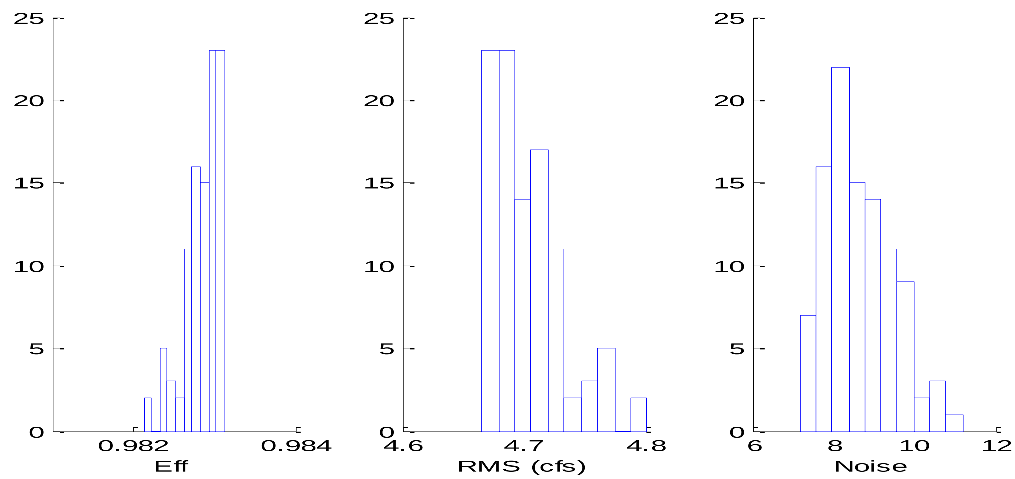

Figure D.1. ANN bootstrap analysis result using one previous time step of Paradise turbidity as input.

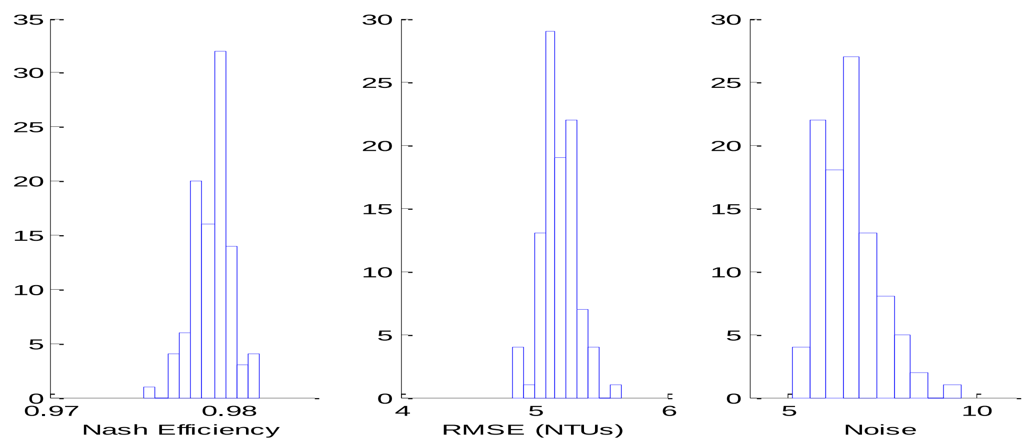

Figure D.2. ANN bootstrap analysis result using four previous time steps of Paradise turbidity as inputs.

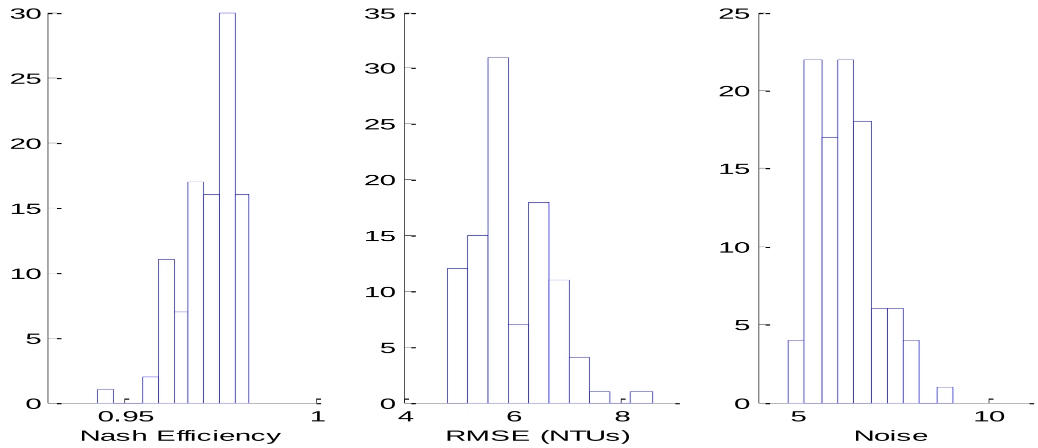

Figure D.3. ANN bootstrap analysis result using four previous time steps of Paradise turbidity and specific conductance as inputs.

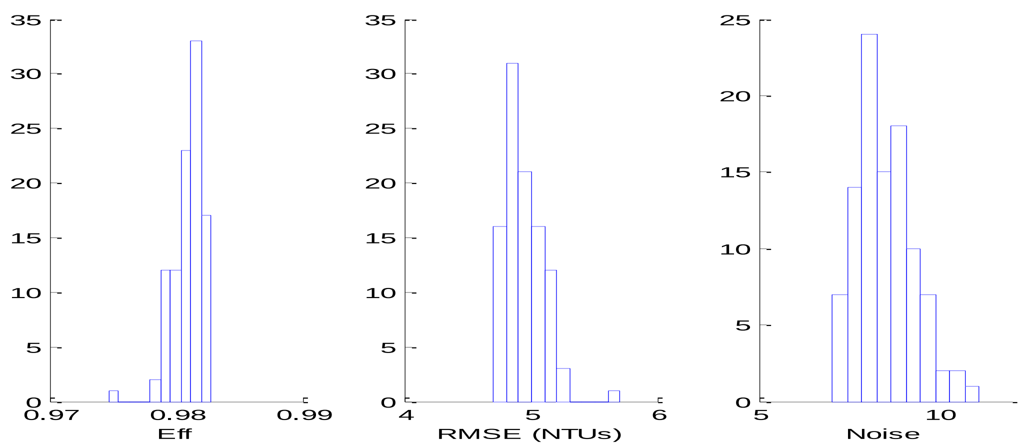

Figure D.4. ANN bootstrap analysis result using one previous time step of Paradise turbidity and pH measurements as inputs.

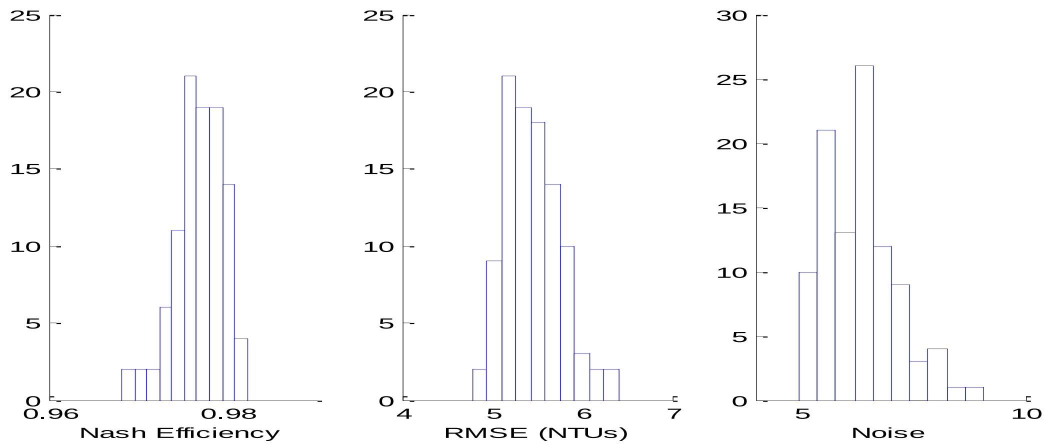

Figure D.5. ANN bootstrap analysis result using four previous time steps of Paradise turbidity and discharge measurements as inputs.

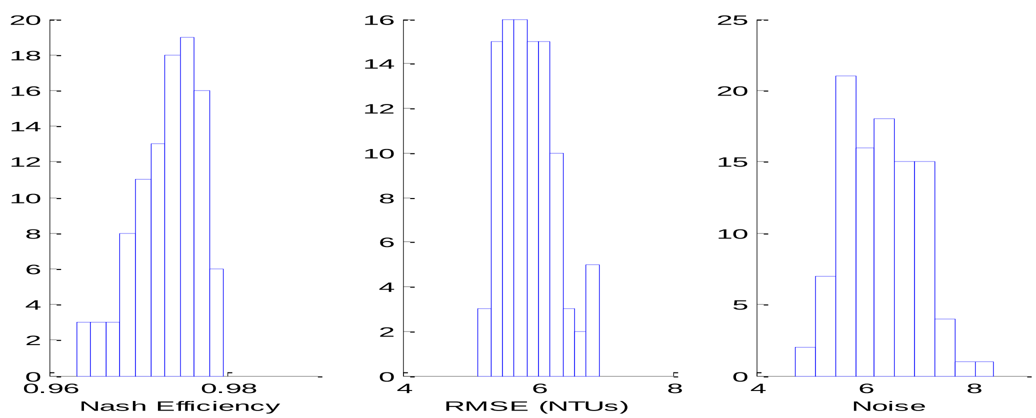

Figure D.6. ANN bootstrap analysis result using four previous time steps of Paradise turbidity and Confluence site turbidity measurements as inputs.

Appendix E

RVM Bootstrap Analysis Results

Appendix E shows results for bootstrap analyses performed on the cases which obtained the best results for test set prediction when using the RVM model.

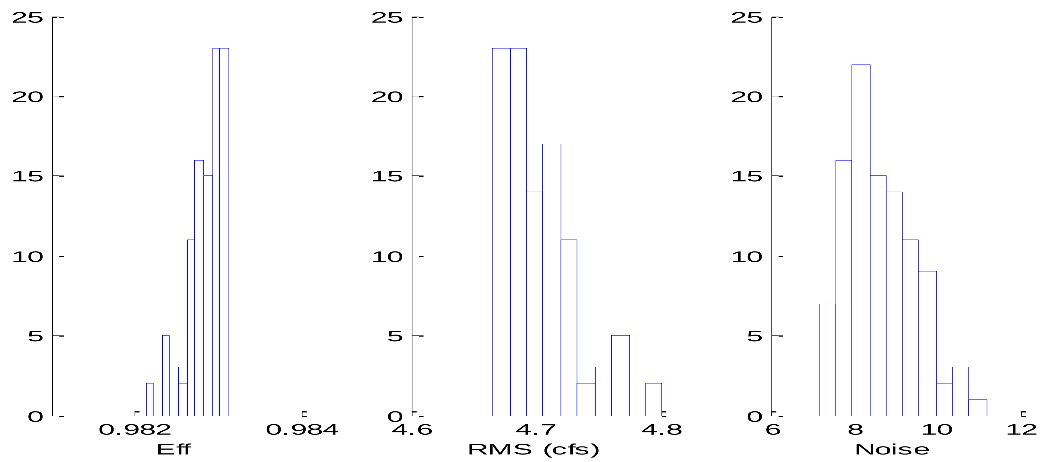

Figure E.1. RVM bootstrap analysis result using one previous time step of Paradise turbidity as inputs.

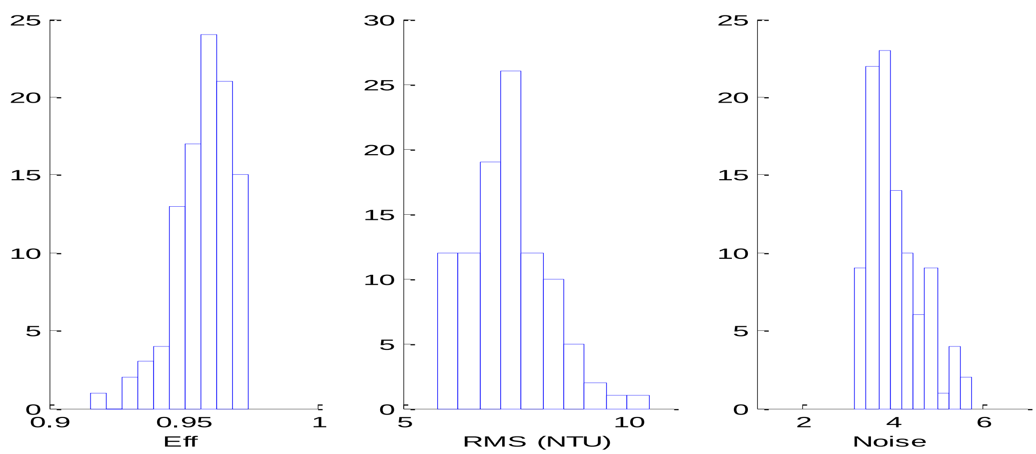

Figure E.2. RVM bootstrap analysis result using four previous time steps of Paradise turbidity as inputs.

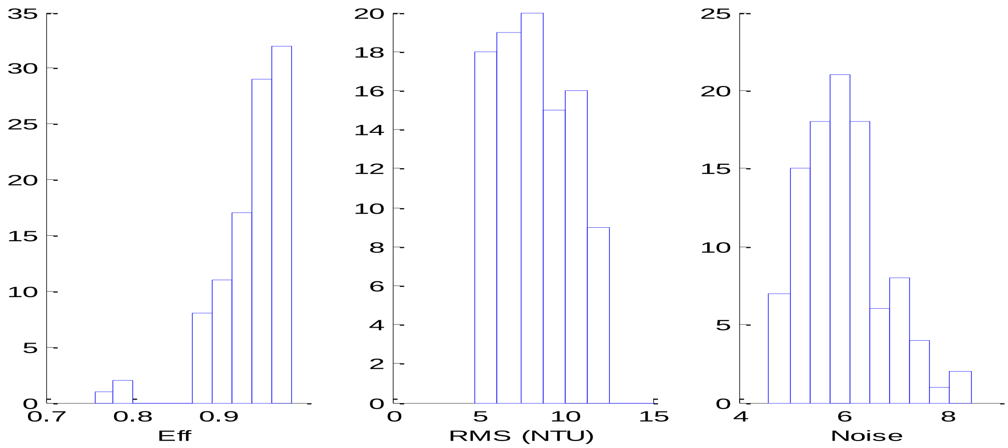

Figure E.3. RVM bootstrap analysis result using four previous time steps of Paradise turbidity and specific conductance as inputs.

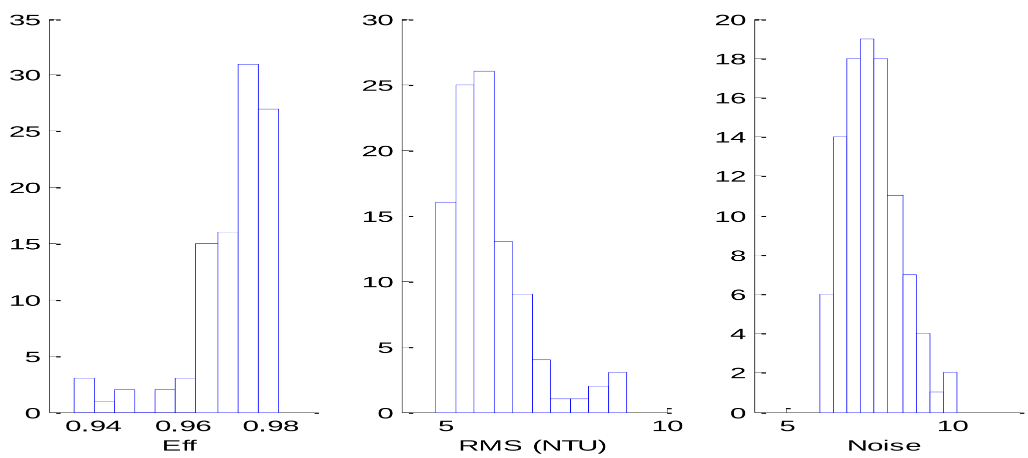

Figure E.4. RVM bootstrap analysis result using one previous time step of Paradise turbidity and pH measurements as inputs.

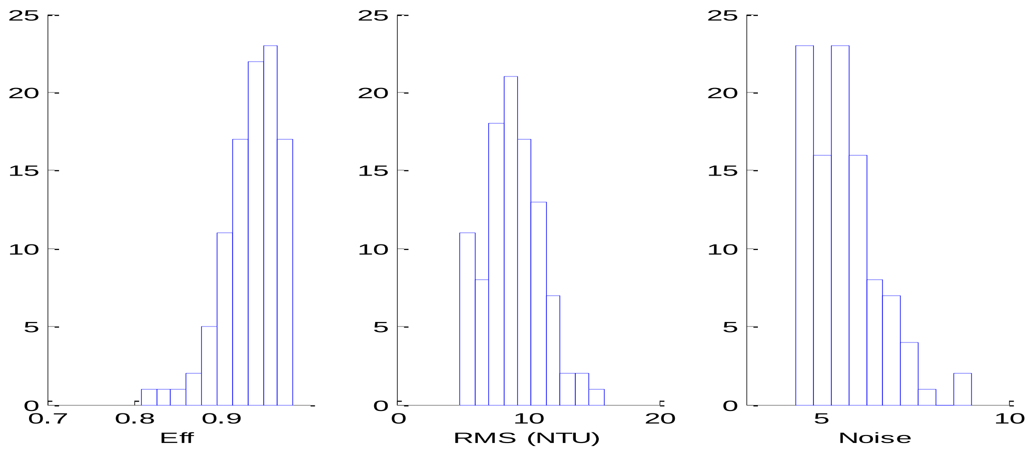

Figure E.5. RVM bootstrap analysis result using four previous time steps of Paradise turbidity and discharge measurements as inputs.

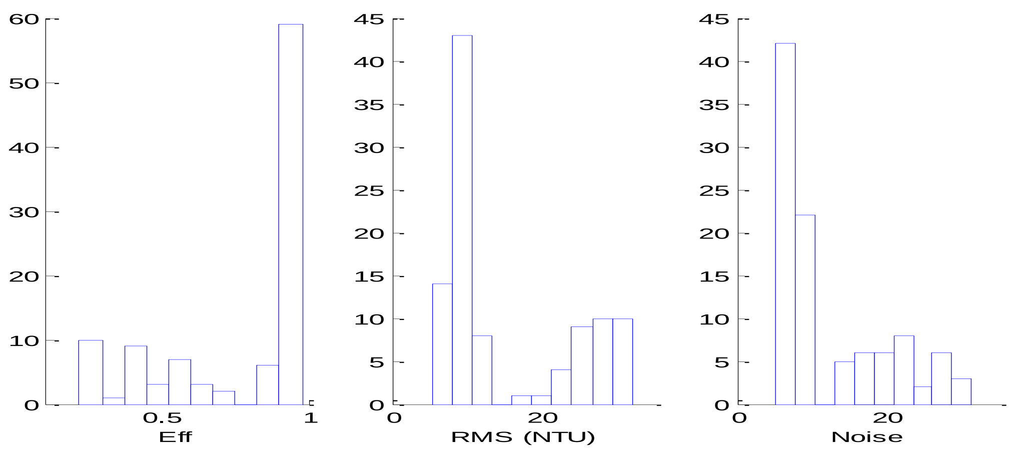

Figure E.6. RVM bootstrap analysis result using four previous time steps of Paradise turbidity and Confluence site turbidity measurements as inputs.